Expert Systems With Applications 221 (2023) 119738

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Review

Condition Monitoring using Machine Learning: A Review of Theory, Applications, and Recent Advances

Onur Surucu[a], Stephen Andrew Gadsden[b,*], John Yawney[a]

[a] Adastra Corporation, Royal Bank Plaza, South Tower, 200 Bay St, Toronto, Ontario, Canada
[b] McMaster University, 1280 Main St West, Hamilton, Ontario, Canada

ARTICLE INFO

Keywords:
Artificial intelligence
Condition monitoring
Data processing
Fault detection
Machine learning
Predictive maintenance

ABSTRACT

In modern industry, the quality of maintenance directly influences equipment's operational uptime and efficiency. Hence, based on monitoring the condition of the machinery, predictive maintenance can minimize machine downtime and potential losses. Throughout the field, machine learning (ML) methods have become noteworthy for predicting failures before they occur. However, the efficacy of the predictive maintenance strategy relies on selecting the appropriate data processing method and ML model. Existing surveys do not comprehensively inform users or evaluate the quality of the monitoring systems proposed. Hence, this survey reviews the recent literature on ML-driven condition monitoring systems that have been beneficial in many cases. Furthermore, in the reviewed literature, we provide an insight into the underlying findings on successful, intelligent condition monitoring systems. It is prudent to consider all factors when narrowing the search for the most effective model for a particular task. Therefore, the tradeoff between task constraints and the performance of each diagnostic technique are quantitatively and comparatively evaluated to obtain the given problem's optimal solution.

1. Introduction

Production plants are expected to run twenty-four hours a day to meet market demand. Unexpected equipment breakdowns cause tremendous economic stresses through significant process downtime. Most companies require that these interruptions are anticipated in advance to take the necessary precautions before stoppages occur unexpectedly. A non-intrusive procedure for tracking and detecting potential faults in systems is obligatory for mechanical and electrical devices. The manufacturing industry has reported a considerable increase in the frequency of accidents due to poor and dangerous maintenance practices (Rao, 1998). Every year, industry in the U.S. spends approximately $200 billion on maintaining plant equipment and facilities while poor maintenance causes losses of up to $60 billion (Mobley, 2002). As a solution, there are three traditional maintenance strategies (Tavner et al., 2020):

1. Breakdown maintenance.
2. Planned maintenance.
3. Condition-based maintenance or predictive maintenance.

Method (i) runs the equipment until it breaks, then replaces it. Method (ii) applies periodical maintenance at regular intervals with or without machinery monitoring. The final method (iii) proposes monitoring machinery's health, and it suggests applying the proper maintenance based on the identified diagnosis. Such diagnostic techniques implicitly yield higher plant efficiency, and reduced replacement expenditure and financial losses caused by unexpected breakdowns. According to Davies's market research, the outcome of predictive maintenance was remarkably rewarding for the British market during the'90s (A. Davies, 1998).

Predictive maintenance (PdM) or condition-based monitoring is an advanced diagnostic technique to reveal the operating machinery faults in their incipient phase before any breakdowns occur. Therefore, required maintenance can be done by analyzing the equipment's sensor signals. In a production scenario, this intelligent monitoring procedure can be applied either directly (offline) or indirectly (online) (Serin et al., 2020). The offline monitoring strategy conducts machine-aided periodic onsite inspection that requires operations to be interrupted. Conversely, online monitoring continuously checks the equipment through sensors during operations. PdM techniques can be categorized into two

* Corresponding author.
E-mail addresses: onur.surucu@adastragrp.com (O. Surucu), gadsden@mcmaster.ca (S.A. Gadsden), john.yawney@adastragrp.com (J. Yawney).

https://doi.org/10.1016/j.eswa.2023.119738
Received 19 October 2021; Received in revised form 1 July 2022; Accepted 21 February 2023
Available online 24 February 2023

strategies: model-based or data-driven (Jaber, 2017a). The model-based approaches build a mathematical model of the system based on empirical data and regard any deterministic change between the output of the actual and the mathematical system as indicating a fault (Windmann et al., 2015). However, the application of model-based approaches is impractical in real-life settings and is only possible within a specific set of conditions or a controlled environment (Jaber, 2017a; Wong et al., 2020). A data-driven strategy uses intelligent models to recognize fault indicators from machinery's lifecycle data (W. Zhang et al., 2019). According to a recent report, machine learning (ML) techniques are selected as an intelligent model for PdM (Lin et al, 2019). In Fig. 1, we visualized a pipeline workflow of a complete PdM with an ML model. The explanation of each process mainly belongs to T. Marwala (Marwala, 2012; Tavner et al., 2020).

Data acquisition step converts, amplifies, corrects the outputs from multiple sensors, and finally stores them in a computer. Sensors are designed to convert the physical environmental inputs into electrical signals. Physical characteristics of the industrial asset are acquired utilizing sensors installed on the equipment. Here are the typical sensor measurements in PdM: vibration, acoustic emission, strain, temperature, and current (Liton Hossain et al., 2018). The robustness and reliability of the diagnosis depend on the quality of the measured data. According to studies on sensor implementation techniques, it has been proven that a suitable sensor placement strategy is crucial in real-world applications where candidate sensor locations are ranked (Gao et al., 2006). Traditionally, to carry out data acquisition tasks in real-world applications, a data acquisition instrument (e.g., an accelerometer) is used to acquire the data from the equipment (T M et al., 2019). At the next step, Analog to Digital conversion, correction, and amplification of data are typical follow-up processes depending on the signal type (Marwala, 2012). Finally, the processed data are stored in a storage device.

Data processing step is employed to convert raw signals to a useful shape, where the trends and patterns can be easily identified (Stachowiak et al., 2005). Accuracy of the decision algorithm is highly dependent on the quality of the input data. Therefore, feature extraction and selection methods are standard approaches to achieve optimal data vectors. Feature extraction is responsible for reducing dimensions, handling missing values, and correcting irregularities in the acquired data. Feature selection reduces the extracted features by removing the input vectors' redundant attributes. This way, the ML model can diagnose without being exposed to any deceptive or false data.

Condition Prognosis and Diagnosis step is responsible for identifying and classifying the fault using ML models. According to T. Marwala, the ideal procedure for estimating the machine's well-being is subdivided into five stages (Marwala, 2012). Initially, the processed data are employed to train the intelligent algorithm that performs *fault detection*. Later, in the *fault classification* phase, more information is extracted from the physical defect, including the nature, extent, and type of the failure. The next step is to identify the *fault location* in the tool. Fault *quantification* is the next stage where the detected fault's magnitude is estimated and quantified. The final stage is to predict the *remaining life* of the machinery that is being monitored. These are the ideal steps for condition monitoring, yet the procedure varies depending on the specific requirements and the system's nature. Most of the studies seek only one or two of the above features.

This paper presents a survey on condition monitoring using ML methods. The objective is to summarize and review recent developments in this field. The following databases were used to perform a literature search: Web of Science, Engineering Village, Springer, and Google Scholar served as the main databases. The article topics were filtered based on the following keyword index {condition monitoring} OR {fault detection} OR {predictive maintenance}. Further keyword filtering is applied to find the relevant ML-driven PdM studies published between 2009 and 2021. This survey paper's remaining structure is outlined as follows: Section 2 describes common data processing methods for converting raw signals into a meaningful data vector. Section 3 demonstrates the diagnostic and prognostic ML techniques that portray promising results for condition monitoring. Section 4 highlights and interprets the underlying findings of the successful methods reviewed in the previous sections. Last, Section 5 summarizes each paper surveyed based on scholarly reviews. In summary, this survey provides more insights into the ML methods compared to other survey papers in this field, and a discussion in-depth concludes our underlying findings on the reviewed literature.

2. Data processing

2.1. Feature extraction techniques

Feature extraction reduces the dimension of the initial input data into a feature set of a (desirably) lower dimension that contains most of the vital information of the original data. An ML system's robustness depends on the quality of the extracted features and the monitoring system's reliability. Unfortunately, not all features are meaningful or possess relevant information about a machine's condition. Hence, removing redundant and irrelevant features is imperative. After a signal is captured, the characteristics and fault indicating features are extracted utilizing statistical-based signal processing techniques. In signal processing, the feature extraction methods are categorized into five groups: the *time domain, frequency domain,* and *time-frequency domain, model-based information extraction,* and *model-based information extraction* (Jaber, 2017b).

Time domain refers to a time series of machine signals that are not transformed into another domain (e.g., frequency domain). The goal is to detect the original signal's statistical characteristics by exploiting the series of discrete data. The standard forms are considered to be the peak-to-valley ratio, root-mean-square (RMS), mean, area under the curve, slope, shape factor, variance, entropy, and the crest factor (Caesarendra et al., 2013). Time domain analysis is one of the most employed techniques in PdM models. Kim et al. employed RMS to extract abnormal patterns from raw vibration signals, where extracted signals increased the accuracy of localized defect detection in low-speed bearings (Kim et al., 2007). In the presence of non-stationary signals, more complicated statistical processes (i.e., kurtosis and skewness) may provide potential solutions by analyzing sharpness or spike patterns (Caesarendra, 2016). Anomaly patterns in the system distort the characteristics and

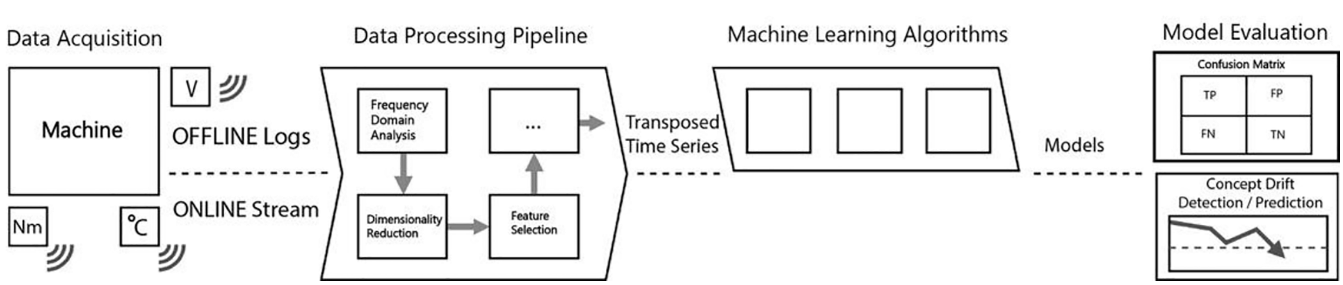

Fig. 1. Sample ML-driven PdM model.

probability distribution function (PDF) of the signal (Caesarendra, 2016). These two features observe the PDF of the signal to identify any changes in the distribution function. Kurtosis measures the relative flatness or spikiness of a signal compared to its non-faulty state (Marwala, 2012). It has proven its usefulness for identifying spontaneous transitions within vibration signals, while the distribution shape of signals is represented through the skewness parameter (Z. Wang et al., 2001). In PDF, a normal distribution equals zero skewness, and depending on the asymmetry of the distribution the skewness value moves towards higher negative or positive values. The study on discriminating motor faults has shown that fault detection can be significantly enhanced by utilizing statistical process control techniques combined with time-domain analysis (Tseng et al., 2014).

Frequency-domain shows how the signal's amplitude is distributed over a range of frequencies. In some cases (e.g., high damping), relevant information may be hidden in the frequency domain rather than the time domain (Marwala, 2012). Therefore, Fourier transform (FT) is essential to represent the signal in a frequency domain using sine waves. This representation provides an alternative perspective where the magnitude of the signals within the given frequency band can be monitored. Among FTs, the discrete Fourier transform (DTF) is considered the fastest (Gray & Goodman, 1995), yet a fast Fourier transform (FFT) is required to compute the DTF (Su & Chong, 2007). The industry has been applying the FFT to analyze many machinery signals: the induction motor current (Yoo, 2019), the machine vibration of rotating machinery (Atoui et al., 2013), U-phase load current of the stator (Pandarakone et al., 2018b). FFT is capable of performing stationary analysis, however when it is alone, it cannot clearly display the nonstationary signals of faulty components, as shown in Fig. 2 (Liton Hossain et al., 2018)(H. Li et al., 2009). The reason behind is non-stationary signals have a tendency of containing random frequency components compared to stationary signals. There are numerous approaches to overcome this critical problem. After implementing FFT, non-stationary signals can be analyzed by expressing them as constants using synchronous sampling methods where the signals are assessed based on a sample clock (Pepper, 2000). Another approach is to use the frequency response function (FRF), which is derived using the response to excitation ratio in the frequency domain (Marwala, 2000). Similar to time-domain analysis, to distinguish any indicator of non-stationary data, the frequency domain uses similar statistical parameters, known as *statistical frequency-domain features* or *frequency parameter indices*. The statistical features exhibit a different reaction when the frequency elements of the system change. Typical frequency parameter indices are described as follows: frequency center (FC), root mean square frequency (RMSF), and root variance frequency (RVF) (Niu, 2017). The RMSF and FC show the position variation of main frequencies, whereas the RVF indicates the power spectrum convergence, which is known as power distribution with the frequency. Similar to remaining time domain analysis, frequency signals can be extracted using spectral skewness (SS), spectral kurtosis (SK), spectral entropy, and Shannon's entropy (Sandoval et al., 2019).

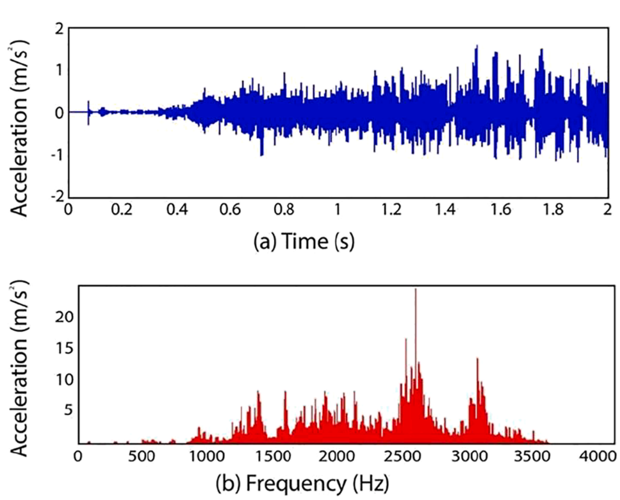

Fig. 2. Time domain and FTT analysis of non-stationary vibration signals of a faulty gear. (a) Time-domain (b) FFT. Adapted from (H. Li et al., 2009).

Time-frequency domain represents the time periods at which various signal frequencies are more dominant. For instance, damage caused by fatigue failure can cause non-stationary signals in the sensor signal. In such cases, it can be beneficial to analyze the behaviour of frequency components in the given time period (Marwala, 2000). In frequency domain analysis, the FT is used to determine the signal's amplitudes and frequencies, yet it is not specified at which intervals the corresponding amplitude occurs (Layer & Tomczyk, 2015). Thus, a domain holding both time and frequency is required while analyzing non-stationary signals. Various approaches have been derived for analyzing the time-frequency domain in fault detection; these are Wavelet Transform (WT), Short-Time Fourier Transform (STFT), and Wigner-Ville Distribution (WVD) (Wigner, 1997). These methods output the two-dimensional functions of time and frequency, known as the *time-frequency domain.* STFT divides the non-stationary signals into small windows of equal time and uses FFT to decompose the original signals at predetermined intervals. Cocconcelli et al. implemented the STFT as a feature extraction method for detecting damage in ball bearings (Cocconcelli et al., 2012), where an abnormal behaviour in sum of the STFT coefficients is regarded as an anomaly or a fault indicator. However, Layer et al. remark that selecting an appropriate STFT window width have been a problem, and an incorrect selection may result in blurred time-frequency data (Layer & Tomczyk, 2015). In wavelet transform (WT), this problem has been overcome by replacing the time window with a wavelet function. This approach exploits short windows at high frequencies and long windows at low frequencies. The WT method is frequently used in pattern recognition of discontinuous and unsteady frequency domain data (Daubechies, 1990). Lastly, WVD is another powerful method used in determining the time-frequency domain of nonstationary instabilities. It provides high-resolution representations in both the time and frequency domains. Unlike STFT and WT, this method represents the time-frequency distributions in a two-dimensional plane (image or matrix form) (Debnath & Shah, 2015). Any changes in energy distribution at the location affect the signal's amplitude. This phenomenon can be seen through lighter shades in the WVD image (Singru et al., 2018). In detecting physical damage to the scrutinized part, a joint time-frequency analysis technique called WVD has been derived for non-stationary signals (H. Li et al., 2006).

Model-based information extraction refers to the estimation of unmeasurable state variables in a dynamic system. Some variables cannot be measured directly in real-world applications but can be predicted utilizing several estimation techniques (Nursalam, 2016 & Fallis, 2013). In some cases, the instrumentation system cannot directly access fundamental quantities that have immense importance for fault detection. In such cases, the "inferential estimation" approach should estimate the essential measurements. The traditional approach is using the Kalman Filter (KF) to predict linear dynamic systems, yet the nonlinear systems are generally estimated using a derivation of KF: the unscented KF (UKF) and extended KF (EKF) (Grimble & Majecki, 2020). Classical KF is known as a "state observer," which provides a recursive solution to the linear filtering estimation problem based on the state-space equation of linear dynamical systems. For example, due to measurement and transmission errors, the actual state of power systems is challenging to measure. Liu et al. implemented Kalman filter (KF) extensions (i.e., EKF and UKF) to predict the dynamic state of power systems (Liu et al., 2020). Another example of an immeasurable signal is the permanent magnet temperature, which is critical for safely operating high speed-rotors. Since the magnets spin during operation, measuring and transmitting temperature data is difficult. Hence, KF is engaged to estimate the permanent magnet's temperature using other machine signals (Feng et al., 2018). As mentioned, KF based strategies are compelling; however, depending on the application, they lack robustness when modelling errors, disturbances, and uncertainties (Gadsden, 2011). Recently, novel estimation strategies, the sliding innovation filter (SIF), and extended SIF (ESIF) outperformed the KF-based methods when there were uncertainties and disturbances (Gadsden & AlShabi, 2020). Additional to KF, this novel state estimator benefits from the switching term, which allows the predictions to slide along a decision boundary layer.

Dimensionality-reduction refers to reducing the dimensions of the input data vector by transforming it into a lower-dimensional space without disrupting its original form. Using high-dimensional data is complicated since interpretation is problematic, storage is expensive, and analysis is complicated. Therefore, a lower-dimensional structure allows us to work with a more compact representation of the data. In this feature extraction method, the objective is to reduce the dataset's dimensionality to improve the learning model's efficiency by transforming the input dataset. According to a review on monitoring large-scale HVAC systems, the most common methods for data reduction are principal component analysis (PCA) and linear discriminant analysis (LDA) (Mirnaghi & Haghighat, 2020). PCA is a technique for compressing high-dimensional data into independent orthogonal components without losing the input vector's correlation. The method utilizes eigenvalues and eigenvectors to perform orthogonal projection onto the principal subspace (Deisenroth et al., 2020). In PdM applications, PCA can be applied to a wide range of types of machinery: Induction motor (Ghate & Dudul, 2010), stirred tank heater (Amin et al., 2019), and hydraulic system (P. Guo et al., 2019). Another common approach for dimensionality-reduction is LDA. The method aims to determine a collection of vectors that best discriminate among classes. Unlike PCA, LDA aims to separate each class of reduced data by utilizing the data labels. This popular method has proven its applicability in induction motors (Jung et al., 2010), transmission lines (A novel transmission line relaying scheme for fault detection and classification using wavelet transform and linear discriminant analysis, 2015), hydraulic systems (Helwig et al., 2015), and many more applications throughout the predictive maintenance industry.

2.2. Feature selection techniques

The accuracy of a machine learning model is highly dependent on the quality of the input data. Redundant attributes must be filtered after feature extraction because, a large dataset with numerous attributes introduces inefficient complexity to the ML model, which ultimately leads to overfitting and less accurate results. Besides, reducing the number of features provides less computational and memory requirements. It is worthwhile to mention that none of these procedures requires data transformation, unlike feature extraction. Selection techniques are divided into three approaches: wrapper, filter, and embedded (Kumar, 2014).

Filter methods rank each data vector, independently of the ML model. In a case study, the minimum redundancy maximum relevance (mRMR) method was proposed and tested to use for tool condition monitoring (Fernandes et al., 2019). The algorithm reduced the feature space from 47 to 32 by ranking the features by their relevance to the objective and penalizing those that are redundant. *Maximum relevance* means finding the subsets of features that best represent the target class, and *minimum redundancy* refers to finding the n most distinct features in the feature set. Another common filter method is Fisher's discriminant ratio (FDR). FDR quantifies the discriminatory power of individual features. Thus, useful features can be selected based on their FDR scores. However, the method can also be used to gain insights into the input data. Bhat et al. analyzed input signals' behaviour by examining the FDR score's sensitivity to different conditions (Bhat et al., 2016). For instance, they discovered that some features' FDR score dropped significantly when a noise was introduced. As a solution, they picked the most stable features against external factors, which significantly increased the classification rate.

Wrapper methods measure the relative usefulness of features by examining the learning algorithm's performance. For example, genetic algorithms (GA) are implemented as wrappers in the condition monitoring applications to find the optimal group from the input dataset. For instance, Baraldi et al. used GA to find the optimal feature subset among 46 nuclear power plant sensor signals (Baraldi et al., 2011). According to the results, the learning algorithm had a ten times improved performance with the GA's selected feature subset.

Embedded methods select the best features while the learning algorithm' executes. Usually, these methods are integrated into the learning algorithm. A typical example is the random forest method, which is detailly explained in section 3: ensemble learning (Jović et al., 2015).

3. Machine learning methods for condition monitoring

ML is a branch of artificial intelligence specialized in building algorithms that learn from data and continuously improve its performance over time without human intervention (H. Wang et al., 2009). In 2020, trend analysis for the PdM industry reported that the recent trend throughout the field is toward ML-driven solutions (Çinar et al., 2020). In the report, the authors emphasized that the ML is the fundamental element in achieving the following advancements: Stoppage reduction, maintenance cost reduction, spare-part life increases, operator safety, increased production, and repair verification. In order to analyze ML models, we divided them into four subsections: classification, clustering, regression, and ensemble. In each subsection, we provided a brief explanation of ML-driven PdM models, the dataset, and the experimental results.

3.1. Classification analysis

Classification analysis is an ML method for predicting the class value of analyzed data using prior observations. This approach can be supervised or unsupervised (Stetco et al., 2019). Classification is a fundamental task in the machine fault identification and classification framework. In condition monitoring, fault detection can be referred to as binary classification, either as a faulty or healthy case. On the other hand, the fault classification can be regarded as a multi-class classification where the input is to be classified into non-overlapping multiple classes (e.g., degree of the fault). There are several standard measurement techniques (performance metrics) to evaluate the model's performance. If the model uses a classification-based strategy for diagnostic purposes, here are the typical metrics: accuracy, recall, precision, recall, specificity, and F1 (Sokolova & Lapalme, 2009).

3.1.1. Multilayer perceptron

Artificial Neural Networks (ANN) are deep learning computer algorithms that mimic the working principles of the human brain (Shanmuganathan, 2016). The processing elements, known as neurons, consist of weights and biases. Each neuron is designed to transmit a signal to other neurons like the synapses in a human brain. The network structure consists of an arbitrary number of neurons in three main types of layers: the input layer, the hidden layers, and the output layer. ANN architecture is subdivided into two branches: feed-forward neural networks and recurrent or feedback neural networks. The multilayer perceptron (MLP) is one of the simplest supervised learning methods among the feed-forward neural networks, transferring signals in only one direction (forward) from the input layer, through the hidden layers, to the output layer (Gardner & Dorling, 1998). The network has a densely linked formation, where each neuron is interconnected between the entire forward layer. Each neuron is fired by an assigned activation function. This function acts as a mathematical "gate" between the input neuron and the output going to the next layer as a continuous function. In summary, the model is capable of learning complex representations through activation functions. Due to its simplicity, a typical choice for an activation function is the ReLU (rectified linear unit); however, the industry uses other activation functions in many scenarios: Sigmoid, Softmax, Leaky ReLU, PReLU (X.-D. Zhang, 2020). As the last step, the model learns the existing data pattern using the backpropagation (BP) algorithm. The error between the desired output and the network's output is propagated backward by varying the network's weights and biases (Du & Swamy, 2014). Thus, the learning algorithm is capable of improving its accuracy in pattern recognition. In the past, this algorithm has been utilized to perform diagnostic and prognostic predictions for many types of industrial equipment.

In 2011, an MLP structure was implemented to expose the potential physical damage on the bearing of the electric motor based on vibration data (Vijay et al., 2011). In a laboratory environment, the dataset is acquired by recording the vibration signals of faulty and healthy bearings. The MLP had three main layers: an input (10 neurons), one hidden, and an output. The neurons of which the input and hidden layer consist were fired by the sigmoid activation function. The dataset was split into training (80%) and testing (20%) sets. The hidden layer architecture was detected by varying the number of neurons in the hidden layer from 5 to 25. After a qualitative analysis, the lowest MSE results were achieved with 16 neurons in the hidden layer. After completing the training, this MLP achieved a 100% fault identification accuracy on both the testing and training sets (Vijay et al., 2011). In 2010, Ghate et al. proposed an optimal MLP neural network classifier for fault detection in a three-phase induction motor (Ghate & Dudul, 2010). The stator currents were processed using statistical parameters: RMSE, the maximum and minimum values, skewness, and the kurtosis coefficients. Later, PCA reduced the number of features from 13 to 5. The objective of this study was to detect if the given AC currents are faulty or healthy. The authors selected three different machine learning models to perform prediction: MLP, PCA-MLP, and neural network with self-organizing map (SOM-NN). Due to the results, the dimensionality reduction process (i.e., PCA) increased the average fault detection accuracy from 95.33% to 97.25% for SOM-NN, and MLP's accuracy only was boosted from 98.03% to 98.2%. When a uniform or gaussian noise was introduced to the AC currents, PCA significantly increased the robustness of both models (PCA-MLP and PCA-SOM-NN). In 2020, health indicators of the power transmission line were extracted from three-phase voltage and current simulations by utilizing an MLP classifier (Leh et al., 2020). The simulated data contained six input vectors including three-phase voltage and current values. The ANN design for fault detection consisted of an output layer with six neurons, a hidden layer with 14 neurons, and one output layer with one neuron, where the neurons were activated through tan sigmoid and ReLU, respectively. The fault classification design had one hidden layer with six neurons and four neurons in the output layer for individual faults. Due to the confusion matrix, the model detected a fault with an accuracy of 100%. In terms of fault classification, the developed model categorized the test set into four different states with an accuracy of 70% and an RMSE of 0.44 (Leh et al., 2020). In 2020, the stator winding condition was empirically monitored with an efficiently modelled MLP (Verma et al., 2020b). The stator inter-turn fault was manually generated for data collection by taping the winding at 25%, 50%, and 75%. Three currents for individual phases were statistically processed using skewness, kurtosis, median, mean in the time domain. During the modelling phase, various performance characteristics of eleven neural network models such as the number of features, number of epoch runs, training time, activation functions, learning rate, model loss functions, and accuracy concerning each model are quantified. Only a few models could be able to classify healthy motors. The neural network with raw data scored 94.73% accuracy, whereas the neural network with statistical featured data had 98.43% (Verma et al., 2020b).

A supervised ML model's performance highly depends on the quality and quantity of the input vector. In real-world applications, labelling the dataset is a challenging and costly operation. In response to that problem, A. Caraddu et al. proposed a PdM model powered by a weakly supervised learning strategy for marine dual-fuel engines (Coraddu et al., 2021). In this study, the primary consideration was to train ML models based on their proposed weakly learning technique so that the ML models would not suffer when the number of labels is insufficient. This way, the cost of gathering labelled data would be reduced. In this experiment, the authors relied on a Digital Twin of the dual-fuel engine or on anomaly detection algorithms; they compared them against state-of-the-art fully supervised ML models (i.e., NN, Random Forest, kernel method, OC-SVM, GKNN). ML models were run in three scenarios: fully supervised performance estimation, fully supervised health status estimation and weakly supervised health status estimation. Due to the results, the proposed weakly supervised learning method maintained its robustness, even though the labelled data was significantly reduced.

3.1.2. Radial basis function neural network

The Radial Basis Function neural network (RBFNN) is from the feedforward neural network family and is used as a supervised classifier throughout the monitoring industry. The RBFNN has a similar structure to MLP with one hidden layer, which applies the radial basis function (RBF) as an activation function to produce a hidden space with higher dimensions (Alves et al., 2018; Mohammadi et al., 2017). Similar to ANN, when training, the model typically utilizes two different steps: adjusting the kernel function's parameters and optimizing the network's weights and biases (Niros & Tsekouras, 2016). After the tuning process, this network becomes ready to solve classification or regression tasks, and one can easily implement the RBFNN algorithm to fault classification or identification tasks. Fig. 3 illustrates the network structure of an RBFNN model.

In 2010, the RBFNN model was implemented as a fault detection technique to identify cracks or wear on gears by utilizing the vibration data (H. Li et al., 2009). First, during the speed-up process, the vibration signals were sampled from each healthy and faulty gear. According to their qualitative analysis, the order cepstrum had the most notable reaction to the non-stationary signal compared to the FFT, angular resampling technique, and conventional order spectrum. Thus, the order cepstrum method was selected among the remaining methods to eliminate the spectral smearing and modulation effects due to variations in shaft speed. At the first training stage, the model utilized K-means clustering to determine the Gaussian RBF parameters (width and center). In the second training stage, the weights were tuned using a cost function, and the model was tested for each gear state (healthy, cracked, worn). According to the reported results, the RBFNN's success rate for each condition was 100% (H. Li et al., 2009). In 2020, the Welch-RBFNN (W-RBFNN) model was proposed for identifying three different fault types in an induction motor's bearings (Jin et al., 2020). In order to further analyze the effectiveness of the Welch step, they visualized the data vector after each preprocessing step, which can be found in Fig. 4. The dataset was formed by collecting vibration data from a motor-driven mechanical system. Fig. 4(a) visualizes the raw vibration data with their

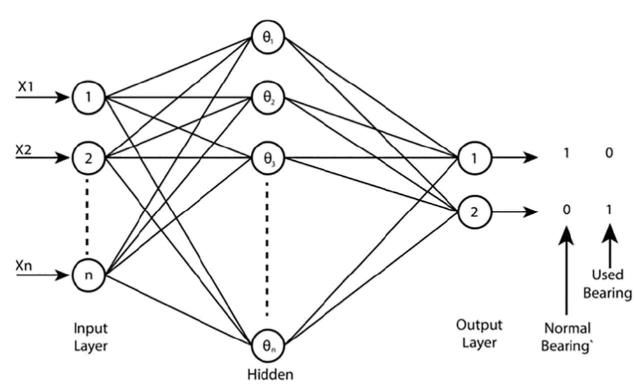

Fig. 3. A sample RBFNN architecture for bearing fault detection. ().
Adapted from Gs et al., 2011

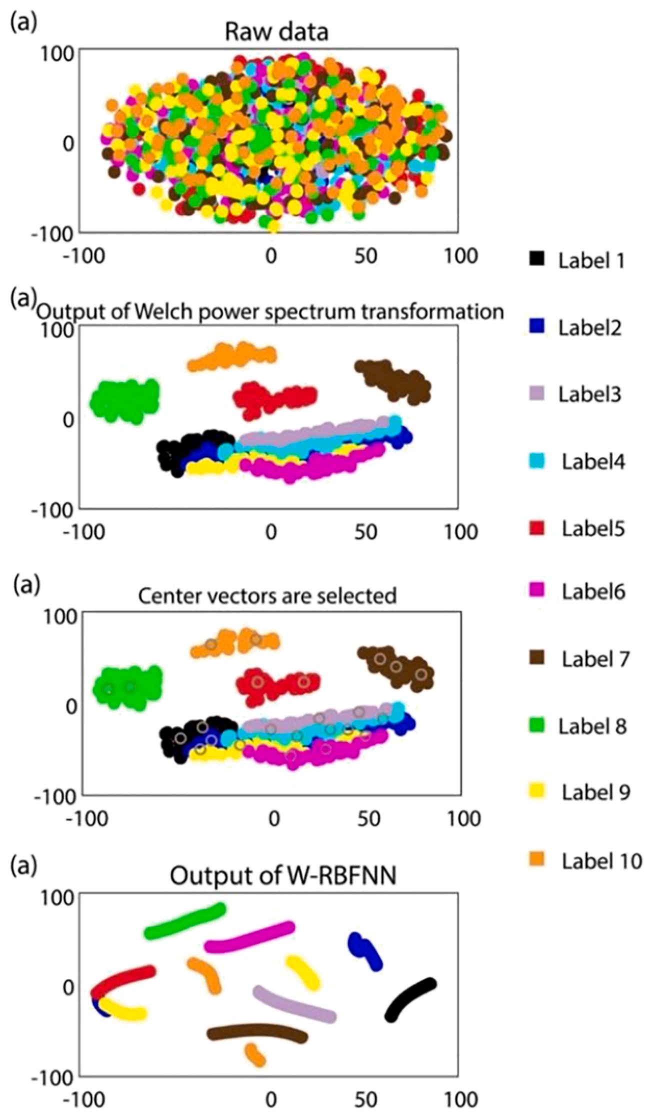

Fig. 4. Data states at each stage are visualized by t-distributed stochastic neighbor embedding. ().
Adapted from Jin et al., 2020

labels. Later, the raw data were transformed using the Welch power spectrum, where the signals were preliminarily separated after the transformation, as shown in Fig. 4(b). Fig. 4(c) displays the selected center vectors after training the model. In the last section, Fig. 4(d) visualizes the output of the W-RBFNN model after being trained. In order to boost the accuracy, Z. Jin et al. used data augmentation to analyze the performance of the proposed model against the training dataset volume. After training the model with datasets of 90 and 12,000 samples, two distinct accuracies were achieved (98.61% and 100%). Other methods MLP, SVM, and WD-CNN were compared using noisy signals to validate the robustness of the model. In the presence of high noise, the W-RBFNN's accuracy decreased to 79.5%, whereas the WD-CNN and MLP accuracy declined respectively from 99.8% to 66.95% and from 99.8% to 31.25%. The RBFNN with the Welch method (W-RBFNN) outperformed the others under different noisy environments similar to those in real-world industrial production (Jin et al., 2020).

3.1.3. Convolutional neural network

The Convolutional Network or Convolutional Neural Network (CNN) is a supervised deep learning method for processing topological datasets and was initially inspired by visual cortex (Hubel & Wiesel, 1968). CNN is an extension of ANN and contains sparse interactions, which execute using a smaller kernel size compared to the input. A standard CNN consists of convolution, pooling, and fully connected layers as presented in Fig. 5 (Datta, 2020). The convolution layer is the first stage and extracts features from an input dataset by convolving the input data with smaller-sized kernels to produce an activation map (Ajit et al., 2020). Hence, a CNN can easily point out local patterns in a dataset, unlike an MLP, which learns global patterns.

In 2020, an experimental study on the condition monitoring of hydraulic pumps using a CNN showed promising results in fault classification (Sun et al., 2020). In a laboratory, various vibration signals were captured from faulty and healthy hydraulic pumps. STFT, WT, and Wigner-Will distribution (WWD) transformed the vibration signals into time-frequency images that contained rich state information. The CNN model included the following features: two kernel sizes (3×3 and 5×5) and the mean-pool method. After comparing each model, the WT-CNN outperformed the SFT-CNN and WWD-CNN models with an accuracy of ~100% by correctly classifying three health states (Sun et al., 2020).

An alternative approach is using a CNN on a 1-D dataset known as 1-D CNN, where the operation was performed on data vectors rather than data matrices. Thus, CNN models can analyze the time-series measurements of sensors by exploiting the network's ability to learn spatial correlations. In 2016, to prove this alternative strategy, a highly accurate fault detection-based real-time condition monitoring system using the adaptive 1-D CNN was proposed by Eren et al. (Ince et al., 2016). Their objective was to prove the efficacy of a simpler CNN architecture. A three-phase induction motor's currents were monitored for the data set. Authors stated that the 1-D model did not require major pre-processing before implementation. Therefore, in the data processing stage before training, the three-phase squirrel-cage induction motor's current signals were downsampled by a factor of 8. The model's structure had three hidden convolutional layers with 2 MLP layers. As an evaluation method, multiple classification performance metrics were utilized to verify the model's applicability by comparing the performance of 1-D CNN against the major ML methods in combination with various data processing techniques such as WP-MLP, WP-RBFN, WP-SVM, FFT-MLP, FFT-RBFN, and FFT-SVM. As stated in the test results, the classification metrics were impressive (~97.2% on average), and the algorithm was vastly faster than its rival methods (Ince et al., 2016). In 2020, Mitiche et al. implemented a 1-D CNN to perform fault detection for high voltage (HV) electrical assets (Mitiche et al., 2020). The electromagnetic interference signals were acquired from real-world power stations. The signals were extracted using spectrum analysis after recording the peak and average power. The model was trained using a ten-fold cross-validation method to fine-tune the model and the Adam optimizer as the backpropagation method. Two stages were prepared for two loss functions: employing the binary cross-entropy loss to overcome the binary classification problem and using cross-entropy loss to overcome the multi-class classification problem. Results show that the 1-D CNN achieved higher accuracy with a lower computation cost compared to the previously proposed 2-D CNN models. The stage with the binary classification task reached an accuracy of 99%, precision of 99%, recall of 99%, specificity of 99%, and an F1 score of 99%. The second stage (multi-class prediction) resulted in the following accuracy (90%), precision (~91%), precision (90%), recall (89%), specificity

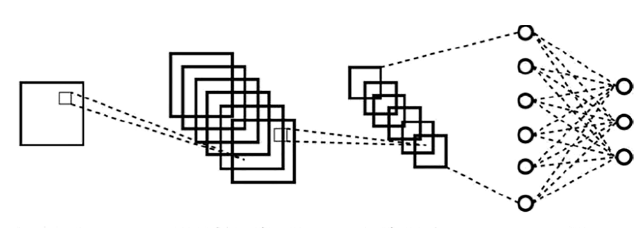

Fig. 5. A general architecture of CNN model.

(97%), and F1 score (89%). Later, the model was successfully implemented into an operating power site (Mitiche et al., 2020).

3.1.4. Autoencoder

An autoencoder is an ANN used for encoding the given data's pattern and reconstructing it with a minimum of difference. As shown in Fig. 6, the encoder structure mainly consists of three structure: the encoder stage consists of a set of linear feed-forward filters (i.e., MLP), the activation stage provides a nonlinear mapping that transforms the encoded coefficients into a range between 0 and 1, and the decoder stage reconstructs the input based on mapped values after the activation stage. Typically, an error term is assigned to measure the error caused by the middle layer. There are several strategies, such as stacked sparse autoencoder, denoising autoencoder, adversarial autoencoder, and many more for deploying autoencoders into the systems.

In a stacked sparse autoencoder (SSAE) network, the objective is to minimize the reconstruction error by using low-dimensional features. The network utilizes the initial layer to transform and compress the input signals. In 2020, Ou et al. proposed order analysis and an SSAE to monitor the state of milling tools (Ou et al., 2020). In an experimental setup, a three-axis CNC machine was used to collect spindle current signals. First, the order analysis extracted the order characteristics from the raw signals. Later, the SSAE model was constructed, which was composed of two hidden layers and a SoftMax classifying layer (output layer). Due to the results, the model classified five states with over 99% accuracy and a training time of 17 s. On the other hand, the model was compared with other conventional ML models: KNN had 84% accuracy, random forest's accuracy was 94%, and support vector machine had 91% accuracy.

If the input signals are noisy in a real-world system, then reconstructing the input data vector becomes difficult. To overcome this problem, an approach is to add noise to input vectors but measure the reconstruction loss against the denoised input. This is known as a denoising autoencoder (DAE). In 2018, a monitoring system based on a denoising autoencoder was proposed for wind turbines (Jiang et al., 2018). A wind turbine model was established to evaluate the proposed monitoring technique, and eight sensor measurements were acquired from the simulation model. The DAE model's objective was to encode manually corrupted signals, convolute the underlying nonlinear correlations, and reconstruct the original (uncorrupted) input. The proposed model consisted of two stages, the encoder, and decoder. To capture the local temporal pattern of nonlinear correlations, overlapping sliding windows (SW) were applied to the input vector. Subsequently, the SW-DAE model regarded the changes in reconstruction error as anomalies or faults. After finding the fault, the proposed model isolated the corresponding fault. Next, the most relevant variables strongly related to the fault were determined using the reconstruction-based contribution approach, where the larger contribution indicates a greater relation to the detected fault. The proposed model was tested in a noisy environment and was compared with other autoencoder models and the PCA approach. The proposed model achieved relatively higher accuracy than other models for seven of the eight faults. On average, the proposed model achieved ~91% accuracy for each fault, and the omission of the sliding window decreased the model's performance to ~83.5% (Jiang et al., 2018).

Another approach is to use an adversarial autoencoder (AAE), which utilizes adversarial loss to regularize the network (Makhzani et al., 2015). Dissimlar to a classical autoencoder structure, the decoder function acts as a generative network by learning to convert the prior distribution to the data distribution. Between the two stages, the discriminator takes the encoder's output and outputs the probability of the data generated by the encoder. During training, the autoencoder updates the encoder and decoder to minimize the reconstruction error, and the discriminative network distinguishes between the original and artificial samples. Thus, the discriminator improves the quality of generated data (Makhzani et al., 2015). In 2020, a novel fault detection method based on an AAE was proposed (Jian & Zhiyan, 2020). A chemical simulation known as the Tennessee Eastman Process was used to produce a dataset to benchmark the proposed model's performance. The proposed method included the following steps: first, the AAE model was tuned using the training dataset, then the parameters in the encoder and decoder computed the anomaly score based on the reconstruction error. Later, a threshold with a certain confidence level was introduced into the system, where an anomaly score higher than the threshold was regarded as the anomaly. The proposed AAE model and an AE model were tested to compare their performance in fault detection. According to the test run, the AAE model achieved slightly higher accuracy compared to the AE model (Jian & Zhiyan, 2020).

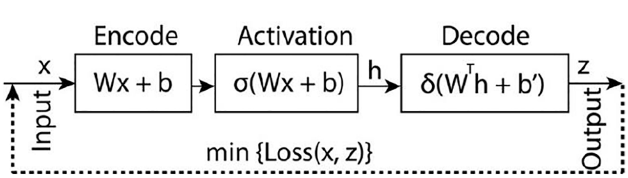

Fig. 6. A sample architecture of an autoencoder.

3.1.5. Bayesian network

The Bayesian network (BN) or Bayesian network classifier (BNC) is an unsupervised or supervised tool that presents and observes multivariate distributions based on computing the Bayesian inference probability. BN utilizes an acyclic graphical model to analyze the conditional dependencies among desired variables. The acyclic graphs are the maps between uncertain observations and certain conclusions. They are key to explaining the cause-effect relationship between corresponding variables. The BN requires parameters that describe the probabilistic relations of each variable to their parents. The model is constructed in three steps: determine the dependencies among the variables, predict the prior probability distribution, and compute the conditional probability distribution (Amin et al., 2019). Due to its relational structure, a BN model can function despite missing entries since the model knows the dependencies among all variables (Heckerman, 2008). In fault detection, they are being used to predict the response values (fault labels) of a set of observations.

In 2020, an unsupervised BNC was integrated as a data-driven approach to monitor the condition of railway catenaries (H. Wang et al., 2020). The periodic inspections of the Beijing-Guangzhou high-speed line were used to form a dataset with 12 features. The input vectors were processed by the feature extraction methods of the time-domain statistical distributions, as well as the power spectrum density and time-frequency representation domains. The BN topology was formed based on the physical relationship between the feature sets. Fig. 7 shows BN's structure as a graph, where the relationship between all events that are directly or indirectly linked to the status of the catenaries' condition (SCC). The parameters were configured using maximum likelihood estimation using the historical dataset. Later, BN

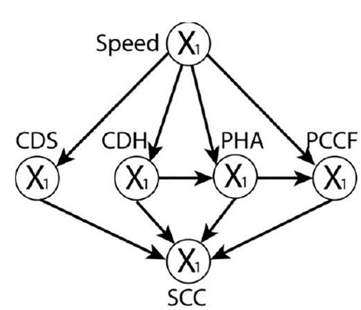

Fig. 7. Graph structure of the BN for catenary condition monitoring, adapted from (H. Wang et al., 2020).

parameters were estimated using the historical data, which ultimately mapped the probabilistic relationship between the given input and desired output. To quantify the SCC, as the last step, a key performance indicator (KPI) was obtained by analyzing the probabilistic relationship based on historical observations. Due to the results, BNC reduced the false alarm rate by up to 66.2% compared with their current practice. This approach has proven its robustness by tolerating the noisy input data (H. Wang et al., 2020). In 2016, using a similar procedure for BN topology formation, the supervised BN classifier (BNC) was used to create a probabilistic boundary to decrease the false alarm rate (He et al., 2016). The experimental data were taken from a 90-ton water-cooled centrifugal chiller of the ASHRAE project 1043-RP. In this case, based on the training sample, eight states, including seven faulty conditions and one normal state were selected. The parameters were chosen using the prior probabilities and conditional probabilities of the training samples. As a novel strategy, He et al. inserted an extra node into the BN, which presented an additional cause-effect relation based on site information (He et al., 2016). Lastly, the probabilistic boundary was implemented by introducing a tolerance variable into the system. After implementing the additional strategies noted above, the model was termed sI-PB-BNC. As reported in the paper, the false alarm rate decreased from 22.7% to 4.7% by integrating the BN into the detection system (He et al., 2016).

Although BNC is a traditional classifier, it is not solely suitable for dynamic systems. A couple of steps must be introduced to overcome the machinery's dynamic behaviour. The classical BNC has a temporal relationship known as STATIS. If the system is dynamic (e.g., a chemical process) a dynamic Bayesian network (DBN) must be introduced to monitor it. Unlike BN, this dynamic network requires an update in each time step, as shown in Fig. 8. The BN can be converted into DBN by the following three steps (Amin et al., 2019; Murphy, 2002):

1) Reshape the BN architecture based on the process dynamics.
2) Add the state of a node to describe the temporal relationships between time slices.
3) Repeat the static BN with time if all the variables exert influence on the process and update the belief of current time step.

In 2019, research was conducted on supervised DBN fault detection in a binary distillation column and a continuously stirred tank heater (CSTH) (Amin et al., 2019). A chemical operation was simulated to generate a train and test set to the DBN model. The chemical process had two main components: a binary distillation column and a continuous stirred tank heater. As a first step, the DBN's relational architecture was constructed using prior knowledge and process flow diagrams. The next step was to estimate the DBN parameters through maximum likelihood estimation, and the fault would be indicated when the novel dynamic Bayesian anomaly index (DBAI) threshold was exceeded. The DBN results were compared with PCA, PCA-T2, and BN classifiers, and the study proves that DBN outperformed all methods for achieving the lowest false alarm rate and highest detection rate (Amin et al., 2019).

3.1.6. Naïve Bayes

Naïve Bayes (NB) classifier's (NBC) structure was derived to avoid the BN's intractable complexity. Similar to BN, the occurrence of an event is predicted by analyzing past observations, but unlike BN, each feature is only dependent on the class known as "parent". In 2018, during research on detecting induction motor bearing failure, a supervised Gaussian NBC proved its robustness by achieving relatively high accuracy (Pandarakone et al., 2018a). In a laboratory, U-phase current signals were acquired from an induction motor at various rotational speeds. FFT and then Frequency Spectrum analysis was performed to extract the relevant information from the induction motor's U-phase load current. SVM and Gaussian NBC were trained to predict three different failure states: healthy (no crack), hole 0.5 mm, and hole 2 mm. The results showed that the Gaussian NBC had an 88.37% accuracy in diagnosing the motor's condition (Pandarakone et al., 2018a). In 2015, an NB classifier was proposed for end-milling, where the algorithm was designed to predict the probability of the tool wear by utilizing the posterior distribution (Karandikar et al., 2015). The end-milling force data were acquired at different spindle speeds using a cutting force dynamometer. Among the input signals, time domain mean cutting force and sum of the frequency domain amplitudes were selected based on the R^2 comparison. In an environment with multiple sensors, a BNC based on Dirichlet distribution and NBC achieved successful predictions on the posterior probabilities of tool wear states. This validated that both models were successful, even though they were computationally inexpensive (Karandikar et al., 2015).

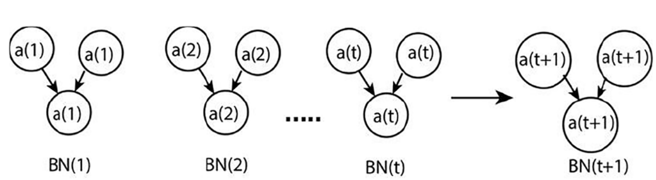

Fig. 8. An exemplary Dynamic Bayesian Network (DBN) architecture.

3.1.7. Support vector machine

Support vector machine (SVM) is a statistical ML method that is commonly applied to most classification problems. This algorithm provides a map between inputs and outputs in the training dataset as a supervised machine learning method. SVMs can process big data and manage multidomain classification industrial problems (Shan Sutharan, 2016). The classification strategy is based on constructing the best hyperplane for separating the hidden classes in the dataset. The algorithm considers data points as p-dimensional vectors. The goal is to find an optimal p + 1 dimensional hyperplane that decides the classification of each data point. For binary classification, the optimal hyperplane is created based on maximizing the distance between two categories (e.g., maximum margins). According to an experiment on milling processes conducted in 2020, SVM and the comprehensive signal analysis methods produced promising outcomes in tool health monitoring (J. Guo et al., 2020). Cutting force and vibration signals were obtained from the "prognostic data challenge 2010" database ((2010) PHM Society Conference Data Challenge, 2021). Due to the experiment, a recent statistical-based signal analysis method known as multifractal detrended fluctuation analysis (MFDFA) achieved favourably high accuracy in detecting the long-range correlation of non-stationary time series. Later, the MFDFA method and the SVM model were combined to find abnormal behaviours in the machinery signals. Due to the comparison against past studies, the proposed model was the most successful, with an accuracy of 95.6%. The authors noted that the model was more successful when the signal had long abnormal patterns.

In some cases, a multi-class classification task requires the fault type of the component to be classified. However, the SVM hyperplane provides a binary classification, yet the extension of SVM obtains promising results on multi-class classification problems. According to a survey particularly on SVM for condition monitoring, the industry generally implements three distinct strategies for SVM multi-class classification: one-against-all (OAA), one-against-one (OAO), and direct acyclic graph (DAG) (Widodo & Yang, 2007). In 2016, an SVM using the OAO strategy was applied to classify the tool wear states based on images of the machined tool surfaces (Bhat et al., 2016). In the experimental setup, the images of machined surfaces were captured after the cutting process. A statistical texture analysis method (gray-level co-occurrence technique) extracted 15 different features from the machined surface images. Next, the Fisher discriminant ratio (FDR) reduced the number of features from 15 to 4. The multiclass SVM (MSVM) model with Gaussian and polynomial kernels were integrated to classify the tool wear states into three different groups (sharp, semi-dull, and dull). The OAO approach with Max-Win voting for both kernel functions was integrated to classify the data points into three classes. The SVM with the polynomial kernel had the best accuracy (94% and 99%). In 2020, the multiclass SVM model successfully diagnosed nonlinear and high dimensional vibration signals of a solenoid pump (Akpudo & Hur, 2020). A testbed was built in a laboratory to collect vibration signals of the pumps in study. As a feature extraction step, mel frequency cepstral coefficient (MFCC) was employed to extract useful trends from the input signals, and the locally linear embedding (LLE) method was used to reduce the dimensionality. The SVM model consisted of a Gaussian kernel for nonlinear estimation, and a Lagrangian formulation for multi-class classification. The SVM model classified the test set into four states (i.e., contaminated fluid, heavy fluid, filter clogged, and normal) and achieved a test score with an F1 score of 89%, recall of 88.5%, and precision of 90%. If the outlier quantity is relatively small, then a decrease in the performance of an MSVM can be expected. In 2018, to address this problem, a new extension for MSVM was introduced and tested to diagnose induction motor faults by embedding Support Vector Data Description (SVDD) to classical multiclass SVM (OAA, OAO, DAG) (Zgarni et al., 2018). In this strategy, hyper-spheres were constructed to discriminate the training samples of each class. The experiment test set was collected by measuring the induction motor's current signals with a healthy and faulty rotor under five different load conditions. The current signals were converted into the frequency domain through the Stationary Wavelet Packet Transform (SWPT) to extract the key features. The proposed classification system (SVDD-DAG SVM) boosted the classification rate from 91% to 100%, and the F1 score rose to 100%. The average rate of classification for other MSVM methods (OAA, OAO) was improved by 5%, which proved its feasibility for all types of MSVM methods (Zgarni et al., 2018).

3.1.8. Fuzzy logic

Fuzzy logic (FL) or Fuzzy set theory is a powerful method for mapping vague inputs to a precise output using linguistic rules. As mentioned before, the real-world information is vague and partially true, which ultimately creates a fuzzy environment. FL plays a vital role in decision-making problems where the extant decision theories are not robust in a fuzzy environment. FL is essentially a derivation of four fundamental elements (Jaber, 2017a)(Marwala, 2012): the membership function (MF) associates each data element of the input space with a corresponding fuzzy membership value in the range between 0 and 1 (Aliev, 2013); fuzzy sets are a class with a continuum of grades of membership (Zadeh & Aliev, 2018); fuzzy logical operators are applied to discover new fuzzy sets from existing fuzzy sets; fuzzy rules are conditional statements that form the input and output relationship of the system which typically includes human descriptive judgments and some sample rules are given below (Cuka & Kim, 2017):

IF {signal is low} AND {error is low} THEN {tool wear is small}.

IF {signal is medium} AND {error is low} THEN {tool wear is medium}.

Finally, in order to converge to a final output, the defuzzification process is required to reduce the system output to a singular value. A rule-based system that mimics a human expert's reasoning process that involves FL is known as fuzzy inference systems (FIS) (Kolman & Margaliot, 2009). Typically, Mamdanior or Sugeno fuzzy rule systems are being widely utilized in the PdM industry (Trillas & Eciolaza, 2015). This powerful technique can be implemented in many industrial systems (e.g., microcontrollers) to enhance its capability.

In 2017, research on fuzzy logic-based online tool condition monitoring was demonstrated and implemented during end-milling operations (Cuka & Kim, 2017). The FL theory was applied to predict the cutting tool's life, and the machinery was programmed to adjust its machining speed with respect to the cutting tool's expected life. Signals from 4 different sensors were extracted respectively using FFT, frequency band analysis, and statistical metrics (Peak-to-peak amplitude, RMSE, power of the signals). The cutting tool's health was divided into four fuzzy sets (small wear, medium wear, accelerated wear, breakage) based on image analysis of the tool's taken picture. The triangle method was selected as a membership function due to its cost-effective computational outcomes. Later, fuzz rules were formed and implemented according to expert knowledge. Finally, the centroid method outputted a single FIS value that determines the classification result. Due to a visual inspection of the cutting tool, the experiment showed that the FL model was 100% accurate (Cuka & Kim, 2017). In 2020, a condition monitoring method for wind turbine condition monitoring based on FIS with an assembled multidimensional membership function was proposed by Furning et al. (Qu et al., 2020). Different than conventional methods, the authors used a nD membership function instead of 2D. The experimental data were collected from a group of wind turbines in China. As a data preparation step, invalid and contaminated data were removed manually. Before constructing the FIS model, the future trend of the wind speed was predicted through regression analysis. Then, based on the regression prediction, the combination of the cubic spline fitting and the 2D membership functions formed the third dimension. The proposed model monitored the wind speeds of the wind turbines and identified anomaly data segments in the trend. After multiple experiments, the proposed method had an earlier anomaly detection capability and fewer false alarms than the classical 2D membership function. In 2017, the authors proposed a fuzzy multi-parametric expert system for diagnosing power systems (Žarković & Stojković, 2017). The experiment examined the performance of two FL-based methods: Mamdani-type and Sugeno-type, by introducing numerous power transform inputs: age, frequency response analysis, the overheating temperature of the hot spot, polarization index, dissolved gas-in-oil analysis, the temperature of isolation, and tg δ. Both models had triangular and trapezoidal membership functions with 27 different fuzzy rules. They were evaluated using a health index, a transformer status indicator, and experiential input from various conferences. Due to the reports, the result for all FLC models achieved similar performance outcomes; however, a higher correlation between the transformer status indexes and the FL outcome was achieved by a Sugeno-type rule system (Žarković & Stojković, 2017). In 2019, an FLC was applied to identify the faults in an induction motor using the stator currents' amplitude (Agyare et al., 2019). The experiment used a simulation dataset to test and train the proposed algorithm, in which an induction motor model was used to acquire signals. RMS values of the stator currents were the test set for the fuzzy controller. Fuzzy rules and logic were developed with the help of expert knowledge. The inputs (low, normal, high, very high, too high current) and the outputs (open phase, damaged, critically damaged, seriously damaged, healthy) were fuzzified into five trapezoidal membership functions. Due to the report of the experiment, the fuzzy logic controller was able to detect the various faults of the 3-phase induction motor (Agyare et al., 2019). In 2020, a novel predictive maintenance system for catenary systems was proposed using a Mamdani-based supervised FC in railway systems (Karaduman & Akin, 2020). A camera and a temperature sensor collected the railway system's input vectors and the temperature signals' noises. Later, the CWD filtered the correlation coefficients (CC) as a preprocessing step by utilizing the captured pantograph images. Mean values of two features were introduced to triangular membership functions with three fuzzy rules. During the training phase, the mean correlation coefficient (MCC) of the training features determined the fuzzy membership functions. After various examinations, the FC model results were compared with the true values. As a fault detector, the fuzzy classifier (93.9%) achieved the highest accuracy compared to an SVM (80.3%) and ANN (76.5%) (Karaduman & Akin, 2020).

The neuro-fuzzy system (NFS) is the method that combines the fuzzy sets and logic with the neural network. In 2020, the NFS and FIS for condition monitoring of induction motors were implemented and compared (Verma et al., 2020a). From an experimental setup consisting of healthy and faulty induction motors, three-phase stator current signals were acquired in the time domain. The FIS model was designed with predetermined rules and membership functions. For the NFS-based model, the membership function type and hyperparameters of the neural network were defined before training the model. The following procedure constructed an adaptive neuro-fuzzy inference system (ANFIS) architecture: the first layer takes three current values (inputs) and computes the corresponding membership functions, constant values are generated for the combination of inputs, and the aggregate output is obtained using all the rules for a given set of inputs, the final outcome is the weighted mean of all the aggregated outcomes. After testing both models, the ANFIS model achieved an accuracy of 93.3%, whereas the Fuzzy logic had an accuracy of 86.3% (Verma et al., 2020a).

3.2. Cluster analysis

Cluster analysis pursues the same objectives as classification analysis. Due to this reason, they are evaluated with the classification performance metrics; however, they do not share the same approach. The objective of clustering is to divide the data points into several groups where similar data points share the same cluster. Clustering-based models are used in supervised or unsupervised learning. According to a recent study on unsupervised methods in condition monitoring, some of the most common clustering techniques are K-means, Fuzzy C-means (FCM), and agglomerative hierarchical clustering (Amruthnath, 2018).

3.2.1. K-means clustering

K-means clustering is an unsupervised recursive method that aims to classify *n* observations into *k* clusters where the new data belongs to the nearest cluster center, and the algorithm updates the cluster centers after each iteration. In 2010, K-means clustering was tested in various scenarios to validate its fault detection and identification performance on rolling bearings (Yiakopoulos et al., 2011). The vibration signals were collected from three industrial installations and one laboratory test case. The real success of this approach was due to the authors including well-chosen frequency-domain parameters (kurtosis, skewness, variance, RMS, FT, WT, and order spectrum), signal envelope (highest and lowest values of the signal), and expert knowledge was employed for interpreting the signal behaviours and underlying causes. They used the system's mathematical dynamic model to distinguish the data. Later, the researchers computed the pre-clusters' initial centroid locations. Each cluster represented a potential bearing fault. Four different distance measurements (all the above metrics and Cosine metric distance) were utilized to compute the distance of new data points. The test results confirmed that the K-means clustering model based on correlation distance achieved a classification rate of 100% on all test sets (Yiakopoulos et al., 2011).

3.2.2. Fuzzy C-means

Fuzzy C-means is an unsupervised dynamic clustering method. The algorithm has a structure similar to K-means clustering; however, it uses a fuzzy membership to assign a degree of membership for every cluster (Nayak et al., 2015). The chosen memberships quantify the strength of the relation between the data values and centroids. In 2017, a study designed a successful fault detection approach for grounding distribution systems using the FCM clustering method (M. Guo & Yang, 2017). The dataset was generated using a simulation of a grounding mathematical distribution model. In the proposed approach, the dataset was processed by applying two feature extraction methods: the Hilbert-Huang transform (HHT) band-pass filter and wave transformation. The HHT band-pass filter reconstructed the instantaneous frequencies that represent the original signals to improve the signals' comparability. The time-frequency matrix was created using an HHT band-pass filter. The polarity distribution matrix (PDM) was constructed from the time-frequency matrix. As the final step of the feature extraction process, the amplitude-polarity feature matrix (APFM) was formed by merging both matrices to assess the characteristic parameters of the intermittent zero-sequence current fault in different feeders. The SVD method transformed the extracted data into a compact form. The FCM method is applied to the normalized singular values to detect the fault feeder without a certain threshold setting. According to four different earth fault cases, the proposed method correctly identified the fault conditions and factors. In 2019, The FCM clustering method was implemented to identify anomalies in a continuous distillation system (Azzaoui et al., 2019). This distillation system has two main modes: the normal mode, where all the system's parameters follow a uniform pattern, and the abnormal mode, in which these parameters deviate during the process. Six indicators of the real-world distillation signals were converted into the time-frequency domain using the stationary WT. The membership matrix and the cluster numbers were initialized and executed by using the time-frequency domain data. According to the model's outcome, the SWT-FCM achieved a classification rate of 74.17%, a sensitivity of 100%, and a specificity of 48%.

3.2.3. Hierarchical clustering

Unlike FCM and K-means, agglomerative hierarchical clustering starts by assigning each data element to an individual cluster. Then, moving up the hierarchy, similar clusters are combined (Murtagh & Contreras, 2011). Initially, there are as many clusters as possible. The measurement methods decide the similarities between the data points. Finally, from the dendrogram graph, the optimal number of clusters is chosen based on the distance between the clusters. In 2015, the hierarchical clustering method was proposed to implement a monitoring system for power transformers (Babnik et al., 2008). The power transformers' radiometric signals were acquired using a helical antenna in a laboratory setup. Each captured data vector was transformed into a frequency domain using FT. Later, the PCA algorithm reduced the dimensions from (5000x2500) to (5000x6), where the first six components contained enough information to represent the dataset. In order to achieve the performance, various hierarchical clustering models are constructed from various distance metrics and linkage methods. Afterward, the data points inside each cluster were qualitatively examined for each hierarchical clustering model. In the experiment analyses, they reported the effect of the variation in distance and linkage method on the final clusters. The average linkage algorithm caused single records to join a larger cluster each time. Ward's linkage method tends to create clusters of similar size. Lastly, a variation in distance metrics did not affect the final results.

3.2.4. Support vector machine

The SVM concept can be adapted to the clustering analysis as well. One-class SVM (OCSVM) constructs a spherical decision boundary in an unsupervised manner to solve one-class classification problems using a specific optimization formula. In 2018, a fault detection system based on OCSVM was applied to a closed-loop system (Z. Li & Li, 2018). A simulation dataset representing three tanks with five parameters was generated and normalized. The OCSVM model with RBF kernel function was trained, and a 95% confidence interval was set as a threshold. The threshold was recalculated for all new data. According to the test results, the OCSVM's effective fault detection rate was ~97% for eight different faults (Z. Li & Li, 2018). However, the condition monitoring industry seeks more robust and accurate performance of SVMs by utilizing additional developments with the available approaches. In 2016, a novel approach using OCSVMs was proposed to resolve performance degradation in classic OCSVMs (Xiao et al., 2016). The strategy is to modify the training set to remove consistent outliers to prevent them from becoming support vectors. The vnuOCSVM model was proposed to overcome this problem, which identifies the potential outliers and remove them from the training. In this way, the decision boundary becomes more robust against outliers. During the training stage, the training samples were normalized to the zero mean and unit standard deviation. Later, the following steps were applied to train the proposed method: the samples located outside the boundaries were regarded as "suspected outliers," and the OCSVM model was retrained without the suspected outliers to compute the final hyperplane, known as "cluster core." Thus, the model observed the distribution of the target class without being affected by the outliers. The hyperplane for vnuOCSVM was farther from the outliers, while the remaining models misclassified some outliers. Datasets from a realistic chemical process (i.e., Tennessee Eastman Process) were used to benchmark the proposed model's fault detection performance against the classical OCSVM models (etaOCSVM, wOCSVM, OCSVM). According to the results, the vnuOCSVM had the lowest false alarm rate, higher fault detection rate, and a larger margin for potential outliers.

3.3. Regression analysis

Regression is the task of estimating continuous output variables based on given observations. This approach can identify the anomaly data by analyzing fault-free patterns where deviations are regarded as abnormalities. Another application is the estimation of the remaining useful life (RUL), where the machinery's lifespan is predicted through its indicators.

3.3.1. Recurrent neural network

The recurrent neural network (RNN) is a class of ANN that is practical for recognizing patterns in sequential data. The RNN structure (Fig. 9) contains multiple feed-forward neural networks that transmit information to the following node, known as the *hidden state*. Differently expressed, the output of the current node is the harmonization of the prior knowledge (hidden state) with its own experience, which thereafter is passed to the next node. Therefore, RNNs use the previous output to determine the next output. However, the algorithm is computationally expensive since the network has two inputs, and the steps are not parallel (Wani et al., 2021). Furthermore, RNNs suffer from two well-documented problems: "gradient vanishing" and "exploding gradients" (Sherstinsky, 2020). Empirical studies show that the RNN with a simple repeating structure (vanilla RNN) cannot learn long-term dependencies if the number of time steps exceeds 10 (Hochreiter & Schmidhuber, 1997). Due to these restrictions, the industry started using RNNs as a feature or feature selection method (Barron et al., 2008; Gugulothu et al., 2018; Huang et al., 2019). Long short-term memory (LSTM) is a novel recurrent network architecture derived from an RNN combined with an optimum gradient-based learning technique (Hochreiter & Schmidhuber, 1997).

LSTM's complicated cell structure with four gates provides long-range dependencies between the elements (Calin, 2020). The network architecture interconnects these memory cells in a cascaded form, as shown in Fig. 10. LSTM is applied when RNNs cannot meet industry requirements. Because of its outstanding success in using sequential inputs, the predictive maintenance industry applies this technique as a regression-based anomaly detection method. In 2018, a novel fault prognosis method using LSTM based on vibration signal of rotating machinery was presented (Xie & Zhang, 2018). The model was trained to predict the performance of the given electrical motor. During the experiment, vibration data were collected using an accelerometer, and the unprocessed data were employed in three other models: LSTM, SVM, and echo state network (ESN). Among all methods, LSTM had the lowest

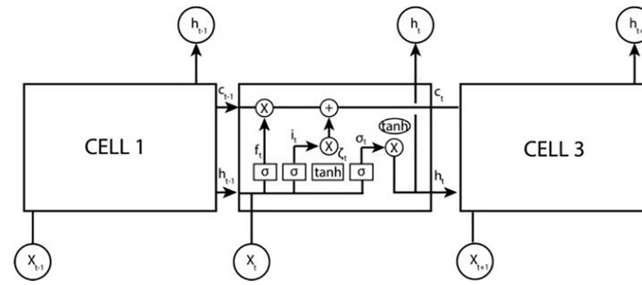

Fig. 10. An exemplary LSTM network structure, where the memory cells are linked in a cascaded form (x and h are the input and the output, respectively).

RMSE value (0.018). The model has also proven its robustness by achieving successful results under different conditions. In 2019, Fabrizio showed the effect of hyperparameter tuning of LSTM on the RMSE value, which then resulted in promising outcomes in the estimation of remaining useful life (Bruneo & De Vita, 2019). Differently from grid search analysis, they trained several models equal to the Cartesian product's cardinality, which can be computed by multiplying the single cardinalities of each set. Thus, they gained an insight into which hyperparameter affects the result most. A dataset consisting of the histories of 100 NASA engines was used to train and test an SVM, a deep neural network (DNN), and LSTM. The training set was normalized between −1 and 1, and the unrelated features were later removed from the dataset. The models were trained to predict the RUL value of the given battery. The hyperparameters were tuned as a function of the model's RMSE. Due to the outcomes, fine-tuned LSTM outperformed the other models with a relatively low RMSE value (11.42) (Bruneo & De Vita, 2019).

In 2020, a study on fault detection in rotating machinery using DL methods, LSTM proved its effectiveness by showing accurate diagnosis under various rotation speeds (Lee et al., 2020). In this research, a motor testbed was used to capture the acceleration signals of the system. RMS and Kurtosis processed the raw data, and the processed data was transformed into the time-frequency domain through continuous wavelet transformation (CWT). Each model was trained using the Adam optimizer, and a grid search method was used to choose each model parameter. The researchers built the structure for both the CWT with CNN (CWT-CNN) and time-series LSTM (TS-LSTM) models. At a constant RPM setting, the CWT-CNN model and TS-LSTM had an accuracy of nearly 100%. Unfortunately, both models suffered from changes in the equipment's speed. As a solution, before processing the raw data, the time-series signals were scaled and smoothed, which included the following steps: scaling the time-domain data, transforming the signals into the frequency domain using DFT, filtering out the high-frequency components utilizing low pass filter, and removing redundant amplitudes, converting back to the time-domain using inverse DTF. At a varying RPM setting, a third, scaled and smoothed TS-LSTM, was tested with unseen RPM settings along with the previous models. According to the results, CWT-CNN had an accuracy of 83.2%, while the scaled and smoothed TS-LSTM produced a similar accuracy with less variation in performance (Lee et al., 2020).

However, LSTMs are not only restricted to regression tasks, and the network can perform a many-to-one classification by changing the activation function on the output layer. In 2018, an alternative approach was implemented using LSTM for fault detection in a simulated chemical process (Xavier & De Seixas, 2018). A realistic simulation of an industrial plant, the Tennessee Eastman process was used as a benchmark for this study. Instead of using LSTM as a regression-based forecasting method, the authors turned the fault detection model into a many (each sequence) to one (unique fault) classification problem. Hence, the model no longer needed any detection threshold. The comparison between LSTM and 14 statistical-based data-driven approaches confirmed that the novel model outperformed 13 of the other methods with an accuracy

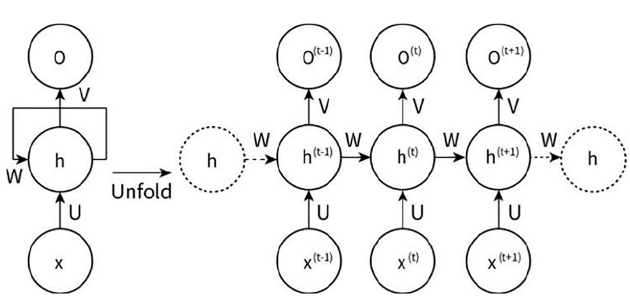

Fig. 9. An exemplary structure of RNN.

of 97% and used significantly lower parameters (Xavier & De Seixas, 2018).

3.3.2. Support vector regression

Apart from classification problems, the SVM concept can be used as the regression algorithm known as support vector regression (SVR). The transition from SVM to SVR is accomplished by introducing an ε-insensitive region around the function known as the ε-tube. Thus, in a supervised manner, an optimization problem is reformulated along the tube region to find the best approximation of the continuous-valued function where the flattest tube that contains most of the training instances represents the best approximation (F. Zhang & O'Donnell, 2019). In 2014, an SVR-based PdM model was proposed to monitor performance degradation in vessel propulsion (Coraddu et al., 2014). A realistic simulator of a naval propulsion system is utilized to generate a naval vessel propulsion dataset for the SVR model. Due to expert analysis, the indicator of the asset degradation was the following signals: ship speed, compressor decay, and gas turbine decay. In preprocessing step, ship speed was expressed with probability density functions, and the remaining features were selected using expert knowledge. For the forecasting task, two well-known ML models were employed: SVR and regularized least squares (RLS). Both models were tuned using the grid search analysis and the k-fold cross-validation method. Later, in order to assess the models' learning capability, they were trained with different sample sizes (10, 25, 50, 100, 250, 500). Due to the results, SVR achieved a slightly higher MSE value in every case; it gained a low error rate after being trained with 100 or more data points. Therefore, the authors proved that their model could support monitoring operations in the maritime domain. In 2018, ε-SVR was constructed to estimate the rotor power curve of a wind turbine through an operational wind turbine's historical data set (Pandit and Infield, 2018). Later, a well-known conventional benchmark (i.e., the binning method) modelled the rotor power curve for comparative analysis. Rotor speed is one of the key performance indicators for wind turbines and represents the relationship between power production and the hub-height wind speed. The signals from real-world wind turbines formed the raw dataset. The corrected wind speed was obtained using an air density correction formula to construct the rotor curves. The results of both methods were compared with the actual measurements. According to the graphical analysis, SVM had a better fit and a faster response compared to the binned method (Pandit and Infield, 2018). In 2017, the new approach of SVR combined with a feature extraction method (MFCC) was introduced for tool-wear condition monitoring based on the RUL (Benkedjouh et al., 2017). Raw time-domain force signals of six different flute cutters were converted to cepstral representation by following the MFCC steps: converting the data into the frequency domain using DFT, computing the Mel-frequency spectrum and logging the power at each of the Mel-frequencies, taking the discrete cosine transform, and removing the dependencies using cepstral coefficients. The MFCC-SVM estimated the wear curves to assess the RUL of the worn tool and achieved R^2 of 97% and RMSE of 0.0398 on average for six cutting tools (Benkedjouh et al., 2017). In 2019, SVR was implemented as a monitoring framework to detect the abnormal conditions of coal mills in thermal power plants (Hong et al., 2019). The support vectors detected the deviations in corresponding performance indicators based on their training set experiences. The parameters were derived from the grid search strategy, and the abnormality threshold (control limit) was chosen based on the Pauta criterion and the Gaussian kernel-density estimation method. The proposed method was applied to three real failure cases at coal mills (the separator failure, the fluctuation of the outlet temperature, and the co-firing cases). Euclidean distance was used to evaluate the similarities between the data points, and the accumulated anomalies were collected for the root causes related variables identification. For every case, the SVR results highly matched the expert's conclusions, and the average performance metrics of the model indicated promising outcomes: a TPR of 95.2%, an FPR of 9.1%, and an MSE of 1.35 (Hong et al., 2019). In 2020, a novel approach with SVM was proposed mainly to estimate the degradation trend in noisy data. The novel approach proved its effectiveness and robustness by producing accurate outcomes in RUL estimation on the NASA Li-ion battery dataset (Ben Ali et al., 2020). The novel approach's structure consists of two main components: an enhanced SVR or incremental support vector regression (ISVR) and quantum-behaved particle swarm optimization (QPSO). ISVR employs the same principles as SVR; however, it follows different steps to perform the regression task. As an optimizer, the QPSO method determines the hyper-parameters of the ISVM. Finally, the QPSO-ISVM and other standard regression methods (Linear regression and Polynomial regression) were trained and tested to compare the proposed method's capabilities. The models were trained to predict the degradation in the batteries' capacities. During validation, the capacity degradation of the two real Li-ion batteries was compared with each model's outcome. Based on MAPE and RMSE evaluation, QPSO-ISVM outperformed the other methods with a significantly low RMSE (0.0202 and 0.0255) and MAPE (0.823% and 4.2%) on both batteries.

3.4. Ensemble learning

Ensemble learning combines several machine learning algorithms (supervised or unsupervised) to reduce overall variance by fusing their outputs (Thomas Rincy & Gupta, 2020). The main objective of ensemble learning is to compensate for each other's weaknesses, ultimately improving accuracy. For this purpose, various ensemble methods (i.e., Bagging, Stacking, AdaBoost) were established (Zhou, 2012).

3.4.1. Stacking

The stacking method refers to the combination of heterogeneous weak learners (learns in parallel), where weak learners' predictions are used to train the meta learner. In 2019, a stacking ensemble method combining an SVM and random forest was proposed for monitoring the hydraulic systems (P. Guo et al., 2019). First, the feature extraction (mean value, kurtosis, skewness, etc.) methods transformed 15 different time-domain signals (temperature, pressure, flow, etc.) into a useful form while PCA reduced the dimension of the input vectors. The feature set was selected through Pearson correlation coefficients. Constructing the proposed stacking method included the following steps: the dataset was divided into k number of partitions, a fixed kernel SVM and k-fold cross-validation were used to obtain and test the k classifiers, the predicted labels were combined into a single array to be used as a training set for the second layer, and the mean values of the k original test set's predicted values were passed into the meta learner (random forest model) with multiple decision trees. For a relative comparison, the classical methods (ANN, LDA) and the proposed model were tested by introducing the same dataset. According to the experimental results, the proposed stacking model outperformed the conventional methods with an identification accuracy of 88.6% in fault detection for different components (cooler, valve, pump, accumulator) of the hydraulic system (P. Guo et al., 2019). In 2011, a stacking ensemble method consisting of EKF and FIS was proposed to perform unsupervised fault detection and diagnosis on industrial gas turbines (Salar et al., 2011). The gas turbine system consisted of multiple stages: a frontal compressor, frontal turbine, rear compressor, and rear turbine. Four signals were acquired from the hydraulic system: the turbine exit temperature, fuel flow, compressor outlet pressure, and compressor outlet temperature. The flow capacity signals were passed into the EKF method to estimate the health parameters of each stage. The EKF method successfully estimated the flow capacity for each state based on the state-space model of the system. Second, based on the EKF prediction, the FIS analyzed the deviations in the system to classify and locate (rear or front) the system's physical fault by utilizing predetermined fuzzy rules. The EKF-FIS model was verified through data from simulations of the actual defects in various fault scenarios; it showed a high correct classification rate of 96.15% (Salar et al., 2011).

3.4.2. Bagging

The bagging or bootstrap method trains the learners in parallel, and the outputs are fused using a deterministic averaging process (Kramer, 2013). The random forest (RF) algorithm is an ensemble learning method that constructs a forest of decision trees in parallel. In 2020, in the assessment of diesel engine conditions, ensemble learning based on bagging extension (random forest method) achieved higher accuracy than shallow classifiers (Shao et al., 2020). A mathematical model simulated a large marine diesel engine under various engine loads with 15 distinct features. A decision tree model was constructed based on the control statements. In the RF structure, each internal node represented a feature, each branch represented the test's output, and each leaf node represented a class label for classification. The random forest classifier considered the category with the most votes as the output result. The testing and training dataset contained 15 features with 100 sets of data for each working condition. According to the paper, in comparison with singular classifiers (KNN and SVM), this approach outperformed the other methods with an accuracy of 95% by classifying the faults into five states (Shao et al., 2020). Again in 2020, apart from classical approaches, a combination of the time-domain analysis, frequency-domain analysis, and variational mode decomposition (VMD) was introduced to the random forest (RF) ensemble method for diagnosing tool wear (Yuan et al., 2020). The spindle motor current was monitored during a real-time cutting operation under various spindle speeds, cutting depths, and feed rates. The degree of tool wear was optically measured using a microscope. Three distinct models (RF, SVM, RBFANN) were tested under ten different operations, and the average recognition accuracy of the various conditions was the evaluation criteria. According to the outcome, the RF model outperformed the other models in all test conditions with an average recognition accuracy of 95% (Yuan et al., 2020). In 2013, a bagging model composed of support vector data description (SVDD) methods was proposed for batch process monitoring (Ge & Song, 2013). Based on 17 signals, the models were implemented in a case study of an industrial semiconductor etching process. The training data was processed and partitioned into sub-datasets for each sub-model. As a first strategy, a voting-based approach decided the final evaluation results by voting on each monitoring result. The objective was to establish a statistics-based binary rule and count the individual models' violations based on each sub-model output. As a second strategy, each monitoring result was probabilistically combined using the Bayesian fusion strategy. The fault detection rate of both ensemble SVDDs was compared to the classic SVDD, and the performance of the SVDD was enhanced by incorporating bagging ensemble methods (Ge & Song, 2013).

3.4.3. Boosting

Unlike Bagging, the Boosting method compensates for the weakness of the ensembles by following a sequential rather than a parallel process (Dong et al., 2020). In 2020, a gradient boosting tree was proposed for fault diagnosis in low voltage smart distribution grids (Sapountzoglou et al., 2020). The boosting tree consisted of multiple classical trees trained with identical training sets; however, each tree specialized in a particular characteristic of the input-output relation. The boosting tree used the following structure in sequence: the first tree determined a decision boundary based on a voltage value; the second tree corrected the misclassified samples of the first tree and determined another boundary based on the current signals; the third tree corrected the first and second trees' estimations. Thus, the model's prediction was based on the serial combination of the three trees. During the experiment, data from Portugal's semi-rural LV distribution grid consisting of 230,688 data points were utilized as the dataset. The distribution grid consisted of various elements (conductor, resistances, reactances) to connect the nodes. Voltage and current sensors monitored each node's corresponding values. Later, the authors split the dataset into three partitions: the training set (learning the pattern), the validation set (optimal hyper-parameter adjustment), and the testing set (performance evaluation). The fault detection model caught the occurrence of faulty current phases with an accuracy of 100%. However, the variation in fault resistance affected the accuracy of classification and location. The fault types were successfully assigned to four categories with an accuracy of 98% for low faulty resistance and 86.7% for high faulty resistance. The model identified the defective branch among the others with an accuracy that varied from 95.8% (low resistance) to 84.1% (high resistance). In 2020, an ensemble-boosted regression trees (EBRT) model predicted potential hazards by monitoring the fire-resistant hydraulic fluids contamination levels in coal mines (Uma Maheswari et al., 2020). A dataset consisting of 73 hydraulic variants from real-world underground coal mines was used to train and test the proposed model. The proposed model used the gradient boosting technique; it consisted of multiple regression tree models where the tree-shaped structure outputs a continuous value by iteratively portioning the dataset into smaller groups. In the proposed EBRT model, each regression model was selected sequentially, and the dataset was divided into simple trees where each tree learned individually. The weights were respectively updated, and the results were aggregated to agree on the final output. The proposed model achieved a 31.68 RMSE in estimating hydraulic fluid values. Therefore, by using this model, the potential risk of imminent hazards can be detected in advance.

4. Discussion

In the previous section, we reviewed various ML-driven PdM methods without giving our insights. In this section, we highlighted each model's limitations and strengths and provided our underlying findings of a successful condition monitoring model. For a general analysis, the results of the empirically tested methods were categorized and listed in Tables 1-5 based on the type of ML model involved in the condition monitoring system. Tables 1 and 2 list techniques for the classification-based methods, Table 3 summarizes the clustering-based methods, Table 4 represents the regression-based methods, and Table 5 displays the ensemble-based condition monitoring methods. Each column represents a feature of the summarized paper: date, machinery in the study, signals acquired from the machinery, dataset information for the model, processing techniques, results of the test run, and a summary of the research. We must stress that comparing individual performances between models from different cases is not a proper evaluation strategy, since each case does not share the same complexity. As previously discussed, designing real-world problems in a mathematical model is a complicated task in most cases, and many more recent scholars support this statement. For instance, according to the study on RUL estimation of Li-ion batteries, model-based approaches suffer from a deficiency of adequate information, and they were unable to estimate the true states of the given system (Ben Ali et al., 2020). In such cases, data-driven strategies are more feasible, practical, and effective. Due to this reason, data-driven approaches are more frequently used in the field.

As mentioned, ML-driven condition monitoring systems require feature engineering processes to become a robust model (Section 2). The main goal of the data processing step is to extract faulty (anomaly) and healthy patterns in a way that the ML model can easily learn their characteristics. Due to our interpretations based on Tables 1-5, we found that all researchers utilized at least one preprocessing method in their proposed models, and many scholars stressed the importance of this crucial process. For example, according to the FMC model's results, the WT increased the model's performance by 11% in classification rate and 47% in sensitivity (Azzaoui et al., 2019). So we know that a collection of high-quality input vectors must always be present for any type of monitoring system to achieve good performance. Unfortunately, sensor signals may have accuracy degradation, drift deviation, and missing values. Thus, interpretation and computation are required to obtain a high-quality dataset for the learning algorithm. The first option can be using a more straightforward method such as interpolation. Due to poor vibration and acoustic-emission sensor measurements, Benkedjouh et al. determined the maximum wear curves by interpolating the integer

Table 1
Summary of reviewed Literature On Classification-Based Methods For Condition Monitoring.

Year	Machinery	Acquired signals & Dataset	Data Processing	ML model	Result	Summary
2020	Railway Catenaries	12 different signals (1.546×10^6 samples)	Kurtosis, RMS, wavelet entropy, power spectrum density, FT Hilbert spectrum	BN (unsupervised)	80.3% accuracy in fault detection.	A BN approach for condition monitoring of high-speed railway catenaries (H. Wang et al., 2020).
2016	Chiller	6 temperature signals, (22400 samples)	Manual feature selection	sI-PB-BN (supervised)	92.5% accuracy in fault classification (7 cases)	Fault diagnosis of chiller using sI-PB-BN (He et al., 2016).
2019	Distillation column	6 signals (500 samples)	Unsp.[1]	DBN (supervised)	93% accuracy in fault detection	Fault detection and pathway analysis using DBN (Amin et al., 2019).
2018	Induction motor	U-phase 3 load currents and voltages (320 samples)	FFT	NB (supervised)	88.3% accuracy in fault classification (3 cases)	Detecting induction motor bearing failure using NBC (Pandarakone, Gunasekaran, et al., 2018).
2019	Hydraulic Brake system	Vibration data (550 samples)	Mean, STD, variance, kurtosis, median, mode, skewness, sum, standard error, max, min, range, RMS, shape factor	NB (supervised)	89.23% accuracy in fault detection	Real-time condition monitoring of hydraulic system using NB and BN (T M et al., 2019).
2020	Milling tool	Cutting force and vibration signals (270 samples)	MFDFA	SVM (supervised)	95.6% accuracy in fault detection.	Tool condition monitoring in a milling process using MFDFA and SVM (J. Guo et al., 2020).
2016	Cutting tool	Machined surface images (250×250)	GLCM, FDR	MSVM (supervised)	96.75% accuracy in fault classification (3 classes)	Tool condition monitoring by SVM classification of machine surface images in turning (Bhat et al., 2016).
2020	Solenoid pump	Vibration signals (1000 samples)	MFCC	MSVM (supervised)	92.5% accuracy in fault classification (4 cases)	Solenoid pump fault detection based on MFCC, LLE and SVM (Akpudo & Hur, 2020).
2018	Induction motor	Three current signals (225 samples)	SWPT, RMS	SVDD-DAG SVM (supervised)	100% accuracy in classification (3 cases)	SVDD-DAG-SVM for induction motor condition monitoring (Zgarni et al., 2018).
2017	End milling	Cutting force, current, vibration, machining signals (unsp.[1] samples)	FFT, frequency band analysis, Peak-to-peak amplitude, RMSE, power of the signals	FIS (unsupervised)	Quantitative evaluation through graphs.	Fuzzy logic-based tool condition monitoring of end-milling (Cuka & Kim, 2017).
2020	Wind turbine	Cut-in and cut-out wind speed (unsp.[1] samples)	Unsp.[1]	FIS (unsupervised)	Graph-based comparison with true signals.	Wind turbine condition monitoring based on multidimensional membership function using FIS (Qu et al., 2020).
2017	Power transformer	6 indicators (unsp.[1] samples)	Frequency response analysis	FLC (unsupervised)	Quantitative evaluation through graphs.	Condition monitoring based on FC for power transformers (Žarković & Stojković, 2017).
2019	Induction motor	Three-phase stator currents (unsp.[1] samples)	RMS	FLC (unsupervised)	Graph-based comparison with true signals.	FLC-based condition monitoring for a three-phase induction motor (Agyare et al., 2019).
2020	Pantograph-catenary systems	Image (unsp.[1] samples)	CWD, correlation coefficients	FC (supervised)	93.9% accuracy in fault classification (3 cases).	An approach based on predictive maintenance using the FC in pantograph-catenary systems (Karaduman & Akin, 2020).

[1] Unspecified information.

values between 66 and 165 (Benkedjouh et al., 2017). In some cases, the number of faulty machinery can be way less than healthy ones. Thus, data augmentation methods are the cheapest for such cases.

As proposed by Kerdprasop et al., in the event of faulty data being scarce, the number of failure cases can be duplicated equally as the healthy cases. Thus, the model has a higher chance of analyzing the faulty cases, which ultimately lowers the false alarm rate and increases the detection rate. If the system contains too many signals, another simple solution is to select a criterion (e.g., %50 or less missing data) to eliminate misleading signals and then conduct a manual selection.

Alternatively, combining a data-driven and model-based approach may lead to greater success if the system behaviour is known. Also, if an

Table 2
Summary of Reviewed Literature On Classification-Based Deep Learning Methods For Condition Monitoring.

Year	Machinery	Acquired signals & Dataset	Data Processing	ML model	Result	Summary
2011	Rolling element Bearing	raw vibration data (10×4800 samples)	RMS, variance, mean, kurtosis, skewness	MLP (supervised)	100% accuracy in fault detection.	ANN based condition monitoring of rolling element bearing (Vijay et al., 2011).
2010	Induction motor (stator winding)	Stator current (13×2500 samples)	RMS, skewness and kurtosis, PCA (feature selection)	MLP (supervised)	98.25% accuracy in fault classification (4 classes).	Optimal MLP classifier for fault classification of three-phase induction motor (Ghate & Dudul, 2010).
2020	Power Transmission Line	Three-phase voltage and current (6×1000 samples)	Unsp. [1]	MLP (supervised)	100% accuracy in fault detection. 70% accuracy in fault classification (4 classes).	Fault detection and classification using ANN for power transmission line (Leh et al., 2020).
2020	Induction motor (stator winding)	Three-phase stator currents (3×1M samples)	Mean, average, RMS	MLP (supervised)	98.4% accuracy in fault identification.	An efficient MLP model for real-time fault detection in industrial machine (Verma, Nagpal, et al., 2020).
2021	Marine dual fuel engines	Engine input parameters (unsp[1])	Expert analysis	Kernel Method (best score), MLP, RF, OC-SVM	~97% accuracy in fault identification.	Authors trained ML models based on their proposed weakly learning technique so that the ML models would not suffer when the number of labels is insufficient (Coraddu et al., 2021)
2009	Gear	Vibration data (16384 samples)	Order cepstrum.	RBFNN (supervised)	100% accuracy in fault classification (3 classes).	Gear fault classification under speed-up conditions on order cepstrum and RBFNN (H. Li et al., 2009).
2020	Rolling bearing	Vibration data (6306 samples)	Welch power spectrum	W-RBFNN (supervised)	98.93% accuracy in fault classification (10 cases)	Fault diagnosis method of rolling bearings based on Welch power spectrum transformation with RBFNN (Jin et al., 2020).
2016	Induction motor (bearing)	Three-phase motor current (240 samples)	No data processing	1D-CNN (supervised)	97% accuracy in fault detection.	Real-time induction motor fault detection by 1-D CNN (Ince et al., 2016).
2020	High voltage electrical asset	EMI signals (13,360 samples)	Peak and average	1D-CNN (supervised)	99% accuracy in fault detection.	1D-CNN based real-time fault detection system for power asset diagnosis (Mitiche et al., 2020).
2020	Milling tool	Three-phase spindle current signals (1040 samples)	Order analysis	SSAE (unsupervised)	99% accuracy in fault classification (5 classes).	Order analysis and SSAE based milling tool wear condition monitoring (Ou et al., 2020).
2018	Wind turbine	8 signals (8200 samples)	Sliding Window	SW-DAE (unsupervised)	~91% accuracy in fault detection.	Wind turbine fault detection using a DAE with temporal information (Jian & Zhiyan, 2020).
2020	Tennessee Eastman Process	52 process variables (unsp.[1] samples)	Unsp.[1]	AAE (semi-supervised)	Graph-based quantitative evaluation.	A novel fault detection method based on AAE (Makhzani et al., 2015).
2020	Induction motor	Three-phase current (314 samples)	Unsp.[1]	Neuro-fuzzy (supervised)	93.3% accuracy in fault classification (6 classes)	Neuro-fuzzy system for fault detection of stator winding (Verma, Jain, et al., 2020).
2018	Tennessee Eastman Process	52 process variables (unsp.[1] samples)	Unsp.[1]	LSTM (supervised)	97% accuracy in fault classification (7 cases)	Fault classification in a chemical process using LSTM (Xavier & De Seixas, 2018).

[1] Unspecified information.

expert exists, there is a further benefit to examine the preprocessed data. For example, Karandikar et al. selected the relevant features based on their correlation to the value of R^2. Yet, the selected attributes were further examined and approved by a human expert, and each chosen attribute was rationally explained with respect to the system's behaviour (Karandikar et al., 2015). This correction process resulted in a tremendous decrease in the error rate. An alternative approach is to correct the feature set manually based on prior knowledge if the input vector is not sufficiently representative of the system's characteristics. Another strategy for using expert knowledge is to model the system's behaviour and use it to boost the learning algorithm's performance. Yiakopoulos et al. initialized cluster locations for a K-means clustering model that used a mathematical model of the physical behaviour of defective rolling-element bearings (Yiakopoulos et al., 2011). As an

Table 3
Summary Of Reviewed Literature on Clustering-Based Methods For Condition Monitoring.

Year	Machinery	Acquired signals & Dataset	Data Processing	ML model	Result	Summary
2011	Rolling element bearing	Vibration signals (8192 samples)	(kurtosis, skewness, variance, RMS, FT, WT, and order spectrum), signal envelope.	K-means (unsupervised)	100% accuracy in fault classification (3 cases)	Rolling element bearing fault classification in industrial environments based on K-means clustering (Yiakopoulos et al., 2011).
2018	Three-phase diode rectifiers	Three-phase voltage signals (unsp.[1] samples)	FFT	K-means (unsupervised)	100% accuracy in fault classification (8 classes).	A diode open circuit fault detection in a three-phase rectifier based on K-means clustering (Rahnama et al., 2018).
2017	Resonant grounding system	Transient zero-sequence current (unsp. samples)	HHT band-pass filter, wave transformation, SVD	FCM (unsupervised)	Graph-based quantitative evaluation.	FCM and SVD based earth fault detection in resonant grounding distribution systems (M. Guo & Yang, 2017).
2019	Continuous distillation system	6 system signals (unsp.[1] samples)	Stationary WT	SWT-FCM (unsupervised)	74% accuracy, 100% sensitivity, and 48% specificity in fault detection.	Abnormality detection in a continuous distillation system using FCM (Azzaoui et al., 2019).
2015	Power transformer	Radiometric signals (6×5000 samples)	PCA	Hierarchical Clustering (unsupervised)	Qualitative analysis.	Hierarchical clustering-based health monitoring of power transformers (Babnik et al., 2008).
2018	Closed-loop system	2 water flow rate and 3 water level signals (300 samples)	Simulation data	OCSVM (unsupervised)	~98% fault detection rate.	OCSVM based closed-loop system condition monitoring (Z. Li & Li, 2018).
2016	Tennessee Eastman Process	52 process variables (14850 samples)	Unsp.[1]	vnuOCSVM (unsupervised)	71.1% accuracy in fault detection.	Robust vnOCSVM for fault detection (Xiao et al., 2016).

[1] Unspecified information.

outcome, the enhanced K-means clustering model was not sensitive to the initial clustering point, which ultimately increased the algorithm's robustness. Model-based information extraction methods (e.g., EKF, ESIF) can be powerful strategies to estimate essential health parameters from available signals when available sensors are contaminated. Salar et al. successfully predicted the flow capacity of each stage in a multi-stage gas turbine system based on the available measurements (Salar et al., 2011). The diagnostic algorithm achieved a high accuracy rate in such a complex environment by utilizing corrected sensor values. After completing the data processing step, the most intuitive way of validating processed features is to compare the raw data visually with the processed data. Ou et al. utilized PCA to display the raw data; they projected the processed input vectors in a 3D graph. Their data processing method was visually inspected to validate its performance (Ou et al., 2020). They once more confirmed that the final dataset was ready for the training step.

In the model selection stage, the trade-off between a task's constraints and the results each diagnostic technique produces must be carefully inspected. Available data is a significant factor when selecting a machine learning model. For example, if the available data points contain an associated label or output, a supervised approach can be utilized to perform its task. Another standard is the amount of available data in the dataset. According to the comparison in Tables 1 and 2, deep learning methods typically require larger datasets for classification models. In the reviewed papers on condition monitoring systems with a large number of signals, we can also generalize that the researchers generally preferred ensemble methods over sole ones. If the machinery system consists of multiple parts or pipelines, an ensemble method can be an appropriate solution to isolate the corresponding fault. For example, gas turbines are complex systems involved in power generation and typically consist of individual components. Salar et al. introduced a hybrid approach called the EKF-Fuzzy method; they isolated each turbine section during the fault detection process (Salar et al., 2011). Similar to gas turbines, voltage distribution grids consist of multiple interconnected modules. In low voltage distribution grids, the proposed ensemble method was more viable than the conventional methods. The increase in fault resistance caused a degradation in the singular methods, but the proposed ensemble of models remained successful (Sapountzoglou et al., 2020). The PdM model is sometimes required to control the machinery's output. Suppose the model is required to control the system's reaction and function as a controller. In that case,n FL-based decision system is a powerful option that automatically makes critical decisions (interrupting the operation, adjusting the operation speed, etc.) in the event of a hazard (Cuka & Kim, 2017). However, we should note that FL-based strategies can only be implemented in the presence of an expert. After the model selection, a critical factor that influences the model's performance is parameter selection. For the SVM model, kernel functions must be selected according to the characteristics of the available dataset. For example, the Gaussian kernel function exhibits better classification results with large amounts of data, while RBF is more effective on small data samples (J. Guo et al., 2020). Another approach to boost the ML model's performance is to change the model's architecture with respect to the system's known characteristics. For example, experts know that the leading cause of vibration in the compressor is due mainly a frost. In response, He et al. implemented an extra node into a BN to present site information that indicates a fault event (He et al., 2016). Subsequently, the rate of undetected refrigerator leaks (27.6%) and overcharging (9.4%) decreased to 7.6% and 5.4%, respectively (He et al., 2016).

The training process involves tuning hyperparameters of the ML

Table 4
Summary Of Reviewed Literature on Regression-Based Methods For Condition Monitoring.

Year	Machinery	Dataset	Data Processing	ML model	Result	Summary
2018	Rolling bearing	Vibration signals (2000 samples)	Unsp.[1]	LSTM (supervised)	0.018 RMSE in vibration data estimation.	LSTM approach for rotating machinery fault prognosis (Xie & Zhang, 2018).
2019	Turbofan engine	21 signals (33,727 samples)	Manual feature selection	LSTM (supervised)	11.42 RMSE in RUL estimation.	LSTM for predictive maintenance in smart industries (Bruneo & De Vita, 2019).
2020	Motor system	Acceleration signal (480,000 samples)	RMS, Kurtosis, CWT, scaling and smoothing	LSTM (supervised)	83.2% accuracy in anomaly detection.	Development of a speed invariant deep learning model with application to condition monitoring of rotating machinery (Lee et al., 2020).
2014	Naval propulsion system	Performance degradation signals (up to 500 samples)	Expert Analysis	SVR (supervised)	Up to ~ 0.09 MSE in RUL estimation.	an SVR-based PdM model was proposed to monitor performance degradation in vessel propulsion (Coraddu et al., 2014).
2018	Wind turbine	Wind speed (626 samples)	Air density correction	ε-SVR (supervised)	Graph-based quantitative comparison.	Comparative analysis of binning and SVR for wind turbine rotor speed-based power curve use in condition monitoring (Pandit & Infield, 2018).
2017	Cutting tool	6 force signals (6×315 signals)	MFCC	SVR (supervised)	0.0398 RMSE on average for 6 cutting tools	Tool condition monitoring based MCC-SVR (Benkedjouh et al., 2017).
2019	Coal mills	12 variables (230 samples)	Manual feature selection	SVR (supervised)	0.059 MSE for ventilation pressure estimation.	Abnormal condition monitoring and diagnosis for coal mills based on SVR (Hong et al., 2019).
2020	Li-ion batteries	Battery capacity (unsp.[1] samples)	Unsp.[1]	QPSO-ISVR (supervised)	0.023 RMSE and ~ 6% MAPE for capacity degradation estimation on average of two test cases.	Reliable state of health condition monitoring of Li-ion batteries based on QPSO-ISVR (Ben Ali et al., 2020).

[1] Unspecified information.

model to reach an accurate conclusion when given new data. Hence, chosen training strategy greatly impacts the overall performance of the learning algorithm. The researchers mostly did not give much information about this process since it is a standard process in the industry. Thus, from our observations of the literature, 3D RMSE plots are among the most intuitive approaches to determining optimal parameters. Fig. 11 shows that the optimal hyperparameter yielding the lowest RMSE value is the lowest blue part of the graph.

Another important factor to consider is the training duration of the ML algorithm. Following the comparison of 11 different MLP models, it is reported that the training duration is not always correlated to performance (Verma et al., 2020b). Similarly, increasing the number of data points is not always yield better performance. In 2020, Jin et al. evaluated the model's performance as a function of the training set's size. In the experiment, the W-RBFNN method was trained with two different dataset sizes (90 and 12,000 samples), and the model's performance only improved from 98.61% to 100% (Jin et al., 2020). Therefore, they found that the proposed model's computational cost could be vastly decreased by sacrificing only ~1.4% of its accuracy.

Before deploying the ML model into a real-world application, the model's performance and reliability are tested through a test set. Due to our analysis, some of the testing approaches produce more realistic results for real-world applications. Additional strategies could be applied besides the classical test run. For example, after testing the learning algorithm with the original dataset, a secondary test set may involve the same dataset with more noisy components, as Jin et al. have already done (Jin et al., 2020). Another option can be providing missing data for a long duration, short duration, or varying the environmental noise (H. Wang et al., 2020). In this way, we would be able to quantify the robustness of the learning algorithm.

The main characteristics of condition monitoring systems were highlighted in the summary tables. Most of the system signals (except simulations) were extracted using a feature extraction strategy. However, a few deep learning models did not require a feature engineering step, which are 1-D CNN, autoencoders, and LSTM. So, we can conclude that deep learning models have high enough complexity to decompose patterns without the help of a data processing step. However, their performance becomes more stable when the feature engineering step is involved. Another factor to consider when comparing two proposed ML models is their learning type. Due to Tables 1-3, the performance of classification models is usually higher than clustering ones. We can generalize that the clustering task is more challenging than supervised classification. As a result, the researchers increased the intensity of the feature engineering step to improve the accuracy of clustering-based anomaly detection algorithms. On the other hand, classification models mostly had only a single preprocessing step, since the performance is high enough. We saw the same effect on the dataset size as well. Simple classifiers (i.e., BN, NB and SVM) are not strongly affected by the dataset size. Some researchers only utilized 90 data points to train their model. In regression analysis, LSTM models achieved promising results in RUL prediction. However, they require large datasets for training. Ensemble models are typically applied to complex systems with multiple components and signals.

Due to our analysis, the researchers are exploring ways to further develop the existing ML models. Taking the SVM concept as an example of how many modifications the SVM method has: the MSVM methods for multi-class classification, the OC-SVM method as a clustering approach, and the SVDD-DAG SVM method to build hyper-spheres. Furthermore, 1D CNN has been derived from the classical CNN to adapt the original concept for time-series datasets. Additionally, the BN structure was modified to be implemented into dynamic systems, known as DBN. Another example is the proposal of vnuOCSVM, where the model was designed to resolve the performance degradation of the OCSVM method.

Based on the reviewed articles, we want to share our observations on

Table 5
Summary Of Reviewed Literature on Ensemble Methods For Condition Monitoring.

Year	Machinery	Acquired signals & Dataset	Data Processing	ML model	Result	Summary
2019	Hydraulic System	18 sensors (18×2205 samples)	Mean, square root amplitude, skewness, kurtosis, kurtosis index, skewness index, waveform indicator, PCA, Pearson correlation coefficients	Ensemble SVM (stacking) (supervised)	88.6% accuracy in fault detection.	Health condition monitoring of hydraulic system based on ensemble (stacking) SVM (P. Guo et al., 2019).
2011	Gas Turbines	5 health parameter estimation (unsp[1]. samples)	EKF	EKF-FIS (stacking) (unsupervised)	96.15% accuracy in fault classification (8 classes)	A hybrid EKF-Fuzzy approach to fault detection and isolation of industrial gas turbines (Salar et al., 2011).
2020	Diesel Engine	15 signals (100×15 signals)	Unsp.[1]	RF (bagging) (supervised)	95.1% accuracy and 90% F1 score in fault detection.	Intelligent fault diagnosis of diesel engine based on RF ensemble method (Shao et al., 2020).
2020	CNC machine	Spindle motor current (unsp.[1] samples)	Mean, RMS, kurtosis, margin factor, frequency centroid, spectral mean square, frequency variance, frequency band energy, VMD	RF (Bootstrap) (supervised)	95.19% accuracy in fault classification (3 cases).	Tool wear condition monitoring by combining variational mode decomposition and Ensemble learning (Yuan et al., 2020).
2013	Semi-conductor etch process	17 signals (1388 samples)	Manual feature selection	BagSVDD (supervised)	46.1% accuracy and 0.027 false alarm rate in fault detection	Bagging SVDD model for batch monitoring (Ge & Song, 2013).
2020	Low voltage distribution grid	72 signals (230,688 samples)	Unsp.[1]	Gradient boosting ensemble tree (supervised)	~90% accuracy in classification (4 classes)	Fault diagnosis in low voltage smart distribution grids using gradient boosting trees (Sapountzoglou et al., 2020).
2020	Hydraulic fluids	Viscosity, fire point, auto-ignition temperature, boiling point, vapor pressure (5×73 samples)	Unsp.[1]	Ensemble boosted tree regression model	31.68 RMSE in fault detection.	Hydraulic fluid monitoring based on ensemble boosted regression tree model (Uma Maheswari et al., 2020).

[1] Unspecified information.

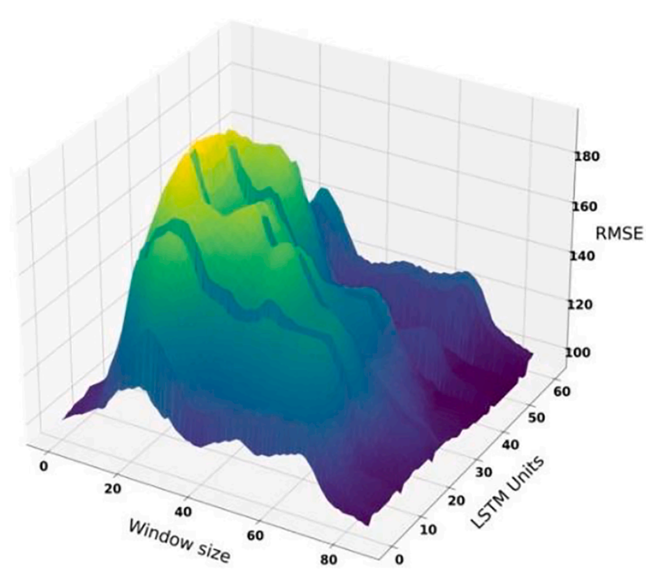

Fig. 11. An exemplary 3D RMSE plot for tuning the hyperparameters of the LSTM model.

the research gaps in the field. The biggest problem of each application is the number of available data. Thus, augmenting the available data should be the main consideration. Still, the researchers do not employ data augmentation techniques in their proposed PdM models. Thus, a PdM model with data augmentation step (e.g., bootstrapping, generative adversarial network) may eliviate the data scarcity problem. Our second recommendation is to include recently developed ML models rather than conventional ML models. One example is using a multihead attention layer in a sequential ML model (Vaswani et al., 2017). Lastly, the successful studies in different fields can be integrated into the PdM. For instance. Due to Uber's research, the LSTM model can be utilized to encode the input data (Laptev et al., 2017). As a result, the prediction performance is significantly boosted compared to the classical pre-processing methods (e.g., time series analysis). Finally, a rapidly developing field, known as physics informed machine learning, should be explored further for meaningful condition monitoring strategies that make use of the underlying physics of a system coupled with ML techniques.

5. Conclusion

Currently within the literature, there are few resources that comprehensively inform users or evaluate the quality of proposed condition monitoring systems. That being the case, this survey paper extensively reviewed recent ML-driven condition monitoring techniques. Based on the reviewed literature, we provided an insight into the underlying findings on successful condition monitoring systems, and we shared our observations on the research gaps in the PdM field. Section 2 demonstrated how to properly process data for a model to extract all possible knowledge from the available information. The systems in this study and their features must be critically considered, and comprehensive investigations should precede decisions on the appropriate feature extraction and selection methods. Section 3 presented recent ML techniques for effective predictive maintenance while providing comparative analysis and explanatory insights into each technique's advantages or limitations. Most techniques excel in a certain task but often perform poorly in others. Consequently, no generalization can be made on the overall best technique. In Section 4, the underlying findings of the reviewed condition monitoring systems were elucidated with our observations. Based on our analysis, it is prudent to consider all factors when narrowing the search for the most effective model for a particular task. For instance, deep learning methods may perform well using nonlinear or high-dimensional data; however, an ensemble method may outperform a deep learning model. Furthermore, a simple physics-based model coupled with a signal processor may yield the best and most cost effective condition monitoring system. Therefore, the tradeoff between a task's constraints and each diagnostic technique's capabilities must be evaluated quantitively and comparatively to determine the optimal solution for a given problem.

CRediT authorship contribution statement

Onur Surucu: Investigation, Visualization, Writing – original draft, Writing – review & editing. **Stephen Andrew Gadsden:** Supervision, Conceptualization, Writing – review & editing. **John Yawney:** Supervision, Writing – review & editing.

Declaration of Competing Interest

The authors appreciate funding support from Natural Sciences and Engineering Research Council of Canada (NSERC) under Alliance grant # ALLRP 566259-21 (Gadsden).

Data availability

No data was used for the research described in the article.

References

0 PHM Society Conference Data Challenge - PHM Society. (2021). In *PHM Society*.

A. Davies. (1998). *Handbook of Condition Monitoring: Techniques and Methodology*.

A novel transmission line relaying scheme for fault detection and classification using wavelet transform and linear discriminant analysis. (2015). In *Ain Shams Eng. J.* (Vol. 6, Issue 1, pp. 199–209). Elsevier. https://doi.org/10.1016/j.asej.2014.10.005.

Agyare, O. R., Asiedu-Asante, A. B., & Biney, A. R. (2019). Fuzzy Logic Based Condition Monitoring of a 3-Phase Induction Motor. *IEEE AFRICON Conference*, *2019-Septe*. https://doi.org/10.1109/AFRICON46755.2019.9133780.

Ajit, A., Acharya, K., & Samanta, A. (2020). A Review of Convolutional Neural Networks. International Conference on Emerging Trends in Information Technology and Engineering. *Ic-ETITE, 2020*, 1–5. https://doi.org/10.1109/ic-ETITE47903.2020.049

Akpudo, U. E., & Hur, J. W. (2020). Intelligent Solenoid Pump Fault Detection based on MFCC Features, LLE and SVM. *2020 International Conference on Artificial Intelligence in Information and Communication, ICAIIC 2020*, 404–408. https://doi.org/10.1109/ICAIIC48513.2020.9065282.

Aliev, R. A. (2013). A Generalized Fuzzy Logic-Based Decision Theory. In *SpringerLink* (pp. 127–189). Springer. https://doi.org/10.1007/978-3-642-34895-2_4.

Alves, M. A. D., Galotto, L., Pinto, J. O. P., García, R. C., Teixeira, H., & Campos, M. C. M. (2018). RBF neural networks modeling methodology compared to non-parametric auto-associative models for condition monitoring applications. *Proceedings: IECON 2018 - 44th Annual Conference of the IEEE Industrial Electronics Society*, 5406–5411. https://doi.org/10.1109/IECON.2018.8591107.

Amin, M. T., Khan, F., & Imtiaz, S. (2019). Fault detection and pathway analysis using a dynamic Bayesian network. *Chemical Engineering Science, 195*, 777–790. https://doi.org/10.1016/j.ces.2018.10.024

Amruthnath, N. (2018). *A Research Study on Unsupervised Machine Learning Algorithms for Early Fault Detection in Predictive Maintenance*. *August 1993*, 355–361.

Atoui, I., Meradi, H., Boulkroune, R., Saidi, R., & Grid, A. (2013). Fault detection and diagnosis in rotating machinery by vibration monitoring using FFT and Wavelet techniques. *2013 8th International Workshop on Systems. Signal Processing and Their Applications, WoSSPA, 2013*, 401–406. https://doi.org/10.1109/WoSSPA.2013.6602399

Azzaoui, H., Mansouri, I., & Elkihel, B. (2019). Methylcyclohexane continuous distillation column fault detection using stationary wavelet transform & fuzzy C-means. *Materials Today: Proceedings, 13*, 597–606. https://doi.org/10.1016/j.matpr.2019.04.018

Babnik, T., Aggarwal, R. K., Member, S., Moore, P. J., & Member, S. (2008). Principal Component and Hierarchical Cluster Analyses as Applied to Transformer Partial Discharge Data With Particular Reference to Transformer Condition Monitoring. 23 (4), 2008–2016.

Baraldi, P., Canesi, R., Zio, E., Seraoui, R., & Chevalier, R. (2011). Genetic algorithm-based wrapper approach for grouping condition monitoring signals of nuclear power plant components. *Integrated Computer-Aided Engineering, 18*(3), 221–234. https://doi.org/10.3233/ICA-2011-0375

Barron, J. W., Moustapha, A. I., & Selmic, R. R. (2008). Real-Time Implementation of Fault Detection in Wireless Sensor Networks Using Neural Networks. *Fifth International Conference on Information Technology: New Generations (Itng 2008)*, 378–383. https://doi.org/10.1109/ITNG.2008.187.

Ben Ali, J., Azizi, C., Saidi, L., Bechhoefer, E., & Benbouzid, M. (2020). Reliable state of health condition monitoring of Li-ion batteries based on incremental support vector regression with parameters optimization. *Proceedings of the Institution of Mechanical Engineers. Part I: Journal of Systems and Control Engineering*. https://doi.org/10.1177/0959651820950849.

Benkedjouh, T., Zerhouni, N., & Rechak, S. (2017). Tool condition monitoring based on mel-frequency cepstral coefficients and support vector regression. *2017 5th International Conference on Electrical Engineering - Boumerdes, ICEE-B 2017*, *2017-Janua*, 1–5. https://doi.org/10.1109/ICEE-B.2017.8192024.

Bhat, N. N., Dutta, S., Vashisth, T., Pal, S., Pal, S. K., & Sen, R. (2016). Tool condition monitoring by SVM classification of machined surface images in turning. *International Journal of Advanced Manufacturing Technology, 83*(9–12), 1487–1502. https://doi.org/10.1007/s00170-015-7441-3

Bruneo, D., & De Vita, F. (2019). On the use of LSTM networks for predictive maintenance in smart industries. *Proceedings - 2019 IEEE International Conference on Smart Computing, SMARTCOMP 2019*, 241–248. https://doi.org/10.1109/SMARTCOMP.2019.00059.

Caesarendra, W. (2016). *Vibration and acoustic emission-based condition monitoring and prognostic methods for very low speed slew bearing*. University of Wollongong.

Caesarendra, W., Kosasih, B., Tieu, K., & Moodie, C. A. S. (2013). An application of nonlinear feature extraction - A case study for low speed slewing bearing condition monitoring and prognosis. *IEEE/ASME International Conference on Advanced Intelligent Mechatronics, 2013*, 1713–1718. https://doi.org/10.1109/AIM.2013.6584344

Calin, O. (2020). *Recurrent Neural Networks BT - Deep Learning Architectures: A Mathematical Approach* (O. Calin, Ed.; pp. 543–559). Springer International Publishing. https://doi.org/10.1007/978-3-030-36721-3_17.

Çinar, Z. M., Nuhu, A. A., Zeeshan, Q., Korhan, O., Asmael, M., & Safaei, B. (2020). Machine learning in predictive maintenance towards sustainable smart manufacturing in industry 4.0. *Sustainability (Switzerland), 12*(19). https://doi.org/10.3390/su12198211

Cocconcelli, M., Zimroz, R., Rubini, R., & Bartelmus, W. (2012). STFT Based Approach for Ball Bearing Fault Detection in a Varying Speed Motor. *Condition Monitoring of Machinery in Non-Stationary Operations, 41–50*. https://doi.org/10.1007/978-3-642-28768-8_5

Coraddu, A., Oneto, L., Ghio, A., Savio, S., Anguita, D., & Figari, M. (2014). Machine learning approaches for improving condition-based maintenance of naval propulsion plants. *Proceedings of the Institution of Mechanical Engineers, Part M: Journal of Engineering for the Maritime Environment, 230*(1), 136–153. https://doi.org/10.1177/1475090214540874

Coraddu, A., Oneto, L., Ilardi, D., Stoumpos, S., & Theotokatos, G. (2021). Marine dual fuel engines monitoring in the wild through weakly supervised data analytics. *Engineering Applications of Artificial Intelligence, 100*, Article 104179. https://doi.org/10.1016/j.engappai.2021.104179

Cuka, B., & Kim, D. W. (2017). Fuzzy logic based tool condition monitoring for end-milling. *Robotics and Computer-Integrated Manufacturing*, *47*(December 2016), 22–36. https://doi.org/10.1016/j.rcim.2016.12.009.

Datta, S. (2020). *A Review on Convolutional Neural Networks* (pp. 445–452). https://doi.org/10.1007/978-981-15-4932-8_50.

Daubechies, I. (1990). The Wavelet Transform, Time-Frequency Localization and Signal Analysis. *IEEE Transactions on Information Theory, 36*(5), 961–1005. https://doi.org/10.1109/18.57199

Debnath, L., & Shah, F. A. (2015). The Wigner{ extendash}Ville Distribution and Time{ extendash}Frequency Signal Analysis. In *SpringerLink* (pp. 287–336). Boston, MA: Birkh{ifmmodeddot{a}elseä i}user. https://doi.org/10.1007/978-0-8176-8418-1_5.

Deisenroth, M. P., Faisal, A. A., & Ong, C. S. (2020). Mathematics for Machine Learning. *Cambridge University Press*. https://doi.org/10.1017/9781108679930

Dong, X., Yu, Z., Cao, W., Shi, Y., & Ma, Q. (2020). A survey on ensemble learning. *Frontiers of Computer Science, 14*(2), 241–258. https://doi.org/10.1007/s11704-019-8208-z

Du, K. L., & Swamy, M. N. S. (2014). Neural networks and statistical learning. In *Neural Networks and Statistical Learning* (Vol. 9781447155). https://doi.org/10.1007/978-1-4471-5571-3.

Feng, G., Lai, C., Tjong, J., & Kar, N. C. (2018). Noninvasive Kalman Filter Based Permanent Magnet Temperature Estimation for Permanent Magnet Synchronous Machines. *IEEE Transactions on Power Electronics, 33*(12), 10673–10682. https://doi.org/10.1109/TPEL.2018.2808323

Fernandes, M., Canito, A., Bolón-Canedo, V., Conceição, L., Praça, I., & Marreiros, G. (2019). Data analysis and feature selection for predictive maintenance: A case-study in the metallurgic industry. *International Journal of Information Management, 46*(431), 252–262. https://doi.org/10.1016/j.ijinfomgt.2018.10.006

Gadsden, S. (2011). *SMOOTH VARIABLE STRUCTURE FILTERING: THEORY AND APPLICATIONS*.

Gadsden, S., & AlShabi, M. (2020). The Sliding Innovation Filter. *IEEE Access, PP, 1*. https://doi.org/10.1109/ACCESS.2020.2995345
Gao, R. X., Yan, R., Sheng, S., & Zhang, L. (2006). *Sensor Placement and Signal Processing for Bearing Condition Monitoring*. 167–191. https://doi.org/10.1007/1-84628-269-1_7.
Gardner, M. W., & Dorling, S. R. (1998). Artificial neural networks (the multilayer perceptron) - a review of applications in the atmospheric sciences. *In Atmospheric Environment, 32*(14–15), 2627–2636. https://doi.org/10.1016/S1352-2310(97)00447-0
Ge, Z., & Song, Z. (2013). Bagging support vector data description model for batch process monitoring. *Journal of Process Control, 23*(8), 1090–1096. https://doi.org/10.1016/j.jprocont.2013.06.010
Ghate, V. N., & Dudul, S. V. (2010). Optimal MLP neural network classifier for fault detection of three phase induction motor. *Expert Systems with Applications, 37*(4), 3468–3481. https://doi.org/10.1016/j.eswa.2009.10.041
Gray, R. M., & Goodman, J. W. (1995). The Fourier Transform. In *Fourier Transforms: An Introduction for Engineers* (pp. 53–113). US: Springer. https://doi.org/10.1007/978-1-4615-2359-8_2.
Grimble, M., & Majecki, P. (2020). *Nonlinear Estimation and Condition Monitoring: State-Space Approach* (pp. 597–641). https://doi.org/10.1007/978-1-4471-7457-8_13.
Gs, V., Pai, P., Sriram, N. S., & Rao, R. (2011). *Bearing Diagnostics – A Radial Basis Function Neural Network Approach*.
Gugulothu, N., Vishnu, T. V., Gupta, P., Malhotra, P., Vig, L., Agarwal, P., & Shroff, G. (2018). On practical aspects of using RNNs for fault detection in sparsely-labeled multi-sensor time series. *Proceedings of the Annual Conference of the Prognostics and Health Management Society, PHM*. 1–10.
Guo, J., Li, A., & Zhang, R. (2020). Tool condition monitoring in milling process using multifractal detrended fluctuation analysis and support vector machine. *International Journal of Advanced Manufacturing Technology, 110*(5–6), 1445–1456. https://doi.org/10.1007/s00170-020-05931-5
Guo, M., & Yang, N. (2017). Electrical Power and Energy Systems Features-clustering-based earth fault detection using singular-value decomposition and fuzzy c-means in resonant grounding distribution systems. *International Journal of Electrical Power and Energy Systems, 93*, 97–108. https://doi.org/10.1016/j.ijepes.2017.05.014
Guo, P., Wu, J., Xu, X., Cheng, Y., & Wang, Y. (2019). Health condition monitoring of hydraulic system based on ensemble support vector machine. *2019 Prognostics and System Health Management Conference. PHM-Qingdao, 2019*, 1–5. https://doi.org/10.1109/PHM-Qingdao46334.2019.8942981
He, S., Wang, Z., Wang, Z., Gu, X., & Yan, Z. (2016). Fault detection and diagnosis of chiller using Bayesian network classifier with probabilistic boundary. *Applied Thermal Engineering, 107*, 37–47. https://doi.org/10.1016/j.applthermaleng.2016.06.153
Heckerman, D. (2008). A Tutorial on Learning with Bayesian Networks. In *SpringerLink* (pp. 33–82). Springer. https://doi.org/10.1007/978-3-540-85066-3_3.
Helwig, N., Pignanelli, E., & Schütze, A. (2015). Condition monitoring of a complex hydraulic system using multivariate statistics. *2015 IEEE International Instrumentation and Measurement Technology Conference (I2MTC) Proceedings*, 210–215. https://doi.org/10.1109/I2MTC.2015.7151267.
Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation, 9*(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735
Hong, X., Xu, Z., & Zhang, Z. (2019). Abnormal Condition Monitoring and Diagnosis for Coal Mills Based on Support Vector Regression. *IEEE Access, 7*, 170488–170499. https://doi.org/10.1109/ACCESS.2019.2955249
Huang, Y., Chen, C. H., & Huang, C. J. (2019). Motor fault detection and feature extraction using rnn-based variational autoencoder. *IEEE Access, 7*, 139086–139096. https://doi.org/10.1109/ACCESS.2019.2940769
Hubel, D. H., & Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *The Journal of Physiology, 195*(1), 215–243. https://doi.org/10.1113/jphysiol.1968.sp008455
Ince, T., Kiranyaz, S., Eren, L., Askar, M., & Gabbouj, M. (2016). Real-Time Motor Fault Detection by 1-D Convolutional Neural Networks. *IEEE Transactions on Industrial Electronics, 63*(11), 7067–7075. https://doi.org/10.1109/TIE.2016.2582729
Jaber, A. A. (2017a). Design of an Intelligent Embedded System for Condition Monitoring of an Industrial Robot. *Springer International Publishing*. https://doi.org/10.1007/978-3-319-44932-6
Jaber, A. A. (2017b). *Signal Processing Techniques for Condition Monitoring*. 53–73. https://doi.org/10.1007/978-3-319-44932-6_3.
Jian, W., & Zhiyan, H. (2020). A Novel Fault Detection Method Based on Adversarial Auto-Encoder. *2020 39th Chinese Control Conference (CCC)*, 4166–4170. https://doi.org/10.23919/CCC50068.2020.9188749.
Jiang, G., Xie, P., He, H., & Yan, J. (2018). Wind Turbine Fault Detection Using a Denoising Autoencoder With Temporal Information. *IEEE/ASME Transactions on Mechatronics, 23*(1), 89–100. https://doi.org/10.1109/TMECH.2017.2759301
Jin, Z., Han, Q., Zhang, K., & Zhang, Y. (2020). An intelligent fault diagnosis method of rolling bearings based on Welch power spectrum transformation with radial basis function neural network. *JVC/Journal of Vibration and Control, 26*(9–10), 629–642. https://doi.org/10.1177/1077546319889859
Jović, A., Brkić, K., & Bogunović, N. (2015). A review of feature selection methods with applications. *2015 38th International Convention on Information and Communication Technology, Electronics and Microelectronics, MIPRO 2015 - Proceedings*, 1200–1205. https://doi.org/10.1109/MIPRO.2015.7160458.
Jung, D., Lee, S., Wang, H., Kim, J., & Lee, S. (2010). Fault detection method with PCA and LDA and its application to induction motor. *Journal of Central South University of Technology, 17*, 1238–1242. https://doi.org/10.1007/s11771-010-0625-y
Karaduman, G., & Akin, E. (2020). A New Approach Based on Predictive Maintenance Using the Fuzzy Classifier in Pantograph-Catenary Systems. *IEEE Transactions on Intelligent Transportation Systems, 1–11*. https://doi.org/10.1109/tits.2020.3042997
Karandikar, J., McLeay, T., Turner, S., & Schmitz, T. (2015). Tool wear monitoring using naïve Bayes classifiers. *International Journal of Advanced Manufacturing Technology, 77*(9–12), 1613–1626. https://doi.org/10.1007/s00170-014-6560-6
Kim, E. Y., Tan, A. C. C., Yang, B.-S., & Kosse, V. (2007). Experimental study on condition monitoring of low speed bearings : time domain analysis. *Fifth Australasian Congress on Applied Mechanics (ACAM 2007), December*, 108–113.
Kolman, E., & Margaliot, M. (2009). *Knowledge-Based Neurocomputing: A Fuzzy Logic Approach* (Vol. 234). https://doi.org/10.1007/978-3-540-88077-6
Kramer, O. (2013). Ensemble Learning. In *SpringerLink* (pp. 25–32). Springer. https://doi.org/10.1007/978-3-642-38652-7_3.
Kumar, V. (2014). Feature Selection: A literature Review. *The Smart Computing Review, 4* (3). https://doi.org/10.6029/smartcr.2014.03.007
Laptev, N. P., Yosinski, J., Li, L. E., & Smyl, S. (2017). *Time-series Extreme Event Forecasting with Neural Networks at Uber*.
Layer, E., & Tomczyk, K. (2015). Wavelet Transform. In *SpringerLink* (pp. 97–105). Springer. https://doi.org/10.1007/978-3-319-13209-9_5.
Lee, W. J., Xia, K., Denton, N. L., Ribeiro, B., & Sutherland, J. W. (2020). Development of a speed invariant deep learning model with application to condition monitoring of rotating machinery. *Journal of Intelligent Manufacturing, Kim, 2019*. https://doi.org/10.1007/s10845-020-01578-x
Leh, N. A. M., Zain, F. M., Muhammad, Z., Hamid, S. A., & Rosli, A. D. (2020). Fault Detection Method Using ANN for Power Transmission Line. *Proceedings - 10th IEEE International Conference on Control System, Computing and Engineering, ICCSCE 2020, August*, 79–84. https://doi.org/10.1109/ICCSCE50387.2020.9204921.
Li, H., Zhang, Y., & Zheng, H. (2009). Gear fault detection and diagnosis under speed-up condition based on order cepstrum and radial basis function neural network. *Journal of Mechanical Science and Technology, 23*(10), 2780–2789. https://doi.org/10.1007/s12206-009-0730-8
Li, H., Zheng, H., & Tang, L. (2006). *Wigner-Ville Distribution Based on EMD for Faults Diagnosis of Bearing BT - Fuzzy Systems and Knowledge Discovery* (L. Wang, L. Jiao, G. Shi, X. Li, & J. Liu, Eds.; pp. 803–812). Springer Berlin Heidelberg.
Li, Z., & Li, X. (2018). Fault detection in the closed-loop system using one-class support vector machine. *Proceedings of 2018 IEEE 7th Data Driven Control and Learning Systems Conference, DDCLS 2018, 1*, 251–255. https://doi.org/10.1109/DDCLS.2018.8515960.
Lin, R. H., Xi, X. N., Wang, P. N., Wu, B. D., & Tian, S. M. (2019). Review on hydrogen fuel cell condition monitoring and prediction methods. *International Journal of Hydrogen Energy, 44*(11), 5488–5498. https://doi.org/10.1016/j.ijhydene.2018.09.085
Liton Hossain, M., Abu-Siada, A., & Muyeen, S. M. (2018). Methods for advanced wind turbine condition monitoring and early diagnosis: A literature review. *Energies, 11* (5). https://doi.org/10.3390/en11051309
Liu, H., Hu, F., Su, J., Wei, X., & Qin, R. (2020). Comparisons on Kalman-Filter-Based Dynamic State Estimation Algorithms of Power Systems. *IEEE Access, 8*, 51035–51043. https://doi.org/10.1109/ACCESS.2020.2979735
Makhzani, A., Shlens, J., Jaitly, N., & Goodfellow, I. J. (2015). Adversarial Autoencoders. *CoRR, abs/1511.0*. http://arxiv.org/abs/1511.05644.
Marwala, T. (2000). *Fault Identification Using Neural Networks And Vibration Data. August*, 192.
Marwala, T. (2012). Condition monitoring using computational intelligence methods: Applications in mechanical and electrical systems. *Condition Monitoring Using Computational Intelligence Methods: Applications in Mechanical and Electrical Systems, 9781447123*(2001), 1–235. https://doi.org/10.1007/978-1-4471-2380-4
Mirnaghi, M. S., & Haghighat, F. (2020). Fault detection and diagnosis of large-scale HVAC systems in buildings using data-driven methods: A comprehensive review. *Energy and Buildings, 229*. https://doi.org/10.1016/j.enbuild.2020.110492
Mitiche, I., Nesbitt, A., Conner, S., Boreham, P., & Morison, G. (2020). 1D-CNN based real-time fault detection system for power asset diagnostics. *IET Generation, Transmission and Distribution, 14*(24), 5816–5822. https://doi.org/10.1049/iet-gtd.2020.0773
Mobley, R. K. (2002). *An Introduction to Predictive Maintenance*. Butterworth-Heinemann.
Mohammadi, M. R., Sadrossadat, S. A., Mortazavi, M. G., & Nouri, B. (2017). A brief review over neural network modeling techniques. *2017 IEEE International Conference on Power, Control, Signals and Instrumentation Engineering (ICPCSI)*, 54–57. https://doi.org/10.1109/ICPCSI.2017.8391781.
Murphy, K. (2002). Dynamic Bayesian Networks: Representation, Inference and Learning. In *Ph.D*.
Murtagh, F., & Contreras, P. (2011). Methods of Hierarchical Clustering. *CoRR, abs/1105.0*.
Nayak, J., Naik, B., & Behera, Dr. H. (2015). Fuzzy C-Means (FCM) Clustering Algorithm: A Decade Review from 2000 to 2014. In *Smart Innovation, Systems and Technologies* (Vol. 32, pp. 133–149). https://doi.org/10.1007/978-81-322-2208-8_14.
Niros, A., & Tsekouras, G. (2016). *A Radial Basis Function Neural Network Training Mechanism for Pattern Classification Tasks* (pp. 193–206). https://doi.org/10.1007/978-3-319-24211-8_8.
Niu, G. (2017). Data Acquisition and Preprocessing. In Data-Driven Technology for Engineering Systems Health Management: Design Approach, Feature Construction, Fault Diagnosis, Prognosis, Fusion and Decisions (pp. 49–99). Springer Singapore. https://doi.org/10.1007/978-981-10-2032-2_4.
Nursalam, 2016, metode penelitian, & Fallis, A. G. (2013). Condition Monitoring of Rotating Electrical Machines. In *Journal of Chemical Information and Modeling* (Vol. 53, Issue 9).

Ou, J., Li, H., Huang, G., & Zhou, Q. (2020). A Novel Order Analysis and Stacked Sparse Auto-Encoder Feature Learning Method for Milling Tool Wear Condition Monitoring. *Sensors, 20*(10). https://doi.org/10.3390/s20102878

Pandarakone, S. E., Gunasekaran, S., Mizuno, Y., & Nakamura, H. (2018). Application of Naive Bayes Classifier Theorem in Detecting Induction Motor Bearing Failure. *Proceedings - 2018 23rd International Conference on Electrical Machines, ICEM 2018*, 1761–1767. https://doi.org/10.1109/ICELMACH.2018.8506836.

Pandarakone, S. E., Masuko, M., Mizuno, Y., & Nakamura, H. (2018b). Deep Neural Network Based Bearing Fault Diagnosis of Induction Motor Using Fast Fourier Transform Analysis. *IEEE Energy Conversion Congress and Exposition (ECCE), 2018*, 3214–3221. https://doi.org/10.1109/ECCE.2018.8557651

Pandit, R., & Infield, D. (2018). Comparative analysis of binning and support vector regression for wind turbine rotor speed based power curve use in condition monitoring. *Proceedings - 2018 53rd International Universities Power Engineering Conference, UPEC 2018*. https://doi.org/10.1109/UPEC.2018.8542057.

Pepper, S. H. (2000). Synchronous Sampling and Applications to Analytic Signal Estimation. *56th ARFTG Conference Digest: Metrology and Test for RF Telecommunications, ARFTG 2000*. https://doi.org/10.1109/ARFTG.2000.327436.

Qu, F., Liu, J., Zhu, H., & Zang, D. (2020). Wind Turbine Condition Monitoring Based on Assembled Multidimensional Membership Functions Using Fuzzy Inference System. *IEEE Transactions on Industrial Informatics, 16*(6), 4028–4037. https://doi.org/10.1109/TII.2019.2957409

Rahnama, M., Vahedi, A., Alikhani, A. M., Takorabet, N., & Fazelbakhsheshi, B. (2018). A novel diode open circuit fault detection in three phase rectifier based on k-means method. *IEEE International Conference on Industrial Technology (ICIT), 2018*, 600–605. https://doi.org/10.1109/ICIT.2018.8352246

Rao, B. K. N. (1998). *Condition monitoring and the integrity of industrial systems BT - Handbook of Condition Monitoring: Techniques and Methodology* (A. Davies, Ed.; pp. 3–34). Springer Netherlands. https://doi.org/10.1007/978-94-011-4924-2_1.

Salar, A., Sedigh, A. K., Hosseini, S. M., & Khaledi, H. (2011). A hybrid EKF-fuzzy approach to fault detection and isolation of industrial gas turbines. *Proceedings of the ASME Turbo Expo, 3*, 251–260. https://doi.org/10.1115/GT2011-45878

Sandoval, D., Leturiondo, U., Pozo, F., Vidal, Y., & Salgado, O. (2019). *Trends in condition monitoring for pitch bearings., 2018*(2), 1–12. https://doi.org/10.1784/cm.2019.113

Sapountzoglou, N., Lago, J., & Raison, B. (2020). Fault diagnosis in low voltage smart distribution grids using gradient boosting trees. *Electric Power Systems Research, 182* (February), Article 106254. https://doi.org/10.1016/j.epsr.2020.106254

Serin, G., Sener, B., Ozbayoglu, A. M., & Unver, H. O. (2020). Review of tool condition monitoring in machining and opportunities for deep learning. *International Journal of Advanced Manufacturing Technology, 109*(3–4), 953–974. https://doi.org/10.1007/s00170-020-05449-w

Shan Suthaharan. (2016). Machine Learning Models and Algorithms for Big Data Classification Thinking with Examples for Effective Learning. In *Integrated Series in Information Systems* (Vol. 36).

Shanmuganathan, S. (2016). Artificial neural network modelling: An introduction. In *Studies in Computational Intelligence* (Vol. 628). https://doi.org/10.1007/978-3-319-28495-8_1.

Shao, M., Wang, J., & Wang, S. (2020). The Intelligent Fault Diagnosis of Diesel Engine Based on the Ensemble Learning. *Journal of Physics: Conference Series, 1549*(4). https://doi.org/10.1088/1742-6596/1549/4/042106

Sherstinsky, A. (2020). Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) network. *Physica D: Nonlinear Phenomena, 404*, Article 132306. https://doi.org/10.1016/j.physd.2019.132306

Singru, P., Krishnakumar, V., Natarajan, D., & Raizada, A. (2018). Bearing failure prediction using wigner-ville distribution, modified poincare mapping and fast fourier transform. *Journal of Vibroengineering, 20*(1), 127–137. https://doi.org/10.21595/jve.2017.17768

Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing and Management, 45*(4), 427–437. https://doi.org/10.1016/j.ipm.2009.03.002

Stachowiak, G. P., Podsiadlo, P., & Stachowiak, G. W. (2005). A comparison of texture feature extraction methods for machine condition monitoring and failure analysis. *Tribology Letters, 20*(2), 133–147. https://doi.org/10.1007/s11249-005-8303-1

Stetco, A., Dinmohammadi, F., Zhao, X., Robu, V., Flynn, D., Barnes, M., … Nenadic, G. (2019). Machine learning methods for wind turbine condition monitoring: A review. *Renewable Energy, 133*, 620–635. https://doi.org/10.1016/j.renene.2018.10.047

Su, H., & Chong, K. T. (2007). Induction machine condition monitoring using neural network modeling. *IEEE Transactions on Industrial Electronics, 54*(1), 241–249. https://doi.org/10.1109/TIE.2006.888786

Sun, S., Sheng, Z., Jiang, W., & Li, Z. (2020). Study on the health condition monitoring method of hydraulic pump based on convolutional neural network. *Proceedings - 2020 12th International Conference on Measuring Technology and Mechatronics Automation, ICMTMA 2020*, 149–153. https://doi.org/10.1109/ICMTMA50254.2020.00041.

T M, A. M., R, J., & Gnanasekaran, S. (2019). Real time condition monitoring of hydraulic brake system using naive bayes and bayes net algorithms. *IOP Conference Series: Materials Science and Engineering, 624*, 12028. https://doi.org/10.1088/1757-899X/624/1/012028.

Tavner, P., Ran, L., & Crabtree, C. (2020). *Condition Monitoring of Rotating Electrical Machines (3rd Edition)*.

Thomas Rincy, N., & Gupta, R. (2020). Ensemble learning techniques and its efficiency in machine learning: A survey. *2nd International Conference on Data, Engineering and Applications, IDEA 2020*. https://doi.org/10.1109/IDEA49133.2020.9170675.

Trillas, E., & Eciolaza, L. (2015). Fuzzy Logic. *Springer*. https://doi.org/10.1007/978-3-319-14203-6

Tseng, C. L., Wang, S. Y., Lin, S. C., Chou, J. H., & Chen, K. F. (2014). A diagnostic system for speed-varying motor rotary faults. *Mathematical Problems in Engineering, 2014*. https://doi.org/10.1155/2014/310626

Uma Maheswari, R., Rajalingam, S., & Senthilkumar, T. K. (2020). *Condition Monitoring of Coal Mine Using Ensemble Boosted Tree Regression Model BT - Intelligent Communication Technologies and Virtual Mobile Networks* (S. Balaji, Á. Rocha, & Y.-N. Chung, Eds.; pp. 19–29). Springer International Publishing.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention Is All You Need. *CoRR, abs/1706.03762*. http://arxiv.org/abs/1706.03762.

Verma, A. K., Jain, A., & Radhika, S. (2020a). Neuro-fuzzy Classifier for Identification of Stator Winding Inter-turn Fault for Industrial Machine. In *SpringerLink* (pp. 101–110). Springer. https://doi.org/10.1007/978-981-15-4775-1_12.

Verma, A. K., Nagpal, S., Desai, A., & Sudha, R. (2020b). An efficient neural-network model for real-time fault detection in industrial machine. *Neural Computing and Applications, 0123456789*. https://doi.org/10.1007/s00521-020-05033-z

Vijay, G. S., P, S. P., & Sriram, N. S. (2011). *Artificial Neural Network based Condition Monitoring of Rolling Element Bearing. February 2016*.

Wang, H., Ma, C., & Zhou, L. (2009). A brief review of machine learning and its application. *Proceedings - 2009 International Conference on Information Engineering and Computer Science, ICIECS 2009*, 9–12. https://doi.org/10.1109/ICIECS.2009.5362936.

Wang, H., Nunez, A., Liu, Z., Zhang, D., & Dollevoet, R. (2020). A Bayesian Network Approach for Condition Monitoring of High-Speed Railway Catenaries. *IEEE Transactions on Intelligent Transportation Systems, 21*(10), 4037–4051. https://doi.org/10.1109/TITS.2019.2934346

Wang, Z., Willett, P., Aguiar, P., & Webster, J. (2001). Neural network detection of grinding burn from acoustic emission. *International Journal of Machine Tools and Manufacture, 41*, 283–309. https://doi.org/10.1016/S0890-6955(00)00057-2

Wani, M. A., Khoshgoftaar, T. M., & Palade, V. (2021). *Deep Learning Applications. Volume 2*. (Vol. 1232).

Widodo, A., & Yang, B. S. (2007). Support vector machine in machine condition monitoring and fault diagnosis. *Mechanical Systems and Signal Processing, 21*(6), 2560–2574. https://doi.org/10.1016/j.ymssp.2006.12.007

Wigner, E. P. (1997). On the Quantum Correction for Thermodynamic Equilibrium. In A. S. Wightman (Ed.), *Part I: Physical Chemistry. Part II: Solid State Physics* (pp. 110–120). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-59033-7_9.

Windmann, S., Eickmeyer, J., Badinger, J., & Niggemann, O. (2015). *Evaluation of Model-Based Condition Monitoring Systems in Industrial Application Cases.*. https://doi.org/10.1007/978-3-662-48838-6_6

Wong, S. Y., Chuah, J. H., & Yap, H. J. (2020). Technical data-driven tool condition monitoring challenges for CNC milling: A review. *International Journal of Advanced Manufacturing Technology, 107*(11–12), 4837–4857. https://doi.org/10.1007/s00170-020-05303-z

Xavier, G. M., & De Seixas, J. M. (2018). Fault Detection and Diagnosis in a Chemical Process using Long Short-Term Memory Recurrent Neural Network. *Proceedings of the International Joint Conference on Neural Networks, 2018-July*. https://doi.org/10.1109/IJCNN.2018.8489385.

Xiao, Y., Wang, H., Xu, W., & Zhou, J. (2016). Robust one-class SVM for fault detection. *Chemometrics and Intelligent Laboratory Systems, 151*, 15–25. https://doi.org/10.1016/j.chemolab.2015.11.010

Xie, Y., & Zhang, T. (2018). A Long Short Term Memory Recurrent Neural Network Approach for Rotating Machinery Fault Prognosis. *2018 IEEE CSAA Guidance. Navigation and Control Conference, CGNCC, 2018*, 1–6. https://doi.org/10.1109/GNCC42960.2018.9018860

Yiakopoulos, C. T., Gryllias, K. C., & Antoniadis, I. A. (2011). Expert Systems with Applications Rolling element bearing fault detection in industrial environments based on a K -means clustering approach. *Expert Systems With Applications, 38*(3), 2888–2911. https://doi.org/10.1016/j.eswa.2010.08.083

Yoo, Y. J. (2019). Fault Detection of Induction Motor Using Fast Fourier Transform with Feature Selection via Principal Component Analysis. *International Journal of Precision Engineering and Manufacturing, 20*(9), 1543–1552. https://doi.org/10.1007/s12541-019-00176-z

Yuan, J., Liu, L., Yang, Z., & Zhang, Y. (2020). Tool wear condition monitoring by combining variational mode decomposition and ensemble learning. *Sensors (Switzerland), 20*(21), 1–21. https://doi.org/10.3390/s20216113

Zadeh, L., & Aliev, R. (2018). *Fuzzy Logic Theory and Applications: Part I and Part II.*. https://doi.org/10.1142/10936

Žarković, M., & Stojković, Z. (2017). Analysis of artificial intelligence expert systems for power transformer condition monitoring and diagnostics. *Electric Power Systems Research, 149*, 125–136. https://doi.org/10.1016/j.epsr.2017.04.025

Zgarni, S., Keskes, H., & Braham, A. (2018). Nested SVDD in DAG SVM for induction motor condition monitoring. *Engineering Applications of Artificial Intelligence, 71* (March), 210–215. https://doi.org/10.1016/j.engappai.2018.02.019

Zhang, F., & O'Donnell, L. J. (2019). Support vector regression. *Machine Learning: Methods and Applications to Brain Disorders, Vc, 123–140*. https://doi.org/10.1016/B978-0-12-815739-8.00007-9

Zhang, W., Yang, D., & Wang, H. (2019). Data-Driven Methods for Predictive Maintenance of Industrial Equipment: A Survey. *IEEE Systems Journal, 13*(3), 2213–2227. https://doi.org/10.1109/JSYST.2019.2905565

Zhang, X.-D. (2020). Neural Networks. In *A Matrix Algebra Approach to Artificial Intelligence* (pp. 441–615). Springer Singapore. https://doi.org/10.1007/978-981-15-2770-8_7.

Zhou, Z.-H. (2012). Ensemble Methods: Foundations and Algorithms. In *Ensemble Methods: Foundations and Algorithms* (Vol. 14). https://doi.org/10.1201/b12207.