Review Article

The Importance of Feature Processing in Deep-Learning-Based Condition Monitoring of Motors

Dileep Kumar Soother,1 Jawaid Daudpoto,2 Nicholas R. Harris,3 Majid Hussain,1 Sanaullah Mehran,1 Intiaz Hussain Kalwar,4 Tanweer Hussain,1 and Tayab Din Memon5,6

1National Centre of Robotics and Automation, HHCMS Lab, Mehran University of Engineering & Technology, Jamshoro 76020, Pakistan
2Department of Mechatronic Engineering, Mehran University of Engineering & Technology, Jamshoro 76020, Pakistan
3School of Electronics and Computer Science, Southampton University, SO32 1PH, Southampton, UK
4Department of Electrical Engineering, DHA Suffa University, Karachi, Pakistan
5Department of Electronic Engineering, Mehran University of Engineering & Technology, Jamshoro 76020, Pakistan
6School of Information Technology and Engineering (SITE), Melbourne Institute of Technology, Melbourne, Australia

Correspondence should be addressed to Nicholas R. Harris; nrh@ecs.soton.ac.uk

Received 18 March 2021; Revised 14 April 2021; Accepted 25 April 2021; Published 8 May 2021

Academic Editor: Dao B. Wang

Copyright © 2021 Dileep Kumar Soother et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The advent of deep learning (DL) has transformed diagnosis and prognosis techniques in industry. It has allowed tremendous progress in industrial diagnostics, has been playing a pivotal role in maintaining and sustaining Industry 4.0, and is also paving the way for Industry 5.0. It has become prevalent in the condition monitoring of industrial subsystems, a prime example being motors. Motors in various applications start deteriorating due to various reasons. Thus, the monitoring of their condition is of prime importance for sustaining the operation and maintaining efficiency. This paper presents a state-of-the-art review of DL-based condition monitoring for motors in terms of input data and feature processing techniques. Particularly, it reviews the application of various input features for the effectiveness of DL models in motor condition monitoring in the sense of what problems are targeted using these feature processing techniques and how they are addressed. Furthermore, it discusses and reviews advances in DL models, DL-based diagnostic methods for motors, hybrid fault diagnostic techniques, points out important open challenges to these models, and signposts the prospective future directions for DL models. This review will assist researchers in identifying research gaps related to feature processing, so that they may effectively contribute toward the implementation of DL models as applied to motor condition monitoring.

1. Introduction

Condition monitoring is described as a continuous process of diagnosis that allows prevention of unintended failure of a system. The basic principle of condition monitoring is to indicate the occurrence of deterioration by taking physical measurements at regular intervals. Subsequently, diagnosis procedures allow the planning of rectification strategies [1–4]. The primary reasons for the application of condition monitoring are increasing availability, prevention of damage, increased reliability, process optimization, increased time between outages, reduced production loss, and better operator information or insights. Its application leads to the development of prognostics, which allows for the estimation of the system’s future health and the prediction of the remaining useful life of the system or system’s components [5–8]. Motors are the backbone of industry; they start degrading due to different reasons such as long period of operation, variations of power supply, or harsh environment; which gradually lead to permanent damage [9–11]. Consequently, it becomes crucial to monitor the operation continuously.

In the past, there has been extensive research relating to anomaly detection, anomaly severity level detection, and
detecting failing elements of the motor. Subsequently, efforts have been made to integrate these diagnosis and prognosis methods, which in turn increases the amount of data. Although condition monitoring system integration improves performance and increases the data volume (providing richer information), it poses different shortcomings such as increased complexity in the information correlating process and increased level of uncertainty [12]. Consequently, this requires novel approaches that can address these shortcomings leading to improved performance. AI-based approaches have been extensively used in the field of condition monitoring for many years [13–16]. With continuous progress in various models of AI, it progressed into machine learning (ML) and subsequently developed into deep learning (DL), which has driven significant impact in the development of modern industry, transportation, medical, and other domains.

The DL-fuelled condition monitoring of motors has set up new horizons in industry 4.0 and has been paving the way for industry 5.0. DL algorithms have impacted almost every area including business [17], medical sciences [18], natural language processing (NLP) [19], robotics [20], transportation [21], the power sector [22], and many other sectors of the modern world. The concept of DL was first coined as “deep belief networks” (DBN) in 2006 and was considered as one of the major breakthroughs in the world of technology [23]. The ability to learn data representation becomes significant with the application of DL models and makes it very attractive in the arena of intelligent diagnosis and prognosis [24, 25]. In comparison to conventional ML models, which can require significant effort in manual feature design and optimization, DL models can automatically extract the representations from the data.

In recent years, DL-based diagnosis and prognosis methods have outperformed conventional machine learning algorithms owing to their generalized nature and many other advantages such as end-to-end implementation, model upgradability, and representation learning using raw data. It does not require human knowledge or intervention in feature designing. Its performance improves with volume of data, but it requires high performance computing hardware such as graphics processing units (GPU) to inherently perform the intense number of necessary matrix multiplication operations. It advocates end-to-end problem-solving rather than dividing the problem into steps. DL-based diagnosis reduces cross-domain discrepancy by learning data representations with multiple levels of abstraction [26–29]. Compared to ML models, DL models can achieve superior performance and their classification accuracies have been tending to 100%. These benefits of DL models have attracted the attention of researchers and they have been extensively applying these models in their domains. Various DL-based condition monitoring methods for motors have been reported by researchers [30, 31]. Meanwhile, extensive adoption of DL methods has enabled industry to progress with better efficiency and sustainability [32]. DL algorithms have efficiently predicted the state of motors in industrial condition monitoring systems with a wide scope. In fact, some models such as GANs can generate data through learning a model of input distribution that is useful in cases when the dataset is small for diagnosis. These algorithms have been in the limelight owing to their merits relating to industry 4.0 and industry 5.0 [27, 33]. DL is reshaping itself through variations in architecture of models, which in turn is reshaping condition monitoring systems by adding more capabilities such as high reorganization accuracy, noise dealing capability, and end-to-end system implementation.

DL-based intelligent diagnosis of various industrial subsystems has witnessed remarkable improvement in performance [28, 34]. However, the disadvantage of DL models lies in feature processing and selection of its parameters such as the learning rate, momentum, and number of neurons in layers. Parameter selection is always a time-consuming and challenging task, which is often achieved through hit-and-trial methods [29, 35]. There have been different reviews and surveys conducted by the researchers relating to condition monitoring of motors and application of DL in this field [11, 36–40]. However, each review has been conducted in different contexts. For example, in [40] authors have surveyed applications of ML and DL in condition monitoring of various machines in the context of vibration data as a key factor for the surveyed studies. Meanwhile, Choudhary et al. [11] have conducted a review of various faults, which occur in induction motors based on various input data types. Authors have also briefly reviewed the techniques used in condition monitoring of induction motors such as Fuzzy Logic, Artificial neural networks, Neuro-Fuzzy inference systems, and support vector machines. They found that noninvasive techniques such as thermal imaging are overcoming the conventional condition monitoring methods. On the other hand, in [41] authors have briefly reviewed various condition monitoring techniques particularly based on current, vibration, and acoustic signals. There is therefore still a need for a comprehensive review on the applications of DL models in condition monitoring of motors. Furthermore, the article presents open challenges and future research directions to promote the application of DL models in engineering scenarios. Compared to the existing reviews, this review focuses on input data and feature-processing techniques used for effective fault diagnosis in the field of DL-based condition monitoring of motors. It surveys and summarizes the recent developments in actual applications of various feature-processing techniques in DL-based condition monitoring of motors. The contribution of this article can be summarized as follows:

(i) Provides an integrated overview of current trends and consolidates the recent work of various researchers related to feature processing in DL-based condition monitoring of motors

(ii) Endeavours to provide an in-depth analysis approach and a valuable road map to engineers and researchers working in this field, which may assist them in signposting the direction of future research
(iii) Presents the merits and limitations of DL for motor condition monitoring methods based on state-of-the-art contributions reported in the literature.

The remainder of this article is constructed as follows. Section 2 reviews feature processing methods used with DL models in condition monitoring of motors, and the techniques used to resolve problems posed by such methods. Section 3 summarizes and discusses performance aspects of models, highlights challenges related to DL, and presents future directions of this field, and Section 4 provides concluding remarks on this review article.

2. Feature Processing for DL-Based Condition Monitoring of Motors

In DL-based motor condition monitoring, data are acquired using various sensors and stored in a database. Typically, vibration, acoustic emission, current and temperature signals are used to monitor the condition of a machine [42]. Each diagnostic method has different capabilities for detecting various types of faults in motors. In the next step, data are preprocessed, and a model is built. Subsequently, the model is trained on data. The end goals or targets are typically defined as fault detection, failure prediction, and remaining useful life (RUL) estimation but are not limited to these cases. Figure 1 illustrates the generalized concept of the DL model pipeline and its comparison with the machine learning (ML) model pipeline. Feature processing is applied to extract buried features in noisy input data through feature extraction and feature selection techniques. After performing feature processing, data are fed to various ML models such as support vector machine (SVM), decision trees (DT), random forest (RF), and K-nearest neighbour (KNN). Moreover, feature processing assists in identifying and removing outliers and redundant information from the dataset and converting raw data to more manageable groups for processing. In addition, it reduces the dimensionality of large datasets, which in turn speeds up the learning process [43–45]. Feature processing, which includes both feature extraction and selection techniques, is very important for deploying reliable ML-based solutions for industry. It not only simplifies the entire process of tackling a domain-specific problem but also makes the process understandable for human experts. Reducing the number of features, or enabling the selection of useful ones, greatly reduces the hardware dependence and the need of highly nonlinear function mapping by the ML models.

Feature extraction techniques include: (i) time-domain techniques, such as statistical parameters, including mean, standard deviation, root mean square (RMS), covariance, kurtosis, and skewness. (ii) Frequency domain features, including fast Fourier transform (FFT), spectral kurtosis, root variance frequency, and mean frequency. (iii) Time-frequency domain features, including wavelet transform (WT), short-time Fourier transform (STFT), Hilbert-Huang transform (HHT), Hilbert transform (HT), and empirical model decomposition (EMD). These techniques basically provide DL models with meaningful representation of raw data and aid in the simplification of decision boundary definition identified by classification models. In the case of regression-based models, these techniques help lower the order of function to be mapped by DL models, which result in computational efficiency and robust real-time deployment. Popular feature selection techniques are principal component analysis (PCA), linear discriminant analysis (LDA), and independent component analysis (ICA) [41, 46]. These techniques are particularly known for their ability to reduce the dimensionality of the dataset. PCA produces an estimated representation of a bulk of features in a concise manner and reduces the feature space. This helps not only in optimization but also aids in enhancing the reliability and explainability of DL models by visualizing the feature space possible.

DL models can learn representations from raw input data and can process accordingly for supervised or unsupervised tasks as shown in Figure 1. Generally, it does not involve steps such as feature extraction and feature selection. Hence, DL-based algorithms have been extensively used in the field of condition monitoring of motors. The various architectures of DL that have been used in motor condition monitoring include deep belief networks (DBN), multilayer perceptron (MLP), autoencoders (AE), deep Boltzmann machine (DBN), convolutional neural networks (CNNs), recurrent neural networks (RNN), and generative adversarial neural networks (GANs). Each of these architectures was developed keeping in mind the ways data could be presented to them. For example, an MLP is designed to learn from tabular data, a CNN is traditionally used to map a 2D field (which in most cases is an image) to an output variable, RNN in the same way was designed to handle sequences of data, making it an attractive paradigm for time-series forecasting or speech-processing tasks. However, these architectures do allow a certain level of flexibility and have been used for alternative tasks. For example, an image could be broken into a sequence of pixels and presented as tabular data to an MLP, but this comes at the cost of explainability. Visualization of a row of image pixels into feature space would be incomprehensible. This lays more emphasis on feature-processing techniques to be used along with a given architecture because proper feature formation is necessary for even considering the trade-off between accuracy and explainability of the models. Researchers have put efforts for optimizing the performance of these methods through different feature-processing techniques which play a key role in the DL-based condition monitoring of motors. In this section, each of these techniques is reviewed with respect to problems, which are addressed by these methods, and how successful they were at the task.

2.1. Multilayer Perceptron (MLP). MLP is one of the most utilised DL model topologies, and also one of the oldest. It is a fully connected neural network consisting of one or more hidden layers. It can easily perform classification tasks with its simple structure. However, it becomes difficult to train the model as tasks become complex owing to the increase in
datasets. It is trained through a supervised learning approach called back-propagation (BP). Figure 2 shows the basic structure of the MLP model.

The MLP has been used in motor condition monitoring applications for a long time owing to its easy training process. Various authors have employed it for motor fault classification tasks. Vieira et al. [47] have employed multiple MLPs with single layer to classify stator winding short-circuit faults of the induction motor using the current data. The dataset was collected by varying the operational conditions of the motor/load/converter. The authors have investigated multiple hypotheses related to the conditions and their effects on the classifier’s accuracy. The MLP model was able to achieve 92.6% accuracy at zero load and 76.9% accuracy at full load. It was observed in the results that variations in the operating frequency did not affect the classifier’s accuracy. However, it was harder to correctly classify the faults with low severity, which led to the observation that the accuracy of the classifier increases as the fault severity level increases.

Palacios et al. [48] have used MLP to carry out motor fault classification using current and voltage amplitudes in induction motors. The authors have used discretized time domain signals for classifying faults such as rotor broken-bar fault (1/2/3/4 broken-bars), bearing faults (inner race grooves, outer race grooves, and rolling element defect), and stator short-circuit fault. Results confirmed a better performance of the method, which achieved 94.6% accuracy compared to other conventional methods including SVM and KNN. It was also observed that the proposed method was computationally more efficient than the comparative methods. In [49], authors have used MLP with mutual
information (MI) for fault classification of the induction motor. MI was employed for feature extraction, and it describes the similarity between time-series data simultaneously acquired from the setup. The method was applied to detect the stator winding short-circuit fault using the features extracted from the current signature data. Experimental results confirmed that the MLP performed better than the radial basis function (RBF) in terms of classifying healthy and faulty conditions of the motor with accuracy over 99%, whereas RBF achieved 96–97% accuracy.

Zolfaghari et al. [50] have used an MLP-based classifier for broken bar fault detection. They have input motor current data after performing FFT and wavelet packet transform, for superior feature extraction, which in turn reduced the burden on the classifier. Furthermore, they have extracted WP-based statistical parameters for intelligent broken-bar severity level detection (one/two/three broken bars). The investigation results confirmed that the method yields promising results in terms of rotor broken-bar severity-level detection, even in the no-load condition with accuracy of 98.8%. In [51], authors have used an MLP model for real-time fault classification of the induction motor. Authors have employed a nonlinear manifold technique called curvilinear component-based analysis (CCA) for feature extraction. The method was employed to classify two major faults: stator interturn fault and rotor broken-bar fault. Results revealed that the MLP model effectively classified the faults in the induction motor with an accuracy of 95.2%.

Bazan et al. [52] used an MLP model with the MI feature extraction method to classify induction motor bearing faults. MI shows the reduction in uncertainty associated with one random variable when combined with information from another variable in simultaneously acquired time-series data. Features were extracted from multiple current signatures under different operating conditions. The results confirmed the robustness of the method in the bearing fault classification compared to the conventional methods, including SVM and KNN. The MLP network achieved 90.5% accuracy with 10% voltage unbalance condition compared to SVM and KNN, which achieved 84% and 83.3%, respectively. Authors have suggested future research work in terms of investigating the real-time implementation of the approach and the potential accuracy trade-off with the speed of the motor. Recent applications of MLP for condition monitoring of motors are summarized in Table 1.

2.2. Autoencoders (AE). AEs are frequently employed as an unsupervised feature extraction technique. They have the capability of reducing the dimensions of input data whilst retaining most of the information. Figure 3 illustrates a deep AE structure and its operation. It has two blocks: the first block is the AE encoder and the second block is the decoder. The encoder network generates low-dimension representations while the decoder network reconstructs the input data from these low-dimension representations. In addition, it uses the reconstruction error as a loss function. Deep AE is initially trained using an unsupervised method that is known as pretraining. The pretraining process allows better convergence as it reduces overfitting and optimises the layers. The process of pretraining of a deep AE model is explained as follows:

(i) A single layer extracts an initial parameter for the following hidden layer and predicts itself using the input vector. Through this technique, it learns about the data without any feedback or labeled data. Subsequently, it stores the learned features as weights for the hidden layers of the network.

(ii) Similarly, the following layer learns about features for succeeding hidden layers and the process is continuous for all the remaining layers.

(iii) Eventually, learned information reaches the output layer typically through a softmax activation function to give a probability distribution.

After this pretraining process, both the networks can be trained simultaneously for classification. Many researchers have argued that AE can be trained without pretraining, but this requires some additional effort in training process or initialization [53].

The recent developments in AE by researchers have presented various derivatives of AE for 1-D and 2-D data such as denoising autoencoders (DAE) [54], variational autoencoders (VAE) [55], and sparse autoencoders (SAE) [56]. Although DAE and SAE have the same structure as a simple AE, they differ in loss function and inputs. Some research has been carried out to improve the performance of AEs using different techniques and have been employed for diagnosis and prognosis of motors. AE and its derivatives have been used in motor condition monitoring with different goals. Sun et al. [57] have employed SAE with DNN for unsupervised feature extraction. The model was fed with vibration data to classify different faults of the induction motor. They employed the “dropout” regularization method to avoid overfitting during the training process. The SAE remained inactive during the testing process. The results confirmed that the approach provided better performance with 97.6% accuracy compared to conventional models such as SVM and linear regression (LR), which achieved 96.4% and 92.7% accuracy, respectively. Liu et al. [58] have used STFT and SAE to extract features from the sound/acoustic emissions signals of a rolling bearing. Considering barriers to feature extraction in the DNN, they have used STFT for fast and effective feature extraction. This enabled fault classification with fewer training data samples and resulted in increased classifier accuracy of the DNN compared to the model without STFT features. The SAE model with the STFT-based input achieved average accuracy of 96.2%.

In [59], authors have performed bearing fault classification using AE and extreme learning machines (ELM). ELM is employed with AE owing to its advantage of high training speed. No explicit feature extraction technique was involved with the employed method. The AE-ELM model input data were frequency features extracted from vibration data. This approach yielded better classification performance
than SAE and results also revealed that it was faster than SAEs. Both the AE-ELM and SAE had an accuracy of 100%. However, the SAE model had a slower response time of 20s in the online learning compared to AE-ELM, which took 0.6s. It was concluded that AE-ELM can be applied for real-time fault diagnosis owing to its faster response and higher accuracy. Chen et al. [60] have used deep SAE with noise added vibration data for rolling bearing fault severity level classification and life stage prediction. The input data along with added noise were fed to the deep SAE, constraining the overfitting problem caused by limited training data. They have used the two different vibration datasets for classifying different bearing severity levels and life stages. The investigation was also carried out with and without noise data. The comparative analysis between the deep SAE with and without noise revealed that the model with noisy data effectively overcame the overfitting problem and achieved 98.3% accuracy, whereas deep SAE without noisy data achieved 93.7% accuracy. The authors in [61] employed stacked SAE with hybrid features for feature extraction and DNN for classification of bearing fault severity level. This approach allowed extracting more discriminative information, which in turn raised the classification accuracy. In addition, the input to the DAE was a combination of time domain features, wavelet energy features, and power spectrum features extracted from the vibration data. This hybrid feature pool was used to overcome nonlinearity in the data. The results of the investigation revealed that the proposed method outperformed the vibration spectral-imaging-based DNN method with an accuracy of 99.1%. Lu et al. [62] have used a stacked denoising autoencoder (SDA) for rolling bearing fault classification using raw vibration data. They have compared the performance of SDA with SAE by inputting the data with varying signal-to-noise ratio (SNR) values. From the comparative investigation results, it was observed that the SDA outperformed the SAE, SVM, RF, and AE with maximum accuracy of 99.8% at the SNR of 20 dB owing to its learning capacity for complex and nonlinear mapping relationships. Sun et al. [63] have employed SAE and DNN for rolling bearing fault classification. They have input vibration data to the model after compressing it. Data compression was used for handling the significant amount of data more efficiently. The investigation of the method confirmed the higher classification accuracy of 97.4% with data compression compared to the DNN model without data compression, which achieved 96.7% accuracy.

Shao et al. [64] have used ensemble deep AEs for the classification of various roller bearing faults. They used 15 standard AEs, each with different activation functions. This approach overcame the manual feature extraction and limitations of the individual model. These AEs were used for extracting features from vibration data. The diagnosis results revealed that the proposed method performed more robustly and produced more effective classification results than the individual AE models with different activation functions.

### Table 1: Application of MLP in condition monitoring of motors.

| Year | Input data and features | DL model | Application | Remarks |
|------|-------------------------|----------|-------------|---------|
| 2016 | Current signature       | MLP [47] | Induction motor fault classification | Operating frequency of the motor did not affect the accuracy of the classifier. However, increase in the accuracy was observed as severity of the fault increased |
| 2017 | Current signature and voltage signal amplitudes | MLP [48] | Induction motor fault classification | The method yielded better classification results and it was computationally efficient compared to the traditional methods |
|      | Current signature + MI  | MLP [49] | Induction motor fault classification | MI feature extraction method allowed MLP to produce promising results compared to RBF |
| 2018 | Current signature + FFT + WPT + statistical parameters | MLP [50] | Rotor broken bar fault classification | Superior feature extraction allowed to detect the fault severity level even with no-load condition |
|      | Current signature + CCA | MLP [51] | Induction motor fault classification | The method allowed to correctly classify the faults with higher accuracy owing to reduced dimensions of the data by CCA |
| 2019 | Current signature + MI  | MLP [52] | Bearing fault classification | MI feature extraction method allowed to classify the bearing faults more effectively compared to conventional methods |

### Figure 3: Deep AE structure.
DBN, and CNN, with an accuracy of 97.1%. In [65], authors have used an unsupervised deep AE to predict the bearing decay state in an inverter-fed induction motor. They input stator current data as a compact representation of the bearing state. The acquired data of the artificially introduced bearing faults with varying loads was segmented using a sliding window of 24s. Subsequently, they have extracted time and frequency domain features and then fed them to the models. The investigation results demonstrated that the deep AE produced clearer clusters of different bearing conditions with higher classification accuracy than a shallow neural network (SNN).

In [66], the authors have used ensemble stacked autoencoders (ESAE) for bearing fault classification. They have applied FFT to raw vibration data and fed to the model. The proposed method was compared with time-domain features’ extraction methods. The comparative results confirmed the superior performance of the ESAE with FFT technique compared to the ESAE with other feature extraction techniques such as average, kurtosis, and variance with a minimum mean absolute error (MAE) of 0.0672. Skyvik et al. [67] have employed deep AE to classify different faults in induction motors such as interturn faults, bearing faults, and broken rotor bar faults. For preprocessing, they employed Welch’s method to estimate spectral density. Subsequently, they have applied FFT to current data and then fed that to the model. The performance comparison with SVM and KNN models confirmed that the proposed scheme performed much better than the traditional schemes with an accuracy of 96.1%. Zhao et al. [68] have constructed an optimal hybrid DL model, which consists of SAE and GRU, to classify the rolling bearing faults more accurately. In addition, they have used the grey-wolf optimising algorithm to enhance the performance of the model. The proposed model was capable of extracting features from raw vibration data. The results confirmed that the model achieved robust and effective results with an accuracy of 97.1%, which is higher than other traditional models, including ANN, SVM, SAE, and gated recurrent units (GRU).

Zhu et al. [69] have presented a novel DL model called stacked pruning denoising autoencoder (SPDAE) for rolling bearing fault diagnosis. The model reduces information loss by introducing new channels to interconnect the layers. The pruning operation was carried out on nonsuperior units in the model to reduce the number of training parameters, which in turn enhanced training efficiency and precision. To ensure the uniqueness of the dimensions extracted from SPDAE, the features from the same layer were fused. Raw vibration data with noise was fed to the model for the bearing fault classification. Comparative results confirmed the superiority of the proposed model compared with the conventional models, such as ANN, SAE, and DAE. The SPDAE model achieved 100% accuracy on the bearing dataset. The authors have suggested future studies investigating noise addition, pruning operations, and feature fusion operations, which may improve the performance of the model. State-of-the-art research related to condition monitoring of motors using AEs is summarized in Table 2.

2.3. Restricted Boltzmann Machines (RBM). The RBM are Markov-based special-type models, consisting of two layer neural networks [70]. They employ an unsupervised training method based on a greedy layer-wise process. They have the capability to learn missing data patterns. RBM-based networks pose difficulty in the training process and also in tracking the loss function. There are two types of RBM-based DL models, which are Deep Boltzmann Machine (DBM) and Deep Belief Network (DBN). DBN is a semidirected model, while DBM is a completely undirected graph model, as shown in Figure 4(b). The RBM-based DL models are reported as follows:

2.3.1. Deep Boltzmann Machine (DBM). The DBM can be considered as a stacked RBM, which comprises multiple hidden and visible layers rather than only RBM layers as shown in Figure 4(a), where the blue dotted line shows the separation of the layers [71]. Each layer is composed of symmetrically coupled stochastic units. DBM has the capacity to learn features from complex data and capture higher-order correlations among the hidden features. The DBM model is trained as a combined model, in comparison to the other type of RBM, a deep belief network (DBN) (Section 2.3.2), which is trained layer by layer [72]. Hence, the training process of the DBM requires more computational power than for DBN models.

2.3.2. Deep Belief Networks (DBN). Like DBM, DBN is formed by stacking RBM layers in such a way that output of the n-th layer becomes input to the (n+1)-th layer. Figure 4(b) shows the structure of the DBN. It is a mixed directed and undirected graphical model. It can process large amount of nonlinear data [73]. As an RBM, it is also trained in a greedy layer-wise unsupervised fashion [74]. Fine-tuning is required after the pretraining process, which is performed either on training data label or on a proxy for DBN log-likelihood. Targets are achieved by adding a softmax layer on top of the DBN architecture.

Both the DBN and DBM have been used in various condition monitoring systems for motors. In [75], the authors have employed DBM for roller bearing fault classification. Three types of features (time-domain features, frequency-domain features, hybrid features) were extracted from vibration data and fed to the model. Among these features, hybrid features yielded the better classification performance with an accuracy of 99.5%. Tao et al. [76] have used DBN and a multisensor information fusion technique for bearing fault detection. They have inputted time-domain features from multiple vibration sensors to the model. It was observed in the results that DBN not only adaptively fused the multisensory data but also increased the accuracy up to 97.5%, that is, 10% greater than that of the model with single sensor data. The DBM model was able to achieve highest accuracy among the comparative models, including SVM, KNN, and ANN. In [77], authors have classified bearing degradation states using DBN and the Weibull distribution. Bearing degradation states were classified based on fitted RMS of the vibration data by Weibull distribution, which
was employed to avoid fluctuation in statistical parameters. The results confirmed the effectiveness of the technique through run-to-failure experiment. The model effectively classified the bearing degradation states with an accuracy of 89.9%.

Shao et al. [21] have used a DBN model for the fault classification of induction motors in manufacturing. Different classes of the motor faults included normal motor, unbalanced rotor, stator winding defect, defective bearing, bowed rotor, and broken bar. They utilised frequency domain features extracted from vibration data investigated the effect of the depth of the model on the classification accuracy. The investigation results confirmed the effectiveness of the method for automatic fault classification in manufacturing. In [78], the authors have investigated an improved version of the DBM by varying its energy function. The proposed method addressed the problem of the DBM, which can only process the binary data; by replacing the binary visible units with the Gaussian units, the variation allowed the DBM to process real-value data. This model was employed for bearing fault classification using raw vibration data. Supervised training was performed followed by greedy unsupervised training to initialize the model parameters. Results confirmed the effectiveness of the method in fault classification using real-value data with manual feature extraction with a classification error of 1.9%.

| Year | Input data and features | DL model | Application | Remarks |
|------|-------------------------|----------|-------------|---------|
| 2016 | Raw vibration           | SAE + DNN [57] | Motor fault classification | Integrated "dropout" regularization allowed to avoid overfitting, which in turn improved the classification performance |
|      | Acoustic emissions/sound signals + STFT | SAE [58] | Rolling bearing fault classification | STFT allowed for fast and effective feature extraction, which in turn increased accuracy of the classifier |
|      | Vibration + FFT         | AE + ELM [59] | Bearing fault classification | AE-ELM approach increased the speed of the training process. The approach was much faster than SAE. However, the model did not perform well for online sequential learning |
|      | Raw vibration           | SAE [60] | Rolling bearing fault severity level classification | Deep SAE inputted with added noise vibration data has effectively overcome the overfitting problem posed by small training datasets |
|      | Vibration + time domain features + wavelet energy features + power spectrum features | SAE + DNN [61] | Rolling bearing fault severity level classification | Hybrid features pool was able to tackle the nonlinearity in the vibration data and resulted in more effective classification results |
|      | Raw vibration           | DAE and SAE [62] | Rolling bearing fault classification | The comparative analysis performed between DAE and SAE classifiers showed that the DAE outperformed SAE classifier due to its potentiality of learning complex nonlinear mapping relationships |
|      | Vibration + compression | SAE + DNN [63] | Rolling bearing fault classification | Data compression technique with SAE enables effective fault classification, with huge datasets through an easy approach |
| 2017 | Raw vibration           | Ensemble AE [64] | Rolling bearing fault classification | A novel ensemble of 15 AEs, each with different activation function overcome the limitations of individual AE models and removed manual feature extraction |
|      | Current + time domain and frequency domain features | Deep AE [65] | Unsupervised bearing fault prediction | Results showed effectiveness of the technique in terms bearing fault prediction with clear clustering and high accuracy |
| 2018 | Vibration + FFT         | ESAE [66] | Rolling bearing fault classification | FFT with ESAE has revealed superior performance of the proposed model in comparison with traditional models |
|      | Current + FFT           | AE [67] | Induction motor fault classification | FFT with the AE yielded superior performance compared to the existing models |
|      | Raw vibration           | SAE + GRU [68] | Rolling bearing fault classification | An optimal hybrid DL mode constructed using SAE and GRU can extract rich features from raw vibration data. Results confirmed the superiority of the proposed model |
| 2019 | Raw vibration           | Stacked pruning DAE [69] | Rolling bearing fault classification | A novel model called SPADE with pruning operation increased efficiency and precision of the model by decreasing the training amount of the model |
| 2020 | Raw vibration           | Stacked pruning DAE [69] | Rolling bearing fault classification |
In [79], researchers have presented a method to classify rolling bearing faults using DBM, principal component analysis (PCA), and a least square support vector machine (LS-SVM). DBM was used for feature extraction, PCA was used to reduce the dimensionality of the data, and the LS-SVM was applied for classifying the bearing faults. Acoustic emission signals were used as input to the model owing to its higher sensitivity compared to vibration. The combination of DBM and PCA identified better features, which in turn increased classification accuracy of the model. Experimental results confirmed the effectiveness of the model with an accuracy of 95.4%. Zhao et al. [80] presented a variation mode decomposition- (VMD-) and Hilbert transform- (HT-) based DBN (VHDBN) for rolling bearing fault classification. Bearing vibration signals are decomposed into intrinsic mode functions (IMFs) through VMD. Subsequently, HT extracted the instantaneous frequency and amplitude of the IMFs and constructed a feature matrix that was fed to the DBN model. A combination of VMD and HT allowed the extraction of improved features and achieved diagnosis accuracy of 98%. The investigation results confirmed that the VHDBN algorithm has great advantage over conventional algorithms such as SVM and the DBN with time domain signals.

In [81], the authors have presented a novel architecture-multiscale cascaded DBN (MCDBN) for automatic fault classification in motors. This variation to DBN added parallel learning capability by introducing a multiscale coarse-grained method, which in turn improved the feature extraction performance. The vibration signal was split into subsignals of equal window size using sliding window with data overlap. The technique allowed DBN to learn features from the data at multiscale rather than inherent information. Subsequently, coarse-grained time-series data at different scales was obtained through the coarse-grained process. Experimental results confirmed superiority of the method compared to standard DBM with an accuracy of 99.8%. The authors have suggested that the investigation has provided a promising tool for condition monitoring of industrial motors. Yu et al. [82] have used a two-stage approach by combining DBN and the Dempster–Shafer (D-S) theory for bearing faults and their severity level classification. D-S theory was employed for fusing vibration data from multiple sensors (horizontal and vertical vibration). A genetic algorithm (GA) and a particle swarm optimization- (PSO-) based hybrid algorithm was used to optimize the parameters of DBN during the training. The results revealed that the fused hybrid GA-PSO algorithm has not only upgraded capabilities of the DBN classifier but also enhanced computational efficiency with an overall accuracy of 99.6%. Authors have also used wavelet package decomposition (WPD) for extracting energy features from the vibration data of bearings. Table 3 summarizes applications of DBN and DBM in condition monitoring of motors.

2.4. Convolutional Neural Networks (CNN). A neural network layer that employs a convolution operation includes convolution layers in its structure. Figure 5 illustrates the structure of a CNN. It consists of two blocks; the first block comprises convolution layers and pooling layers that extract features from the data. Multiple stacks of convolutional layer
and pooling layer are employed to extract rich features from data. The second block comprises fully connected layers that predict target variables through learning representations from training data.

The discrete convolution networks are employed to extract representation from 1D or 2D data through learnable filters. The convolutional operation results in output $C$ as given in equation (1), and the output of the convolutional layer is known as the feature map.

$$C(i, j) = \sum_{m} \sum_{n} I(m,n)K(i - m, j - n),$$  \hspace{1cm} (1)

where $I$ and $K$ are the input and filter, respectively. The pooling layer reduces the size of the feature map and assists in avoiding overfitting. Following the pooling operation, the feature map is flattened (1D array) and fed to the classification block [83]. The classification block generates output based on the extracted features.
CNN and its derivatives have been widely used in condition monitoring systems. In [84], the authors have used CNN with DFT for classifying motor faults. The method was used to classify motor faults such as bearing faults and lubricant degradation levels. Compared to traditional methods that rely on manual feature extraction, this method allowed the automatic extraction of features from scaled vibration data. Then, nonoverlapping windows are separated, each containing measurement samples of one second.

Subsequently, the DFT of the windows are calculated. At the end, these frequency decompositions are fed to the deep CNN model. Overall, this method yielded better performance than classical feature engineering techniques such as kurtosis, skewness, mean, and standard deviation. Liu et al. [85] have proposed a novel method called dislocated time-series CNN (DTS-CNN) for fault classification of electric motors. The dislocation layer in the model can extract the relationship between periodic vibration signals with different intervals. This modification of CNN performs more effectively under nonstationary conditions. This method extracts features from the raw data. The model mitigates the overfitting problem through weight sharing and sparse connectivity. The model was used to predict nine different motor faults, and it was confirmed through the results that this model performed better than the standard CNN with an accuracy of 96.3%.

In [86], authors have employed a 1D-CNN for real-time classification of the bearing faults. The method did not require any feature extraction technique, which in turn made it fast and computationally efficient. Motor current signals were fed to the model under constant speed condition. Experimental results confirmed the effectiveness of the method compared to the conventional models including MLP, SVM, and RBF. Guo et al. [87] have investigated a novel adaptive deep CNN (ADCNN) for bearing fault classification and their severity levels. It avoided training process failure due to unsuitable learning rate by the addition of adaptive learning rate and momentum. It also enabled the automatic extraction of features from raw vibration signals. The investigation results confirmed the superiority of the method compared to existing methods such as support vector regression machine (SVRM).

Ding et al. [88] have proposed a novel method called energy-fluctuated multiscale feature (EFMF) learning with deep CNN for the spindle bearing fault classification. A multiscale deep CNN was constructed using different layers such as convolution and pooling layers with sigmoid function. It combined the skipping layer with the last convolution layer, which provided input to the multiscale layer. Meanwhile, wavelet packet energy images (WPI) were constructed using wavelet packet transform (WPT) and phase space reconstruction. This 2D image was able to reveal energy fluctuations of the vibration signals and reconstruct local relationships among the WP nodes. Furthermore, representations were learned by the deep CNN architecture through brightness (frequency energy) variations of the energy-fluctuated images. Taking advantage of local and global features, the model was able to effectively classify the ten different spindle-bearing faults with a maximum accuracy of 98.8%. The model showed outstanding performance compared to the traditional methods, including standard CNN, BPNN, PCA, and LDA. In [89], authors have used an ensemble deep CNN model with the improved Dempster–Shafer (D-S) evidence theory (IDSCNN) for bearing fault classification under varying load conditions. The improved D-S evidence theory was employed for data fusion from the two vibration sensors. It was implemented using a similarity matrix and a modified Gini index. It addressed the two problems of traditional D-S evidence theory: objective evaluation of the basic function of the evidence body, and solving conflicting evidence from different sensors. Deep CNN was fed with the RMS value from FFT of the vibration signals. The experimental results confirmed higher performance of the model compared to the other existing models such as MLP, SVM, deep convolutional neural networks with wide first-layer kernels (WDCNN), and DSCNN.

Zhang et al. [90] have employed a novel method called CNN with training interference (TICNN), which can detect bearing faults with noisy data and under varying load conditions. The authors have used two techniques to introduce antinoise and domain adaptation abilities, which allowed the model to extract features from raw vibration signals. This was done by adding “dropout” layers and very small batch training. In addition, they have initially used wide convolution kernels for suppressing the noise, which is followed by small convolutional kernels that extracted rich representations from the data. It was confirmed through the results that the model performed stable and accurate classification due to ensemble learning with an average accuracy of 95.5%, whereas the comparative methods of SVM, MLP, and DNN only achieved poor performance with average accuracy of 65%, 80%, and 80%, respectively. The authors have suggested that this model would be useful in industrial environments. In [91], the authors have employed CNN as machinery health indicator. The method automatically learned features from the vibration data and constructed health indicators (HIs). HIs were constructed using a specific degradation process. However, the HIs faced a problem of outlier region deviation that was referred to as the trend burr, which negatively affects the performance of the HIs. The method was effectively able to detect and correct the outlier regions. The investigation results confirmed the effectiveness of the method in feature extraction.

Jia et al. [92] have employed a model called deep normalized CNN (DNCNN) to classify the bearing faults using the vibration data. The study addressed two problems of CNN. First the imbalanced fault classification problem was addressed using normalized CNN with weighted softmax, which allowed learning of better features and avoided misclassification. Another problem is that it was not clear what the CNN had actually learned as DL methods are treated as black box models. This problem was addressed using a neuron activation maximization (NAM) technique, which suppressed the useless information by analysing the kernels of the normalized CNN. The model produced effective results by overcoming these two problems and had a maximum accuracy of 99.2%.
Hoang et al. [93] have employed vibration image CNN (VI-CNN) to classify the rolling bearing faults. Authors have converted the vibration signals into gray-scale images by normalizing into range of $[-1, 1]$. Consequently, the amplitude of each sample becomes the intensity of pixels of the vibration image. It was confirmed that the model effectively classified the faults even in varying load conditions with an accuracy of 97.7%. The authors have suggested that the method can robustly perform in industrial environments. Li et al. [94] have employed a novel method by combining CNN and S-transforms to classify the bearing faults. The S-transform was combined with a CNN as an S-layer, which allowed extraction of features from the two accelerometers data. The S-layer automatically converted the vibration data into a 2D time-frequency matrix. The investigation results confirmed superior performance compared to the existing methods, namely, SVM, KNN, linear discriminant (LD), and bagged trees (BT).

In [95], the authors have used CNN based on a capsule network (ICN) for bearing fault classification. The method was intended for strong generalization and used a dynamic routing capsule net with an inception block. The Inception block removed the nonlinearity of the capsule. Before applying these steps, vibration data were converted to time-frequency graph using STFT and fed to the model. Subsequently, the model classified the faults through varying lengths of the capsule. The investigation results confirmed the higher generalization power of the model with an accuracy of 82%, compared to the standard CNN. Hoang et al. [96] have employed CNN with an information fusion technique to classify bearing faults of a motor. They have employed two phase current signatures, which further split into equal samples using a sliding window. Then, 1D signals are converted into 2D matrix by rearranging the array signal. The transformed 2D current signal is effectively a gray-scale image and this is fed into the model. Experimental results confirmed the effectiveness of the technique in detecting the faults with acceptable performance with a maximum accuracy of 99.4%.

Li et al. [97] have used CNN with WPT for rolling bearing fault classification. WPT extracted features from raw vibration signals, which were further transformed again into gray-scale images. The investigation results showed that the model achieved superior fault detection with individual fault detection accuracy of up to 100% because of the richness of the input features. In [98], the authors have used enhanced LeNet-5 CNN for bearing fault classification. STFT was employed to convert vibration data into 2D images, and hierarchical regularization was used to speed up the training process. In addition, the scaled exponential linear unit (SELU) function was employed to avoid excessive dead-nodes during the training process, which in turn enhanced the robustness of the model learning. It was observed through results that the model accelerated the training process by selecting sensitive features even under varying load conditions. Moreover, the model effectively classified the faults with higher accuracy up to 100% compared to the models with individual time or frequency features. In [99], authors have used CNN to diagnose the stator winding faults of induction motors. They have normalized the raw current data and then converted it into a three-dimensional matrix. The method was able to detect the number of faulty stator-winding shorted turns without extensive preprocessing. The investigation results confirmed the applicability of the method in real-time, regardless of motor operating conditions. Furthermore, results showed a high accuracy of up to 100% in the individual fault detection. Research studies on motor diagnosis and prognosis using CNN are summarized in Table 4.

2.5. Recurrent Neural Networks (RNN). RNNs yield better performance with sequential or time-series data, which makes them the most suitable candidate for condition monitoring of motors. According to ref. [100], RNNs are the deepest among all neural networks, and their architecture is as shown in Figure 6. They differ from MLP, which only maps input data to target vectors, while RNNs map the entire history of past inputs to target vectors owing to their memory capability. For supervised tasks, RNNs can be trained by employing back-propagation through time. The transition function of a basic RNN can be defined as in

$$ h_t = \mathcal{H}(x_t, h_{t-1}), $$

where $\mathcal{H}$ is a nonlinear and differentiable transfer function, $x_t$ is current time information, and $h_{t-1}$ is the previous time information.

Although RNNs perform better on time-series data, they face the gradient vanishing problem. This problem was addressed through the advent of long short-term memory (LSTM) in 1997 [101], which has outperformed in various fields. LSTMs have the capability to memorize and forget representations of data. Moreover, gated recurrent units (GRU) and bidirectional LSTMs can enhance model flexibility and capacity. As shown in Figure 7, deep LSTM can increase the representation learning capabilities through propagating the output of one layer as an input to the next layer.

RNN and its derivatives have been extensively used in motor condition monitoring owing to its memorizing power and robust performance. Zhao et al. [102] have employed deep LSTM with time-series data to classify motor faults. The “dropout” layers were used for model regularization. Raw vibration data were used as input to the model. Comparative results confirmed the effectiveness of the LSTM model with minimum root mean error (RMSE) of 10.2% compared to conventional techniques such as MLP and basic RNN. In [103], authors have employed local feature-based GRU (LFGRU) for motor fault classification again using raw vibration data. Local features were extracted from synchronized windows of the multisensory data, and then fed to the model. Experimental results confirmed the robustness of the model in classifying the motor faults with the maximum accuracy of 99.6%. In [104], authors have used RNN for rolling bearing fault prediction. They have investigated the model with two types of features: time domain and frequency domain. FFT was used to convert vibration data into frequency domain features. Experimental results
Table 4: Application of CNN in condition monitoring of motors.

| Year | Input data and features | DL model | Application | Remarks |
|------|-------------------------|----------|-------------|---------|
| 2016 | Raw vibration           | CNN + DFT [84] | Motor fault classification | Owing to the automatic feature extraction, the method produced promising results compared to classical feature engineering methods |
|      | Raw vibration           | DTS-CNN [85] | Motor fault classification | Dislocation layer yielded better performance than standard CNN owing to rich feature extraction |
|      | Raw current             | 1D-CNN [86] | Bearing fault classification | The simple architecture of the model allowed real-time detection of the faults |
|      | Raw vibration           | ADCNN [87] | Bearing fault classification and their severity level classification | Adaptive architecture of the CNN allowed to learn rich features from the data, which in turn increased its performance |
| 2017 | Vibration + WPI         | Multiscale deep CNN [88] | Spindle bearing fault classification | Novel feature learning techniques with multiscale deep CNN allowed the model to outperform other traditional models through learning local and global features simultaneously |
|      | Raw vibration data from two sensors | IDSCNN [89] | Bearing fault classification | Improved data fusion technique produced better classification results by addressing the problems in the conventional feature extraction techniques |
| 2018 | Raw vibration           | TICCN [90] | Bearing fault classification | Dropout layers with the wide and small convolutional layers suppressed the noise of data. Meanwhile, ensemble model allowed to attain high classification accuracy |
|      | Raw vibration           | CNN [91] | Motor health indication | The method was able to overcome outlier regions and effectively learned features from the data |
|      | Raw vibration           | DNCCN [92] | Bearing fault classification | The model produced effective classification results by overcoming data imbalance problem |
| 2019 | Raw vibration           | VI-CNN [93] | Rolling bearing fault classification | 2D vibration image allowed the method to effectively classify the faults without any feature learning and denoising technique |
|      | Raw vibration           | ST-CNN [94] | Bearing fault classification | The method produced higher performance than existing methods owing to automatic feature extraction by S-layer |
|      | Vibration + STFT        | ICN [95] | Bearing fault classification | Inception block resulted in better generalization than CNN |
|      | Raw current signatures  | CNN + IF [96] | Bearing fault classification | Promising results were obtained owing to information fusion |
| 2020 | Raw vibration + WPT     | CNN [97] | Rolling bearing fault classification | Classification performance improved owing to the gray-scale vibration images |
|      | Vibration + STFT        | CNN + SELU function [98] | Rolling bearing fault classification | The method effectively classified the fault owing to the 2D-images of the data and regularization |
|      | Raw current signatures  | CNN [99] | Stator winding fault detection | The method can effectively detect stator winding faults from raw current data without any preprocessing |

Figure 6: RNN structure.
demonstrated that the model predicts the conditions of bearing better with frequency domain features.

Xiao et al. [105] have used deep LSTM to classify various motor faults using vibration data. The authors have not used any feature learning technique for classifying motor faults such as broken-bar, bowed-bar, bowed-rotor, faulty bearing, and voltage imbalance using raw vibration data. Results confirmed the superiority of the model performance with an accuracy of 98.6% compared to the conventional models including SVM, MLP, and standard RNN. In [106], the authors have predicted bearing performance degradation using bottleneck features based on LSTM. A wavelet threshold denoising (WTD) technique eliminated noise from the vibration data. Then, statistical features were extracted and finally fed to the stacked autoencoder (SAE) to obtain bottleneck features. The bottleneck features take depth and nonlinearity of the signals into account. Experimental results showed the effectiveness and superiority of the method in predicting the bearing’s degradation with a resulting minimum RMSE of 0.0891 which compares well to existing methods including SVM and MLP.

Xiao et al. [107] have employed LSTM with weighted batch normalization (BN) for detecting different faults in the induction motor. They have employed two manual feature extractions, namely, empirical statistical parameters and recurrent quantification analysis (RQA) to add antinoise capability in the model. The weighted BN allowed the evaluation of the contribution of each feature learning technique, and validated noise reduction. The results confirmed the robustness and effectiveness of the model in fault classification of the induction motor with an accuracy of 99.3% compared to the other DL models such as SVM, MLP, CNN, and standard LSTM. In [108], authors have classified bearing faults using a hierarchical stacked LSTM. The input layer received the vibration data, and then two stacked LSTMs were able to effectively learn representations from the data without any preprocessing. The experimental results confirmed that the model produced promising results owing to its deep structure with an accuracy of 99% and the model outperformed the state-of-the-art models such as SVM, MLP, 1-layer LSTM, and CNN.

Zhang et al. [109] have employed LSTM to assess the bearing performance degradation and predict the RUL using vibration and temperature data. The authors have used waveform entropy (WE) to identify the running condition of the bearing by computing the local mean of logarithmic vibration energy. In addition, a particle swarm optimization (PSO) method was used to optimize parameters of the LSTM, which in turn improved its learning performance. The authors have divided degradation states into different stages by time. WE has performed effectively under various degradation states while it lagged in some conditions owing to window length. However, it did not reflect any negative impact on the performance of the model. Experimental results showed the effectiveness of the method in indicating the degree of degradation in assessing the degradation states with an accuracy of 93.1%. In [110], the authors have used deep gated recurrent units (DGRU) with ELM to classify the faults of an adaptive rolling bearing. In addition, the authors have used an artificial fish swarm algorithm (AFSA) for GRU parameter optimization. The model consists of two stacked layers of GRU, which learned features from the raw vibration data. Lastly, ELM is applied for accurate classification of the faults based on learned features. The model was able to achieve 94.5% testing accuracy and the results demonstrated robust performance of the model compared to the conventional models such as CNN, DBN, and SAE.

Enshaei et al. [111] have used bidirectional deep LSTM (BiD-LSTM) to classify bearing faults. BiD-LSTM takes sequential data into account in both the forward and backward directions. It was confirmed through experimental results that the deeper BiD-LSTM performed better than the single-layered network and achieved 100% testing accuracy. The model effectively classified faults with high accuracy with raw vibration data. In [112], the authors have employed a hierarchical GRU network (HGRUN) for predicting future health index (HI) and RUL of the bearing. The kernel principle component analysis (KPCA) and exponentially weighted moving average (EWMA) were used to design a modified HI. Firstly, statistical features are extracted (time domain, frequency domain, and hybrid domain), then the KPCA fused these features and transformed it as a unified

![Deep LSTM classifier structure](image-url)
HI. Subsequently, EWMA further modified the HI, which can depict the bearing degradation process. Lastly, HGRUN was developed by stacking multiple GRU layers and inputting with the modified HI. Experimental results confirmed that the method effectively depicted the degradation process of the bearing and can predict the future HI and RUL of the bearing. Comparison showed the superiority of the technique to the existing techniques with the minimum of $13.8 \pm 2.8\%$. In [113], the authors have used the enhanced deep GRU and complex wavelet packet energy moment entropy for early bearing fault classification. Complex wavelet packet energy moment entropy as a monitoring index allows reduction in aliasing and the detection of dynamic changes in the vibration data. Subsequently, deep GRU allows to learn the complex mapping relationships from the vibration data. Lastly, the learning capability of the model was enhanced using a modified training algorithm based on learning rate decay strategy. Experimental results confirmed the effectiveness of the method in early fault detection compared to the other prognosis methods. These recent studies related to motor condition monitoring using RNN and its variants are summarized in Table 5.

2.6. Generative Adversarial Networks (GAN). A Generative Adversarial Network (GAN) is a binomial zero-sum game-theory-based learning model. It comprises two models: a generator model (G) and a discriminator model (D). The structure of a GAN is shown in Figure 8(a). Both the models can have a different type of neural network architecture such as RNN, AE, or CNN. The D-model tries to increase the probability of collected true data (x) and decrease the probability of samples generated by the G-model. The G-model tries to cheat the D-model by generating a sample training set using a noise input (z), gradually improving its performance until the D-model can no longer discriminate between the true data and the generator data. Thus, by employing an antagonism training process, the capacity of both the models is improved simultaneously [114, 115]. The optimization of this two-player game is calculated as given in

$$\min_G \max_D V(D, G) = E_{x \sim P_{data}(x)} [\log D(x)] + E_{z \sim P_z(z)} [\log (1 - D(G(z))].$$

(3)

Considering a supervised learning approach, the GAN models can generate fake labels which are like real data. Figure 8(b) gives a visual illustration of a GAN-based classifier. The classifier receives samples from G-model and the classification error back-propagates through G-model and classifier.

In recent years, researchers have employed GANs and its derivatives in motor condition monitoring. It is often employed to address the data imbalance problem through data augmentation. Shao et al. [116] have developed an auxiliary classifier GAN (ACGAN)-based framework to learn and generate realistic one-dimensional vibration data. Both the generator and discriminator consisted of 1D-CNN, which was allowed to learn representations and generate high-quality artificial data samples. In addition, batch normalization was employed to address the gradient vanishing problem during the training process of the GAN, which in turn assisted in avoiding overfitting. They have employed statistical methods to evaluate the quality of the generated data. Experimental results demonstrated the effectiveness of the model in data augmentation. The model robustly classified the faults by addressing the data imbalance problem with an accuracy of 99.1%. In [117], the authors have employed deep GAN for bearing fault diagnosis using an imbalanced dataset. The authors have used a two-stage approach: the first stage augmented the data through the GAN model; then the second stage classified the faults using a deep CNN model. They have introduced multiple generators in which each generator is dedicated to the specific bearing conditions. Investigation results confirmed the robustness of the approach in data augmentation and fault classification with a maximum accuracy of 99.9%. Authors have verified the approach by applying it on two different datasets. Applications of GANs in motor condition monitoring are summarized in Table 6.

3. Summary, Challenges, and Prospects of DL Models

The effective application of DL models in condition monitoring systems extensively rely on data acquisition, data labeling, feature processing, and model parameter optimization. However, these processes are challenging, time-consuming, and may require domain expert knowledge. In the previous section, we have comprehensively reviewed each DL model and their application in condition monitoring of motors. Although, the reviewed research demonstrated promising leads in condition monitoring of motors, there are various open challenges that are yet to be completely solved. Figure 9 shows a heat map of DL models that have been used along with a variety of feature-processing techniques. From the heat map, it is apparent that the DL models have been effectively performing on raw input data owing to their deep hierarchical architectures. However, time domain features also have been used in various researches with less complex models that show the importance of feature-processing techniques in simplifying the training and testing process of DL Models. On the other hand, Figure 10 shows a 3D map of the number of publications using the type of input data with different DL models for motor fault diagnosis. The majority of the current literature has focused on using vibrational analysis for motor condition monitoring tasks.

The following section presents some challenges in the application of DL models and future directions to improve the performance of these models. It also summarizes the strengths and drawbacks of these models:

3.1. Class Data Imbalance. During the data extraction process, most of the time healthy samples of the data outnumber the ones representing fault conditions. Hence, while
Table 5: Application of RNNs and its variants in condition monitoring of motors.

| Year | Input data and features | DL model | Application | Remarks |
|------|-------------------------|----------|-------------|---------|
| 2016 | Raw vibration           | LSTM [102] | Motor fault detection | The model outperformed the existing models owing to its deep architecture |
|      | Raw vibration from multiple sensors | LFGRU [103] | Motor fault classification | The model was able to effectively classify motor faults owing to deep structure without any extensive feature extraction |
|      | Vibration + time domain features + frequency domain features | RNN [104] | Rolling bearing fault prediction | Model showed better prediction results with frequency domain features |
| 2017 | Raw vibration           | LSTM [105] | Motor fault classification | The results confirmed effectiveness of the technique compared to existing methods owing to gated functions of the model |
|      | Vibration + WTD + time domain features | SAE + LSTM [106] | Prediction of bearing performance degradation | Results confirmed effectiveness of the method in achieving the task compared to the conventional methods |
|      | Vibration + hybrid feature + RQA | LSTM [107] | Induction motor fault classification | Method effectively classified the motor faults due to the antinoise feature extraction techniques |
|      | Raw vibration           | LSTM [108] | Rolling bearing fault classification | The results demonstrated efficiency in the bearing fault classification owing to inherent mechanism and deep architecture |
|      | Vibration + temperature + WE | LSTM + PSO [109] | Bearing degradation assessment and RUL prediction | Results confirmed effectiveness of the method in estimating the bearing life states and degree of degradation |
| 2018 | Raw vibration           | GRU + AFSA + ELM [110] | Adaptive rolling bearing fault classification | Results revealed robust performance of the model owing to optimized parameter selection by AFSA and accurate classification by ELM |
|      | Raw vibration           | BiD-LSTM [111] | Bearing fault classification | Results demonstrated effectiveness of the method owing to its long-term dependency in time series data KPCA and EWMA allowed to effectively track the degradation process and promisingly predicted the HI and RUL of the bearing |
|      | Vibration + time, frequency, hybrid features | HRGUN + KPCA + EWMA [112] | Future HI prediction and RUL prediction of the bearing | |
| 2020 | Raw vibration + complex wavelet packet energy moment entropy | Enhanced deep GRU [113] | Early bearing fault classification | The monitoring health index technique and the modified training algorithm allowed to detect fault at an early stage |

Figure 8: (a) GAN structure. (b) GAN-based classifier structure.
designing a classification model for anomaly detection or fault identification, there is a clear bias toward classes that have more samples of data. There are multiple ways to deal with this bias in classification models; examples include undersampling which is reducing the samples of over-sampled classes, and oversampling which is increasing samples of minority classes by synthesizing new examples. A prebiased model can also be used with weights, which is a clever way to deal with imbalanced data, where individual class weights are calculated, and higher weights would lower their effect on model prediction. Most DL models have performed effectively in different applications but these models have some drawbacks which require more investigation for producing optimal results. The generative models such as AEs and GANs, although harder to train, can provide a way to synthesize authentic data.

3.2. Feature Processing. Feature processing reducing the number of features greatly reduces the hardware dependence and the need for a highly nonlinear function mapping by DL models. However, the bulk of the research is conducted on raw input data with DL models as highlighted in Figure 9, which truly exploits the potential of these models. However, performance evaluation of these models with current evaluation matrices is vague. Either the trade-off between raw and formulated features versus explainability of the models needs to be taken in account or new evaluation matrices need to be developed.

3.3. Model Selection. Model selection for addressing domain-specific problems heavily depends upon the way the problem is formulated. Anomaly detection problems are mostly formulated as classification problems. Another factor that plays a role in selecting a proper model for the problem is the type of data available. The right choice for the model also depends upon the way data is formulated.

3.4. Hyperparameter Tuning. This is one of the most important parts of employing a DL algorithm and includes model tuning in terms of (a) number of neurons per layer, (b) number of layers, and (c) choices regarding initialization, activations, optimizers, learning rate, loss calculation, etc. Although most of the time this is done through multiple trials and error experiments, the gradient vanishing and exploding, idle points, and nonconvex optimizations are currently generating research interest. The development of new frameworks like Keras, Tensorflow, Theano, and Pytorch has stimulated the process of experimentation, and the community of researchers addressing these issues is increasing with increasing progress.

3.5. Generalization Power. The generalization power of a DL model is basically its ability to identify the correct sample and is often defined in terms of overfitting or underfitting. The complexity of a model should be increased only when needed. However, there are multiple approaches by which the overfitting problem of the DL models can be addressed. L1 and L2 regularizations can smooth the training process; a “dropout layer” between fully connected layers works in the same way to reduce the complexity of the model in a controlled fashion; and batch-normalizations techniques have the capacity to reduce overfitting by reducing the effect of some dominant neurons in the network.

3.6. Interpretability of DL Models. DL models, which are inherently black box models, do not provide much insight into their inner workings. For humans to trust these models, they need to be interpretable and explainable. Multiple publications [26, 118] have reported that a neural network could be fooled easily into choosing a wrong category by making minor changes to pixels and neither discriminative nor generative models are an exception. Interpretability alone might not be enough for humans to trust these black box models; they will need explainability. Explanatory Artificial intelligence (XAI) is one of the frontiers of DL research.

### Table 6: Application of GAN in condition monitoring of motors.

| Year | Input data and features | DL model | Application | Remarks |
|------|-------------------------|----------|-------------|---------|
| 2019 | Raw vibration           | ACGAN + batch normalization [116] | Bearing fault classification and data augmentation | Results revealed effectiveness of the model in classification due to data augmentation and batch normalization |
| 2020 | Raw vibration           | GAN + CNN classifier [117] | Bearing fault classification and data augmentation | The two-stage approach yielded promising results in terms of data augmentation and fault classification |

**Figure 9: Heatmap of DL models versus type of input data features.**
3.7. Hardware Challenges. DL models, regardless of their efficiency at solving the task on hand, invite hardware challenges, especially when their deployment in a standalone or embedded system is considered. Deep models due to their complex topology require higher computing power, energy, and memory. Training and optimization of these models is an iterative process during which multiple high-dimensional matrix multiplications are performed. Therefore, much of the research has focused on reducing the cost of multiply and accumulate (MAC) [119]. Key factors taken into consideration while integrating these models with hardware are accuracy, energy consumption, throughput/latency, and cost [120]. Designing an efficient DL architecture that incorporates all these factors is done through optimizing and compressing the DL models through algorithmic techniques or designing application-specific hardware. Algorithmic

![Table 7: Strengths and drawbacks of DL models.](image)

**Table 7:** Strengths and drawbacks of DL models.

| DL model     | Strengths                                                                 | Drawbacks                                                                 |
|--------------|---------------------------------------------------------------------------|---------------------------------------------------------------------------|
| MLP [47, 83] | (i) Can work with clean, balanced, and scaled data, regardless of the data type  |
|              | (ii) Integration in real-time systems is easy and allows one-dimensional data analysis |
|              | (iii) Allows to learn rich representations and reduces dimensionality       |
|              | (iv) Can work as a denoising technique to get cleaner data                  |
|              | (v) Easy implementation                                                    |
| AE [118, 127]| (i) Can achieve higher level of generalization on one-dimensional raw data |
|              | (ii) Allows for one-dimensional data analysis                              |
|              | (iii) Combined optimization of all the layer parameters                    |
| CNN [10, 83] | (i) Fits well for multidimensional data analysis                           |
|              | (ii) Enables for feature extraction from raw data                          |
| RNN/LSTM/GRU [83, 113, 129] | (i) Performs well with time series or sequential data | (i) Requires a lot of tuning to work on dirty or unscaled data |
|              | (ii) Better forecasting ability in time series and sequential data         |
|              | (iii) Computes optimization of all the layer parameters                    |
| GAN [115, 116, 130] | (i) Learns underlying representation of data well | (i) Requires massive training data and training time |
|              | (ii) Seems to achieve a discriminator with less generalization error        |
|              | (iii) Only algorithm that can work in semisupervised or even unsupervised setting to identify under observation clusters |
|              | (iv) Model with high fidelity                                               |

![Figure 10: Three-dimensional map of the number of publications of DL models in terms of type of input data.](image)
techniques usually focus on retaining the accuracy of a DL architecture after performing pruning/compression on it. On the other hand, hardware is designed to make it energy-efficient and reduce the latency. However, codesign of algorithm and hardware has also been explored [121]. Hardware solutions for deploying DL architectures range from general purpose solutions (GPUs) to application-specific solutions (FPGAs, ASICs). The optimum choice of hardware depends upon the application. Since CPUs are least used due to their limited throughput, GPUs are currently the preferred option for training and inference of DL models. FPGAs-based solutions particularly focus on efficient data routing and benefit from sparsity and reuse of data points already fetched in the computation network [122]. Moreover, hardware accelerators using variable bandwidth and reconfigurable data flow paths also have been developed [123–126]. Efficient deployment of DL architectures on edge is an active area of research and many paradigms are yet to be explored for effective performance.

The strengths and drawbacks of these models are summarized in Table 7.

4. Concluding Remarks

In this paper, a comprehensive review is presented related to the usage of various feature-processing techniques with DL models for condition monitoring of motors. This paper has reviewed the literature in terms of input data and feature types used in effective fault diagnosis of motors using DL models. More specifically, this work summarizes the applications of DL models in terms of various feature-processing techniques and how these techniques have solved various problems that allow to achieve high generalization. It was observed that the usage of feature-processing techniques has raised the efficiency of DL models and reduced hardware dependency. Moreover, usage of feature-processing techniques with DL models allowed faster execution owing to the feature extraction and removal of redundant information. It was found from the literature review that the AEs, DBM, DBNs, and MLP architectures have been in the focus of researchers for fault diagnosis and prognosis. On the other hand, CNN and RNN also have received attention for their complex architectures demand expert human knowledge and feature processing for successful implementation. Furthermore, GANs have solved the class imbalance problem in data to some extent. However, more effort is still required for acceptable results.

Moreover, it has been observed that most of the available work focuses on mechanical faults diagnosis using DL models, specifically bearing faults. However, very limited work is conducted related to electrical fault diagnosis and prediction using DL models. Researchers may be facing difficulties in introducing electrical faults owing to danger associated with these faults. However, it is crucial to explore this area because motors also get damaged owing to various electrical faults. A large portion of the DL-based condition monitoring of motors have been conducted in relation to fault classification. However, a few researchers have conducted research related to health index (HI) prediction and RUL estimation of the motor and its components. Meanwhile, data fusion techniques have been successfully used with various models, which allowed the improvement of model classification accuracy. More work is required in this area; through fusing various diagnostic methods such as current and vibration, model performance could be improved further. Consequently, it will uplift condition-based maintenance of motors by employing DL models.

Meanwhile, DL models have also exhibited some deficiencies, which can be viewed as prospective future opportunities for researchers and engineers in this domain. This review has demonstrated that the DL-based diagnosis mostly involve a supervised learning approach; however, the practicality of such an approach is highly challenging and time-consuming in real engineering scenarios. From a future perspective, DL models need to be employed for automatic end-to-end diagnosis, which includes feature learning from data acquisition to motor fault classification or prediction. Moreover, acquiring more data does not necessarily mean that the DL models will produce better results. Thus, feature-processing techniques are essential for better generalization. It was noticed that the majority of research in this domain was focused on fault diagnosis in certain components such as the bearing. Very little research has been conducted on root cause analysis, degradation, and RUL prediction. Considering future opportunities, there is an urgent need for advanced feature-processing techniques that can assist in analyzing the huge amounts of data and yield effective diagnosis and prognosis results. Furthermore, there is great demand for research on explainable AI (XAI), which will overcome the problem of the vague operation of DL algorithms. The authors of this review believe that the practitioners working in this domain would find this article very useful in solving their problems and evaluating the methods. Meanwhile, this review gives valuable pointers to a road map for future research in this domain. The results and discussion may be presented separately, or in one combined section, and may optionally be divided into headed subsections.

Data Availability

The references and data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

The authors thank the “Haptics, Human Robotics, and Condition Monitoring Lab” established in Mehran University of Engineering and Technology, Jamshoro, under the umbrella of the National Center of Robotics and Automation funded by the Higher Education Commission (HEC), Pakistan, and the support of the EPSRC under grant no. EP/
References

[1] A. Hmidet and O. Boubaker, “Real-time low-cost speed monitoring and control of three-phase induction motor via a voltage/frequency control approach,” Mathematical Problems in Engineering, vol. 2020, 2020.

[2] W. Zhang, M.-P. Jia, L. Zhu, and X.-A. Yan, “Comprehensive overview on computational intelligence techniques for machinery condition monitoring and fault diagnosis,” Chinese Journal of Mechanical Engineering, vol. 30, no. 4, pp. 782–795, 2017.

[3] D. Goyal and B. S. Pabla, “The vibration monitoring methods and signal processing techniques for structural health monitoring: a review,” Archives of Computational Methods in Engineering, vol. 23, no. 4, pp. 585–594, 2016.

[4] M. Deighton, Facility Integrity Management: Effective Principles and Practices for the Oil, Gas and Petrochemical Industries, Gulf Professional Publishing, London, UK, 2016.

[5] G. Yao, B. Fan, W. Wang, and H. Ma, “A novel median-point mode decomposition algorithm for motor rolling bearing fault recognition,” Mathematical Problems in Engineering, vol. 2020, 2020.

[6] P. Gangsar and R. Tiwari, “Comparative investigation of vibration and current monitoring for prediction of mechanical and electrical faults in induction motor based on multiclass-support vector machine algorithms,” Mechanical Systems and Signal Processing, vol. 94, pp. 464–481, 2017.

[7] S. E. Pandarakone, Y. Mizuno, and H. Nakamura, “Distinct fault analysis of induction motor bearing using frequency spectrum determination and support vector machine,” IEEE Transactions on Industry Applications, vol. 53, no. 3, pp. 3049–3056, 2016.

[8] N.-H. Kim, D. An, and J.-H. Choi, Prognostics and Health Management of Engineering Systems: An Introduction, Springer, Berlin, Germany, 2016.

[9] H. Chen, B. Jiang, W. Chen, and H. Yi, “Data-driven detection and diagnosis of incipient faults in electrical drives of high-speed trains,” IEEE Transactions on Industrial Electronics, vol. 66, no. 6, pp. 4716–4725, 2018.

[10] J.-H. Han, D.-J. Choi, S.-K. Hong, and H.-S. Kim, “Motor fault diagnosis using CNN based deep learning algorithm considering motor rotating speed,” 2019.

[11] A. Choudhary, D. Goyal, S. L. Shimi, and A. Akula, “Condition monitoring and fault diagnosis of induction motors: a review,” Archives of Computational Methods in Engineering, vol. 26, no. 4, pp. 1221–1238, 2019.

[12] A. K. S. Jardine, D. Lin, and D. Banjevic, “A review on machinery diagnostics and prognostics implementing condition-based maintenance,” Mechanical Systems and Signal Processing, vol. 20, no. 7, pp. 1483–1510, 2006.

[13] S. Yin, X. Li, H. Gao, and O. Kaynak, “Data-based techniques focused on modern industry: an overview,” IEEE Transactions on Industrial Electronics, vol. 62, no. 1, pp. 657–667, 2014.

[14] T. Han, D. Jiang, Q. Zhao, L. Wang, and K. Yin, “Comparison of random forest, artificial neural networks and support vector machine for intelligent diagnosis of rotating machinery,” Transactions of the Institute of Measurement and Control, vol. 40, no. 8, pp. 2681–2693, 2018.

[15] Z. Xu, C. Hu, F. Yang et al., “Data-driven inter-turn short circuit fault detection in induction machines,” IEEE Access, vol. 5, pp. 25055–25068, 2017.

[16] C. T. Kowalski and T. Orlowska-Kowalska, “Neural networks application for induction motor faults diagnosis,” Mathematics and Computers in Simulation, vol. 63, no. 3–5, pp. 435–448, 2003.

[17] A. Habib, F. K. Saddozai, A. Sattar, A. Khan, I. A. Hameed, and F. M. Kundi, “User intention mining in bussiness reviews: a review,” 2018.

[18] D. Shen, G. Wu, and H. I. Suk, “Deep learning in medical image analysis,” Information Fusion, vol. 19, pp. 221–248, 2017.

[19] S. Sun, C. Luo, and J. Chen, “A review of natural language processing techniques for opinion mining systems,” Information Fusion, vol. 36, pp. 10–25, 2017.

[20] H. A. Pierson and M. S. Gasgher, “Deep learning in robotics: a review of recent research,” Advanced Robotics, vol. 31, no. 16, pp. 821–835, 2017.

[21] H. Shao, H. Jiang, H. Zhang, and T. Liang, “Electric locomotive bearing fault diagnosis using a novel convolutional deep belief network,” IEEE Transactions on Industrial Electronics, vol. 65, no. 3, pp. 2727–2736, 2017.

[22] J.-F. Toubeau, J. Bottieau, F. Vallée, and Z. De Grève, “Deep learning-based multivariate probabilistic forecasting for short-term scheduling in power markets,” IEEE Transactions on Power Systems, vol. 34, no. 2, pp. 1203–1215, 2018.

[23] G. E. Hinton and R. R. Salakhutdinov, “Reducing the dimensionality of data with neural networks,” Science, vol. 313, no. 5786, pp. 504–507, 2006.

[24] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” Nature, vol. 521, no. 7553, pp. 436–444, 2015.

[25] R. Roberts, G. Gianconteri, L. Inzerillo, and G. Di Mino, “Towards low-cost pavement condition health monitoring and analysis using deep learning,” Applied Sciences, vol. 10, no. 1, p. 319, 2020.

[26] D.-T. Hoang and H.-J. Kang, “A survey on Deep Learning based bearing fault diagnosis,” Neurocomputing, vol. 335, pp. 327–335, 2019.

[27] J. Villalba-Diez, M. Molina, J. Ordonez-Meró, S. Sun, D. Schmidt, and W. Wellbrock, “Geometric deep lean learning: deep learning in industry 4.0 cyber-physical complex networks,” Sensors, vol. 20, no. 3, p. 763, 2020.

[28] S. Dagan, M. Kumar, M. R. Ayyagari, and G. Kumar, “A survey of deep learning and its applications: a new paradigm to machine learning,” Archives of Computational Methods in Engineering, vol. 23, pp. 1–22, 2019.

[29] Z. Wang, K. Liu, J. Li, Y. Zhu, and Y. Zhang, “Various frameworks and libraries of machine learning and deep learning: a survey,” Archives of Computational Methods in Engineering, vol. 23, pp. 1–24, 2019.

[30] Z. Chen, M. Wu, R. Zhao, F. Guretno, R. Yan, and X. Li, “Machine remaining useful life prediction via an attention based deep learning approach,” IEEE Transactions on Industrial Electronics, vol. 23, 2020.

[31] J. Wang, S. Li, Z. An, X. Jiang, W. Qian, and S. Ji, “Batch-normalized deep neural networks for achieving fast intelligent fault diagnosis of machines,” Neurocomputing, vol. 329, pp. 53–65, 2019.

[32] A. Suresh, R. Udendran, and M. Balamurugan, “Integrating IoT and machine learning - the driving force of industry 4.0,” Internet of Things for Industry 4.0, vol. 23, pp. 219–235, 2020.

[33] S. Nahavandi, “Industry 5.0-A human-centric solution,” Sustainability, vol. 11, no. 16, p. 4371, 2019.
[34] Y. Wang, D. Zhang, Y. Liu, B. Dai, and L. H. Lee, “Enhancing transportation systems via deep learning: a survey,” *Transportation Research Part C: Emerging Technologies*, vol. 99, pp. 144–163, 2019.

[35] S. Shao, R. Yan, Y. Lu, P. Wang, and R. Gao, “DCNN-based multi-signature induction motor fault diagnosis,” 2019.

[36] Y. Lei, B. Yang, X. Jiang, F. Jia, N. Li, and A. K. Nandi, “Applications of machine learning to machine fault diagnosis: a review and roadmap,” *Mechanical Systems and Signal Processing*, vol. 138, p. 106587, 2020.

[37] D. K. Soother, J. Dauplato, and A. Shaikh, “Vibration measurement system for the low power induction motor,” *Engineering Science and Technology International Research Journal*, vol. 2, pp. 53–57, 2018.

[38] R. Zhao, R. Yan, Z. Chen, K. Mao, P. Wang, and R. X. Gao, “Deep learning and its applications to machine health monitoring,” *Mechanical Systems and Signal Processing*, vol. 115, pp. 213–237, 2019.

[39] Z. Zhao, “Deep learning algorithms for rotating machinery intelligent diagnosis: an open source benchmark study,” *ISA Trans.*, vol. 23, 2020.

[40] G. Toh and J. Park, “Review of vibration-based structural health monitoring using deep learning,” *Applied Sciences*, vol. 10, no. 5, p. 1680, 2020.

[41] D. K. Soother and J. Dauplato, “A brief review of condition monitoring techniques for the induction motor,” *Transactions of the Canadian Society for Mechanical Engineering*, vol. 43, no. 4, pp. 499–508, 2019.

[42] A. Choudhary, S. Shimi, and A. Akula, “bearing fault diagnosis of induction motor using thermal imaging,” 2018.

[43] X. Zhang, Q. Zhang, M. Chen, Y. Sun, X. Qin, and H. Li, “A two-stage feature selection and intelligent fault diagnosis method for rotating machinery using hybrid filter and wrapper method,” *Neurocomputing*, vol. 275, pp. 2426–2439, 2018.

[44] Y. Lei, Z. He, Y. Zi, and Q. Hu, “Fault diagnosis of rotating machinery based on multiple ANFIS combination with GAs,” *Mechanical Systems and Signal Processing*, vol. 21, no. 5, pp. 2280–2294, 2007.

[45] V. Stojanovic and N. Nedic, “Robust identification of OE model with constrained output using optimal input design,” *Journal of the Franklin Institute*, vol. 353, no. 2, pp. 576–593, 2016.

[46] V. Bolón-Canedo, N. Sánchez-Maroño, and A. Alonso-Betanzos, “A review of feature selection methods on synthetic data,” *Knowledge and Information Systems*, vol. 34, no. 3, pp. 483–519, 2013.

[47] R. G. Vieira, C. M. Medeiros, and E. T. Silva, “Classification and sensitivity analysis to detect fault in induction motors using an MLP network,” 2016.

[48] R. Palacios, W. Godoy, A. Goedtel, I. da Silva, D. Morriño-Sotelo, and O. Duque-Perez, “Time domain diagnosis of multiple faults in three phase induction motors using intelligent approaches,” 2017.

[49] G. H. Bazan, P. R. Scalassara, W. Endo, A. Goedtel, W. F. Godoy, and R. H. C. Palacios, “Stator fault analysis of three-phase induction motors using information measures and artificial neural networks,” *Electric Power Systems Research*, vol. 143, pp. 347–356, 2017.

[50] S. Zolfaghari, S. B. M. Noor, M. Rezaazadeh Mehrjou, M. H. Marhaban, and N. Mariun, “Broken rotor bar fault detection and classification using wavelet packet signature analysis based on fourier transform and multi-layer perceptron neural network,” *Applied Sciences*, vol. 8, no. 1, p. 25, 2018.
[68] K. Zhao, H. Jiang, X. Li, and R. Wang, “An optimal deep sparse autoencoder with gated recurrent unit for rolling bearing fault diagnosis,” Measurement Science Technology, vol. 31, no. 1, 2019.
[69] H. Zhu, J. Cheng, C. Zhang, J. Wu, and X. Shao, “Stacked pruning sparse denoising autoencoder based intelligent fault diagnosis of rolling bearings,” Applied Soft Computing, vol. 88, 2020.
[70] L. Rao, W. Jin, and R. Pavel, “Enhanced restricted Boltzmann machine with prognostic capability for prognostics and health assessment,” IEEE Transactions on Industrial Electronics, vol. 63, no. 11, pp. 7076–7083, 2016.
[71] F. AlThobiani and A. Ball, “An approach to fault diagnosis of reciprocating compressor valves using Teager–Kaiser energy operator and deep belief networks,” Expert Systems with Applications, vol. 41, no. 9, pp. 4113–4122, 2014.
[72] R. Salakhutdinov and G. Hinton, “Deep Boltzmann machines,” 2009.
[73] Z. Chen, C. Li, and R.-V. Sánchez, “Multi-layer neural network with deep belief network for gearbox fault diagnosis,” Journal of Vibration and Control, vol. 17, no. 5, pp. 2379–2392, 2015.
[74] C. Zhang, J. H. Sun, and K. C. Tan, “Deep belief networks ensemble with multi-objective optimization for failure diagnosis,” IEEE International Conference on Systems, Man, and Cybernetics, vol. 23, 2015.
[75] S. Deng, Z. Cheng, C. Li, X. Yao, Z. Chen, and R.-V. Sanchez, “Rolling bearing fault diagnosis based on Deep Boltzmann machines,” 2016.
[76] J. Tao, Y. Liu, and D. Yang, “Bearing fault diagnosis based on deep belief network and multisensor information fusion,” Shock and Vibration, vol. 2016, 2016.
[77] M. Ma, X. Chen, S. Wang, Y. Liu, and W. Li, “Bearing degradation assessment based on weibull distribution and deep belief network,” 2016.
[78] D. Liu, Q. Wang, J. Tao, G. Li, and J. Wu, “Fault diagnosis method based on improved deep Boltzmann machines,” 2018.
[79] G. Niu, B. Zhang, P. Ziehl, F. Ferrese, and M. Golda, “Rolling element bearing fault diagnosis based on deep belief network and principal component analysis,” Proceedings of the Annual Conference of the PHM Society, vol. 11, p. 1, 2019.
[80] H. Zhao, H. Liu, J. Xu, C. Guo, and W. Deng, “Research on a fault diagnosis method of rolling bearings using variation mode decomposition and deep belief network,” Journal of Mechanical Science and Technology, vol. 33, no. 9, pp. 4165–4172, 2019.
[81] X. Yan, Y. Liu, and M. Jia, “Multiscale cascading deep belief network for fault identification of rotating machinery under various working conditions,” Knowledge-Based Systems, vol. 193, p. 105484, 2020.
[82] K. Yu, T. R. Lin, and J. Tan, “A bearing fault and severity diagnostic technique using adaptive deep belief networks and Dempster-Shafer theory,” Structural Health Monitoring, vol. 19, no. 1, pp. 240–261, 2020.
[83] W. Liu, Z. Wang, X. Liu, N. Zeng, Y. Liu, and F. E. Alsaadi, “A survey of deep neural network architectures and their applications,” Neurocomputing, vol. 234, pp. 11–26, 2017.
[84] O. Janssens, V. Slavkovikj, B. Vervisch et al., “Convolutional neural network based fault detection for rotating machinery,” Journal of Sound and Vibration, vol. 377, pp. 331–345, 2016.
[85] R. Liu, G. Meng, B. Yang, C. Sun, and X. Chen, “Dislocated time series convolutional neural architecture: an intelligent fault diagnosis approach for electric machine,” IEEE Transactions on Industrial Informatics, vol. 13, no. 3, pp. 1310–1320, 2016.
[86] T. Ince, S. Kiranyaz, L. Eren, M. Askar, and M. Gabbouj, “Real-time motor fault detection by 1-D convolutional neural networks,” IEEE Transactions on Industrial Electronics, vol. 63, no. 11, pp. 7067–7075, 2016.
[87] X. Guo, L. Chen, and C. Shen, “Hierarchical adaptive deep convolutional neural network and its application to bearing fault diagnosis,” Measurement, vol. 93, pp. 490–502, 2016.
[88] X. Dang and Q. He, “Energy-fluctuated multiscale feature learning with deep convnet for intelligent spindle bearing fault diagnosis,” IEEE Transactions on Instrumentation and Measurement, vol. 66, no. 8, pp. 1926–1935, 2017.
[89] S. Li, G. Liu, X. Tang, J. Lu, and J. Hu, “An ensemble deep convolutional neural network model with improved D-S evidence fusion for bearing fault diagnosis,” Sensors, vol. 17, no. 8, p. 1729, 2017.
[90] W. Zhang, C. Li, G. Peng, Y. Chen, and Z. Zhang, “A deep convolutional neural network with new training methods for bearing fault diagnosis under noisy environment and different working load,” Mechanical Systems and Signal Processing, vol. 100, pp. 439–453, 2018.
[91] L. Guo, Y. Lei, N. Li, T. Yan, and N. Li, “Machinery health indicator construction based on convolutional neural networks considering trend burr,” Neurocomputing, vol. 292, pp. 142–150, 2018.
[92] F. Jia, Y. Lei, N. Lu, and S. Xing, “Deep normalized convolutional neural network for imbalanced fault classification of machinery and its understanding via visualization,” Mechanical Systems and Signal Processing, vol. 110, pp. 349–367, 2018.
[93] D.-T. Hoang and H.-J. Kang, “Rolling element bearing fault diagnosis using convolutional neural network and vibration image,” Cognitive Systems Research, vol. 53, pp. 42–50, 2019.
[94] G. Li, C. Deng, J. Wu, X. Xu, X. Shao, and Y. Wang, “Sensor data-driven bearing fault diagnosis based on deep convolutional neural networks and S-transform,” Sensors (Basel), vol. 19, no. 12, p. 2750, 2019.
[95] Z. Zhu, G. Peng, Y. Chen, and H. Gao, “A convolutional neural network based on a capsule network with strong generalization for bearing fault diagnosis,” Neurocomputing, vol. 323, pp. 62–75, 2019.
[96] D. T. Hoang and H. J. Kang, “A motor current signal based bearing fault diagnosis using deep learning and information fusion,” IEEE Transactions on Instrumentation and Measurement, vol. 23, 2019.
[97] G. Li, C. Deng, J. Wu, Z. Chen, and X. Xu, “Rolling bearing fault diagnosis based on wavelet packet transform and convolutional neural network,” Applied Sciences, vol. 10, no. 3, p. 770, 2020.
[98] Y. Zhang, K. Xing, R. Bai, D. Sun, and Z. Meng, “An enhanced convolutional neural network for bearing fault diagnosis based on time–frequency image,” Measurement, vol. 10, 2020.
[99] M. Skowron, T. Orłowska-Kowalska, M. Wołkiewicz, and C. T. Kowalski, “Convolutional neural network-based stator current data-driven incipient stator fault diagnosis of inverter-fed induction motor,” Energies, vol. 13, no. 6, p. 1475, 2020.
[100] J. Schmidhuber, “Deep learning in neural networks: an overview,” Neural Networks, vol. 61, pp. 85–117, 2015.
[101] S. Hochreiter and J. Schmidhuber, “Long short-term memory neural computation 9,” 1997.
[102] R. Zhao, J. Wang, R. Yan, and K. Mao, "Machine health monitoring with LSTM networks," 2016.
[103] R. Zhao, D. Wang, R. Yan, K. Mao, F. Shen, and J. Wang, "Machine health monitoring using local feature-based gated recurrent unit networks," IEEE Transactions on Industrial Electronics, vol. 65, no. 2, pp. 1539–1548, 2017.
[104] Q. Cui, Z. Li, J. Yang, and B. Liang, "Rolling bearing fault prognosis using recurrent neural network," 2017.
[105] D. Xiao, Y. Huang, X. Zhang, H. Shi, C. Liu, and Y. Li, "Fault diagnosis of asynchronous motors based on LSTM neural network," 2018.
[106] G. Tang, Y. Zhou, H. Wang, and G. Li, "Prediction of bearing performance degradation with bottleneck feature based on LSTM network," 2018.
[107] D. Xiao, Y. Huang, C. Qin, H. Shi, and Y. Li, "Fault diagnosis of induction motors using recurrence quantification analysis and LSTM with weighted BN," Shock and Vibration, vol. 2019, 2019.
[108] L. Yu, J. Qu, F. Gao, and Y. Tian, "A novel hierarchical algorithm for bearing fault diagnosis based on stacked LSTM," Shock and Vibration, vol. 2019, 2019.
[109] B. Zhang, S. Zhang, and W. Li, "Bearing performance degradation assessment using long short-term memory recurrent network," Computers in Industry, vol. 106, pp. 14–29, 2019.
[110] K. Zhao and H. Shao, "Intelligent fault diagnosis of rolling bearing using adaptive deep gated recurrent unit," Neural Processing Letters, vol. 23, pp. 1–20, 2019.
[111] N. Enshaei and P. Naderkhani, "Application of deep learning for fault diagnostic in induction machine’s bearings," 2019.
[112] X. Li, H. Jiang, X. Xiong, and H. Shao, "Rolling bearing health prognosis using a modified health index based hierarchical gated recurrent unit network," Mechanism and Machine Theory, vol. 133, pp. 229–249, 2019.
[113] S. Haidong, C. Junsheng, J. Hongkai, Y. Yu, and W. Zhantao, "Enhanced deep gated recurrent unit and complex wavelet packet energy moment entropy for early fault prognosis of bearing," Knowledge-Based Systems, vol. 188, p. 105022, 2020.
[114] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, and D. Warde-Farley, "Generative adversarial nets in advances in neural information processing systems (NIPS)," 2014.
[115] H. Alaftani, M. Kavaki-Thorne, and G. Kumar, "Applications of generative adversarial networks (GANS): an updated review," Archives of Computational Methods in Engineering, vol. 23, pp. 1–28, 2019.
[116] S. Shao, P. Wang, and R. Yan, "Generative adversarial networks for data augmentation in machine fault diagnosis," Computers in Industry, vol. 106, pp. 85–93, 2019.
[117] W. Zhang, X. Li, X.-D. Jia, H. Ma, Z. Luo, and X. Li, "Machinery fault diagnosis with imbalanced data using deep generative adversarial networks," Measurement, vol. 152, p. 107377, 2020.
[118] H. Sagha, N. Cummins, B. Schuller, and K. Discovery, "Stacked denoising autoencoders for sentiment analysis: a review," Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, vol. 7, no. 5, p. e1212, 2017.
[119] V. Sze, Y. Chen, J. Emer, A. Suleiman, and Z. Zhang, "Hardware for machine Learning: challenges and opportunities," 2016.
[120] B. Moons, D. Bankman, and M. Verhelst, “Embedded deep learning,” 2019.
[121] H. Song, Efficient Methods and Hardware for Deep Learning, Cadence, San Jose, UK, 2017.
[122] M. Capra, B. Bussolino, A. Marchisio, M. Shafique, G. Masera, and M. Martina, An Updated Survey of Efficient Hardware Architectures for Accelerating Deep Convolutional Neural Networks, Future Internet, London, UK, 2020.
[123] H. Sharma, “Bit fusion: bit-level dynamically composable architecture for accelerating deep neural network," 2018.
[124] S. Ryu, H. Kim, W. Yi, and J. J. Kim, “Area and energy-efficient precision-scalable neural network accelerator with bitwise summation," 2019.
[125] H. Kwon, A. Samajdar, and T. Krishna, "Maeri," ACM SIGPLAN Notices, vol. 53, no. 2, pp. 461–475, 2018.
[126] W. Lu, G. Yan, J. Li, S. Gong, Y. Han, and X. F. Li, "A flexible dataflow accelerator architecture for convolutional neural networks," 2017.
[127] E. Principi, D. Rossetti, S. Squartini, and F. Piazza, "Unsupervised electric motor fault detection by using deep autoencoders," IEEE/CAA Journal of Automatica Sinica, vol. 6, no. 2, pp. 441–451, 2019.
[128] T. Pan, J. Chen, J. Pan, and Z. Zhou, “A deep learning network via shunt-wound restricted Boltzmann machines using raw data for fault detection," IEEE Transactions on Instrumentation and Measurement, vol. 23, 2019.
[129] H. F. R. dos Santos, L. W. C. da Silva, and A. P. B. Sobral, “Forecast of multivariate time series sampled from industrial machinery sensors," Brazilian Journal of Operations and Production Management, vol. 17, no. 1, pp. 1–12, 2020.
[130] A. Odena, "Semi-supervised learning with generative adversarial networks," 2016.