

PAPER • OPEN ACCESS

## Early sub-surface fault detection in rolling element bearing using acoustic emission signal based on a hybrid parameter of energy entropy and deep autoencoder

To cite this article: Yu Wang *et al* 2023 *Meas. Sci. Technol.* **34** 064008

View the [article online](#) for updates and enhancements.

### You may also like

- [Study of Sulfur-Based Electrodes By \*in Situ\* Characterization Techniques](#)  
Quentin Lemarie, Hassane Idrissi, Eric Maire et al.
- [Detection of Degradation of Lithium-Ion Batteries with Acoustic Emission Technique](#)  
Shogo Komagata, Naoaki Kuwata, Rangasamy Baskaran et al.
- [Fatigue crack sizing in rail steel using crack closure-induced acoustic emission waves](#)  
Dan Li, Kevin Sze Chiang Kuang and Chan Ghee Koh

# Early sub-surface fault detection in rolling element bearing using acoustic emission signal based on a hybrid parameter of energy entropy and deep autoencoder

Yu Wang<sup>1,\*</sup> , Rune Harald Hestmo<sup>2</sup> and Alexei Vinogradov<sup>1</sup>

<sup>1</sup> Department of Mechanical and Industrial Engineering, Norwegian University of Science and Technology—NTNU, N-7491 Trondheim, Norway

<sup>2</sup> Kongsberg Maritime AS, Trondheim, Norway

E-mail: [yuwa@ntnu.no](mailto:yuwa@ntnu.no)

Received 26 May 2022, revised 13 October 2022

Accepted for publication 7 March 2023

Published 27 March 2023



CrossMark

## Abstract

Bearings are a crucial component of wind turbines. The acoustic emission (AE) technique offers the advantage of earlier detection of defects and failures of bearings in comparison to traditional vibration techniques. Parameter-based analysis is the most widely used approach to interpret AE waveforms, partly due to the challenges arising in the processing of large amounts of streaming data. In this work, the AE technique is applied to monitor a run-to-failure process of a roller bearing, and it is found that the use of multiple known parameters, such as the root mean square, skewness, crest factor, impulse factor etc, fails to characterise the evolution of the acquired AE signals, thus highlighting the long-standing necessity and significance of developing new AE indicators that are more adequate to detect the failure of rotating machines. We propose a hybrid parameter—the information entropy penalty factor (IEPF)—which uses the advantages of the entropy theory and deep learning methods. The effectiveness of the proposed method has been investigated and demonstrated for roller bearing contact fatigue experiments, and the results show that IEPF can timely and accurately detect the incipient sub-surface faults.

Keywords: sub-surface fault detection, bearings, acoustic emission signal, parameter analysis, energy entropy

(Some figures may appear in colour only in the online journal)

\* Author to whom any correspondence should be addressed.



Original content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](https://creativecommons.org/licenses/by/4.0/). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

## 1. Introduction

Condition monitoring (CM) of rotating machines has been a hot topic for decades. Wind turbines are rotating machines that have evolved to become pivotal components for the generation of green energy. Wind turbines are usually installed in extreme and harsh environments and are prone to a high failure rate. Bearings are essential and highly demanding components of wind turbines. Faults in the bearings can lead to critical failures, breakdowns and consequent losses associated with the downtime of wind turbines. Therefore, CM and timely and accurate fault diagnosis offers substantial benefits to operating equipment with rolling element bearings by identifying incipient damage at as early stage as possible before the faults evolve to a critical stage. This is especially important for machines, where a fault can cause irrecoverable damage to the environment, and not least to avoid losses of human life or health.

Multiple sensing techniques have been employed in bearing CM systems used in industrial settings in general and for wind turbine CM in particular. Monitoring and trending the temperature of a bearing is a simple and cost-effective method to identify a bearing condition. However, in most applications, the temperature measurements are not sensitive enough to detect an early stage of fault development in a roller bearing. Instead, vibration analysis has become the most widespread and market-leading technology due to its simplicity, robustness and multiple uses for custom-built solutions. Various acquisition and analysis tools have been established and proven effective for vibration data. However, the vibration signals induced by tiny defects at the early stage of their development can be easily masked by the uncontrolled mechanical disturbances from the rotating machine. Moreover, the vibration acceleration signals can go undetected in heavy or slowly rotating structures until the fault increases significantly to a large (detectable) scale, by which stage it is often too late for preventive/corrective maintenance and is close to a catastrophic failure. As opposed to vibrations, acoustic emissions (AEs) reflecting the dynamics of the sources evolving under load can be generated even by microscopic flaws, such as breaks of hard non-metallic inclusions, incipient cracks, etc [1]. Moreover, the AE signal tends to increase with the growing scale of the sources. Therefore, the potential of the AE technique for early fault detection enjoys growing recognition in the industrial domain. AE methods have become an important companion of reliable monitoring systems when the impact of wear and friction of rotating components is of concern. AE is commonly defined as a phenomenon whereby transient elastic waves are spontaneously emitted by the rapid stress relaxation within localised sources in material under load. Plastic deformation and fracture associated with the nucleation and growth of cracks represent the primary mechanisms of the sources releasing the elastic strain energy associated with AE transients [2]. In contrast with the vibration signal, the sources generating AE signals are characterised by a much wider frequency range (100 kHz and 1 MHz) [3], which does not overlap significantly with low-frequency mechanical vibration signals caused by imbalance or misalignment of machine

components [4, 5]. A great deal of evidence has been accumulated, suggesting that AE parameters can reveal the faults in rotating equipment before they show up in the vibration acceleration range. Since the early work by Yoshioka *et al* [6], these results have been investigated and confirmed in abundant literature over the last 30 years; see [2, 7, 8] for examples.

With the advent of artificial intelligence, machine learning methods have become more and more extensively applied in the field of fault diagnosis. Deep learning (DL) technology is the most prominent branch of machine learning (ML) methodology, and refers explicitly to artificial neural networks with a multi-layered architecture. A number of DL architectures, such as convolutional neural networks (CNNs) [9, 10], long short-term memory (LSTM) [11] and autoencoder [12, 13], have been applied in the CM field and demonstrated outstanding potential and practicality. However, most of the relevant methods developed in this field are based on artificially seeded defects and supervised circumstances. In the reality of the run-to-failure scenario, only the data characterising the ‘healthy’ status of the object under inspection are accessible before the emergence of the faults. Hence, the detection of defect initiation is fundamentally an unsupervised task, and relevant studies are still scarce. Although some unsupervised DL architectures have been developed, e.g. the stacked autoencoder, deep belief network and deep Boltzmann machine, they are mainly employed with only an auxiliary role on supervised subjects; this is generally followed by a supervised model or an extra fine-tuning procedure, as proposed in [12–15]. Aiming at early fault detection, Lu *et al* proposed a DL-based architecture comprising three network blocks—a basic autoencoder, a feature extraction layer and an LSTM-based autoencoder [16]. Autoencoder is a prevalent unsupervised DL model designed to reconstruct its own input data with the learning objective to minimise the reconstruction error. It is reasonable that the reconstruction error can indicate emerging faults. Since the acquired signal may suffer serious distortion during the run-to-failure process, it can be foreseen that autoencoder will be unable to reconstruct the input correctly, thus leading to increasing reconstruction errors, which serves as a fault indicator; see [17, 18]. The above-cited works are based on vibration data. The application of AE has yet to be tested. In addition, one needs to bear in mind that the tolerance of the neural network to small variations in the data may limit the effectiveness of the entire ML-based approach. Thus, the application of DL models to early fault detection in the run-to-failure process faces serious challenges.

Up to now, the vast majority of existing studies deal with vibration signals, while attempts to pair DL methods and the AE technique are still limited. The parameter-based methods still dominate the philosophy of the AE waveform analysis. Therefore, the relevance of the involved parameters strongly affects the performance of the detectors. The conventional AE features extracted from AE waveforms include, but are not limited to, AE hit parameters such as counts, duration, rise time, counts to peak, amplitude, etc [2, 19, 20], as well as statistical parameters/features such as root mean square (RMS), kurtosis, crest factor, skewness, etc [5, 7, 21], defined in the time domain. In addition, multiple signal processing

techniques involve spectral decomposition techniques, such as Fourier transformation [22], wavelet analysis [23–25], variational mode decomposition [26], etc, to assess the AE signal in the frequency and time–frequency domains. As is commonly seen in the general statistical analysis of random data, different features characterise the AE waveforms from different angles, thus providing a great variability in features, as well as a range of strategies and options for their analysis, interpretation and decision-making.

Raw pseudo-AE waveforms harvest a wealth of mechanical interactions from rotating components, splashing oil, electrical interferences and multiple other noise-like sources of unknown origin. Therefore, the AE signal represents periodical patterns arising in response to the roller movement. To characterise the periodicity and its disturbance embedded in AE signals, a hybrid parameter that combines DL and the information entropy (IE) theory is introduced in this work. As a natural measure of uncertainty and chaos, IE provides new insight into the underlying AE process. There is no standard way to acquire IE from the AE signal. Elforjani and Mba [27] adopted the probabilities of AE events in a given AE signal to obtain the IE value. They showed that IE was more sensitive and representative than the kurtosis and crest factor. Amiri *et al* calculated the AE entropy based on counts [28]. Kahirdeh *et al* proposed three similar IE models using AE counts, accumulated counts or the estimated histogram of the AE signal [29]. However, these methods commonly suffer from shortcomings associated with the AE hit (and corresponding parameters) definition depending heavily on the present amplitude threshold, which introduces irrecoverable uncertainty in low-amplitude and/or overlapping signal detection. Thus, the early AE events are hard to identify because the AE signals caused by incipient faults are usually of low amplitude and can be masked by strong noise. This conclusion concurs with the literature review provided in [30]. Several studies have been proposed to obtain IE from the histogram of the AE signal, as documented in [29–31].

The main contributions of this work can be summarised as follows. (1) To investigate the capacity of the AE technique in sub-surface fault detection of bearings, a laboratory durability test of a roller bearing element was carried out; roller contact fatigue damage was initiated under controlled conditions, and the accompanying AE waveforms were acquired. (2) Aimed at detecting the emerging faults timely and accurately, a health indicator combining the IE theory and autoencoder was proposed to describe the evolution of AE waveforms during the run-to-failure process, which is referred to as the information entropy penalty factor (IEPF). (3) The proposed parameter is demonstrated to be more sensitive to the periodicity and disturbance in the AE signal. (4) The high sampling frequency of AE technology limits the application of DL methods; thus, a moving variance window (MVW) was utilised to reduce the dimensions of raw AE signals. Then, autoencoder was applied to denoise the signal for feature augmentation.

The rest of the paper is organised as follows. The mathematical details of the proposed method are unfolded in section 2. The test rig and the implementation details of the proposed

method are introduced in section 3, along with the experimental results and discussion. Conclusions are formulated in section 4.

## 2. Methodology

### 2.1. Basic theory of autoencoder

Since it is an unsupervised task to detect the onset of early fault during the durability test to failure, a prevalent unsupervised network architecture—autoencoder—is chosen for the present work. The theoretical background of the involved neural network architectures is presented in the following sub-sections. Autoencoder aims at reconstructing its own input data. The basic form of the autoencoder is relatively simple—it is a symmetrical three-layer neural network consisting of input, hidden and output layers representing an encoder and decoder pair. For a given dataset  $\mathbf{X}$ , the mathematical details for encoding and decoding are represented as follows:

$$\text{Encoder: } \mathbf{H} = \text{Activ} (W_e^T * \mathbf{X} + b_e) \quad (1)$$

$$\text{Decoder: } \tilde{\mathbf{X}} = \text{Activ} (W_d^T * \mathbf{H} + b_d) \quad (2)$$

where  $W_e^T$  and  $W_d^T$  stand for the weights of the encoder and decoder, respectively, and  $b_e$  and  $b_d$  are the corresponding biases. The encoder is regarded as a feature extractor, and the output  $\mathbf{H}$  is the latent representation containing the main information of the input data.  $\tilde{\mathbf{X}}$  is the reconstructed data that is decoded from the latent representation  $\mathbf{H}$ .

To minimise the distance between  $\mathbf{X}$  and its reconstruction  $\tilde{\mathbf{X}}$ , the mean square error (MSE) loss function  $J_{\text{MSE}}$  is generally used, which is expressed as

$$J_{\text{MSE}} = \frac{1}{n} \sum_{i=1}^n \left( \frac{1}{2} \|X_i - \tilde{X}_i\| \right)^2 \quad (3)$$

where  $n$  denotes the total number of samples.

### 2.2. Energy entropy (EE)

The EE is a measure of IE, which is based on the change in the energy of the signal. A moving energy window (MEW) is applied to slide over the signal to construct the probability distribution of the energy. Given a recorded AE signal  $\mathbf{X}$ , the moving window is defined as

$$\text{win}_{k,l,s}^{\mathbf{X}} = \mathbf{X} \left[ x_{\text{start}}^{k,s} : x_{\text{start}}^{k,s} + l - 1 \right] \quad (4)$$

$$x_{\text{start}}^{k,s} = (k - 1) \times s + 1, k = 1, 2, \dots, n_k \quad (5)$$

where  $\text{win}_{k,l,s}^{\mathbf{X}}$  represents the area of the signal  $\mathbf{X}$  covered by the moving window;  $k$ ,  $l$  and  $s$  are integers specifying the moving step, window length and moving stride, respectively;  $x_{\text{start}}^{k,s}$  is the start point of the window on the signal  $\mathbf{X}$ . The energy of the overlaid region is extracted at each moving step. The total

number of moving steps is  $n_k = [(N - l) / s] + 1$ , where  $N$  is the length of the recorded AE signal. Therefore, the partial energy of the signal  $E_k$  and its probability distribution  $P_k$  are obtained as

$$E_k = \sum_{x \in \text{win}_{k,l,s}^x} |x|^2 \quad (6)$$

$$P_k = \frac{E_k}{\sum_{k=1}^{n_k} E_k}. \quad (7)$$

The MEW length is recommended to contain information about at least one entire axel revolution of the rotating machine. Thereby, it is determined by the lowest axle rotation frequency and sampling frequency. With the probability distribution, the EE is acquired based on Shannon’s entropy formula:

$$H = - \sum_{k=1}^{n_k} P_k \log_b P_k. \quad (8)$$

The logarithmic base ‘ $b$ ’ defines the unit of the measured information. The units include bits ( $b = 2$ ), nats ( $b = e$ ), and bans ( $b = 10$ ) [32]. In the case of  $P_k = 0$ , the value of  $0 \log_b 0$  is taken to be 0; therefore, the minimum value of entropy is 0.

### 2.3. The proposed method

In this paper, a new fault indicator combining the EE and reconstruction error of autoencoder is proposed, which is referred to as IEPF. The details are presented below.

**2.3.1. Feature augmentation.** To capture transient changes within the signal, an MVW, which calculates the sample variance, is applied to the original signals. For a recorded AE signal  $\mathbf{X}$ , the procedure is formulated as follows:

$$X_{MVW} = \frac{1}{l} \sum_{x \in \text{win}_{k,l,s}^{X_{MVW}}} |x - \mu|^2 \quad (9)$$

where  $\mu$  is the mean of  $x$  within  $\text{win}_{k,l,s}^{X_{MVW}}$ . The function of the MVW is to capture the transient events and highlight some important detailed information about the data. The output is a dimensionless number that measures the dispersion of the data, and thereby, the signal is de-dimensionalised. Additionally, the dimension of the original signal is largely reduced through this process, which makes it easier to be processed by the neural network.

Then, autoencoder is employed in this work to denoise and enhance the main features of the signal. With the target of reconstructing its own input data, autoencoder has been widely used for feature extraction. However, the reconstruction error is inevitable, which compels the network to outline the main features of the input data and neglect some redundant noise. To better reconstruct the input data, the CNN architecture is used to extract detailed information. The applied autoencoder architecture is shown in figure 1.

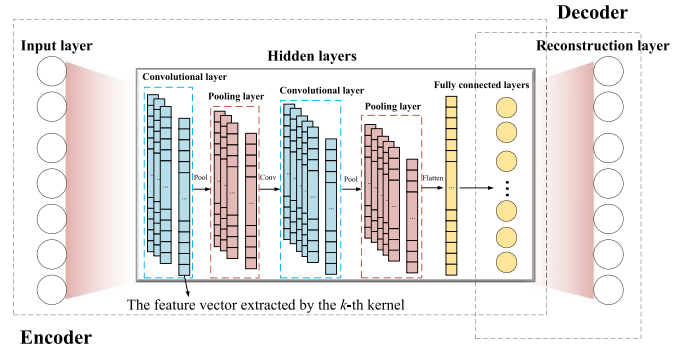


Figure 1. The neural network architecture used in the present work.

**2.3.2. IEPF.** During the run-to-failure process, the acquired AE signal may experience serious changes with the damage propagation through the test piece. It can be foreseen that autoencoder will eventually be unable to reconstruct the deformed signal and cause the MSE value to increase. Several researchers proposed the reconstruction error of autoencoder as an indicator of an early fault or anomaly in the mechanical behaviour of the system [16–18]. However, the neural network has a certain tolerance to waveform changes; i.e. if the discrepancy between signals is not very large, autoencoder can still fit the data with a relatively low reconstruction error. Since MSE is inevitable, the information provided by the reconstructed data itself is incomplete. Therefore, the EE is further adapted to utilise the advantages of autoencoder. To implement this coupling, equation (6) is rewritten as

$$\tilde{E}_k = \sum_{x \in \text{win}_{k,l,s}^{X_{MVW}}} |\widetilde{x}_{MVW} + \delta|^2 \quad (10)$$

where  $\widetilde{x}_{MVW}$  stands for the reconstruction of data covered by  $\text{win}_{k,l,s}^{X_{MVW}}$ , and  $\delta$  is the corresponding MSE value. By plugging  $\tilde{E}_k$  into equation (7), a new probability distribution of the reconstructed data is adopted as  $\tilde{P}_k$ .

The maximum value of Shannon’s entropy is obtained when all elements are identically distributed, i.e. when the dynamic range of  $H$  is limited to  $[0, -\sum_{k=1}^{n_k} \frac{1}{n_k} \log_b \frac{1}{n_k}]$ . The maximum value is greater than 0, increasing monotonically over the range of  $n_k$ . Theoretically, the greater the disturbance of the AE signal, the smaller the value of Shannon’s entropy. However, for the sake of convenience, one can modify the entropy calculation in such a way that it will increase with the variation in the AE signal as

$$IEPF = \sum_{i=1}^{n_k} \left( \tilde{P}_k \log_2 \tilde{P}_k - \frac{1}{n_k} \log_2 \frac{1}{n_k} \right). \quad (11)$$

If the IEPF value is close to 0, it indicates that the signal has strong periodicity and vice versa.



**2.3.3. General procedures.** The general fault detection procedures are summarised as the following five steps:

**Step 1:** Data acquisition. AE sensors are installed on the testing rig machine; the machine runs to failure of the rolling bearing element, and the AE waveform is recorded periodically. The acquired AE signals at the initial stage of the experiment serve as training data, and are utilised for training autoencoder and constructing a fault threshold. The rest of the data are the testing set.

**Step 2:** Signal pre-processing. Two moving windows are applied to the acquired AE signal.

- (1) MVW is first applied to the original signal for dimension reduction and extracting detailed information.
- (2) MEW is applied to the data; instead of extracting the energy feature at each moving step immediately, the covered data are prepared as the input of autoencoder.

**Step 3:** Data reconstruction. The processed signals are fed into autoencoder for denoising.

**Step 4:** Obtain health indicator. Calculate the energy probability distributions of the reconstructions. A fault alarm threshold is constructed on the basis of the IEPF values calculated from the training data.

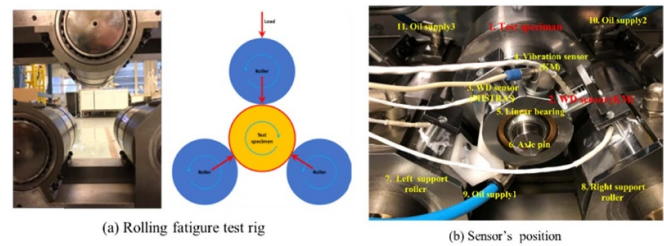
**Step 5:** Decision making. The IEPF value of the testing set exceeding the threshold is considered a fault alarm.

### 3. Experiment and discussion

#### 3.1. Test rig and data acquisition

To monitor the rolling contact fatigue phenomenon occurring in a roller bearing element, a run-to-failure test was carried out using an instrumented special-purpose testing rig designed at SINTEF Industry (Trondheim, Norway). The experimental setup is schematically illustrated in figure 2(a). The test specimen (central roller) is supported by three rollers, and each roller is supported by two needle bearings SKF NA 6914-zw. The wideband differential (WD) sensors (MISTRAS, USA) (only one sensor was used during the test) were connected to the data acquisition system as displayed in figure 2(b). The signal was amplified by 40 dB in the frequency band 20–1200 kHz by a 2/4/6 low-noise preamplifier (MISTRAS, USA).

The AE recording started automatically when the axle rotation frequency was greater than the threshold. After warming up to  $47 \pm 2$  °C, the initial axle rotation frequency was set at 364 rpm at the initial load of 67.1 kN, corresponding to 1807 MPa contact stress. The test was interrupted periodically, as is indicated by the vertical lines in the test diagram represented in figure 6 for ultrasonic inspections performed with an Olympus OMNISCAN SX phase array ultrasonic scanner (PAUT). As the PAUT inspections revealed no faults after initial cycling up to approximately  $3 \times 10^6$  cycles ( $10^6$  axel rotations), the load gradually increased in a stepwise manner up to 91.3 kN (2002 MPa contact stress). The cumulative number of fatigue cycles reached  $2.7 \times 10^7$  cycles. Excessive vibrations were detected in the machine at this load when running



**Figure 2.** Rolling fatigue test rig: photographic image and schematics of the geometry of supporting rollers and the testing roller (a), and a close-up view of the setup instrumented with AE sensors (b).

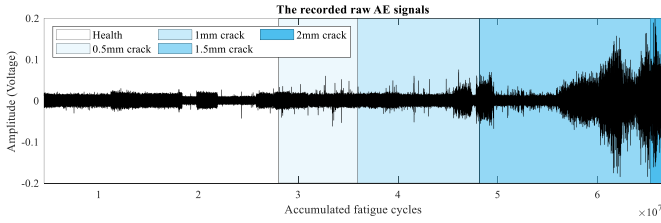
**Table 1.** Number of AE records for different stages of damage propagation.

Health condition	Number of records	Number of fatigue cycles
No damage	542	$2.8 \times 10^7$
0.5 mm crack	377	$3.6 \times 10^7$
1 mm crack	809	$4.8 \times 10^7$
1.5 mm crack	718	$6.5 \times 10^7$
2 mm crack	25	$6.6 \times 10^7$

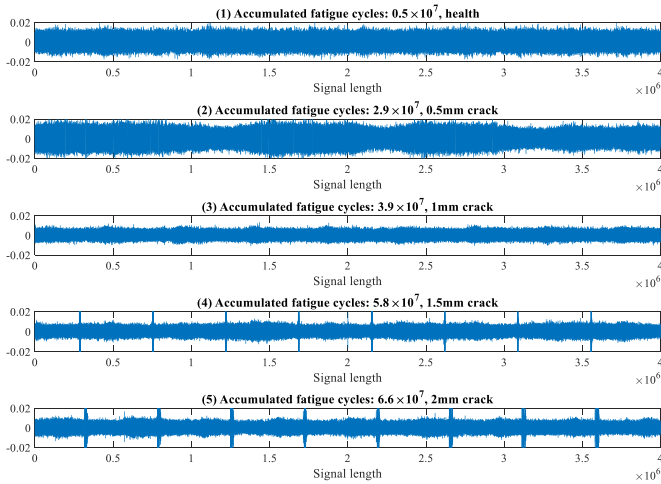
at a rotational speed of 364 rpm. Therefore, the axle rotation frequency was reduced to 256 rpm until the end of the test. The test was continued with a 91.3 kN load, and the first sub-surface crack that was detected by the PAUT was after  $2.8 \times 10^7$  fatigue cycles at approximately 4 mm below the contact surface. The smallest detected crack was estimated to be 0.5 mm long. The continued regular PAUT inspections revealed continuous slow crack growth in the longitudinal direction up to 2 mm length along the roller axis before the test was terminated. The test roller was then sectioned for metallographic inspection and verification of the PAUT results. As predicted by PAUT, three sharp fatigue cracks were observed beneath the surface.

The AE waveforms were continuously recorded at a 2 MHz sampling frequency for 2 s per record using the Kongsberg HSIO-100-A high-speed acquisition module. At the beginning of the test, AE streams were collected every 60 min. After the confirmation of the first sub-surface crack, the time interval between the successive AE acquisitions was reduced to 20 min. In total, 2471 records were qualified for the analysis. The number of records corresponding to different stages of crack growth is presented in table 1. The records are indexed from 1 to 2471 according to the time of acquisition. The recorded raw AE signals are plotted in figure 3 for illustration. An appreciable change in the AE amplitude is first observed after  $4.6 \times 10^7$  fatigue cycles. Ultrasonic inspections revealed a crack of 1 mm at this stage.

Several randomly chosen AE records, which are typically observed during different stages of the damage propagation, are shown in figure 4. The evolution of the AE waveforms can be observed. At first, the AE waveforms exhibited evident periodical characteristics as shown in figure 4(1) due to the routine operation of the rotating machine. After the fatigue cycle was



**Figure 3.** Diagram showing the raw AE signals against the number of fatigue cycles.



**Figure 4.** Examples of randomly chosen representative AE records at different stages of damage propagation.

accumulated to  $2.8 \times 10^7$ , the initial periodic behaviour in the waveforms disappeared—the effect is assumed to be related to the generation of AE signals from the defect. The continuous background quasi-steady AE signal is assumed to be produced primarily by the over-rolling of roughness asperities [23]. If a defect forms on either the surface or in the bulk of the test roller, and if the sensor successfully captures it, the two random processes—the background noise and the defect-induced AE—overlap additively. The AE associated with the defect can be considered as a disturbance, distorting the waveform of the original signal. Assuming that the AE signal recorded from the healthy stage is  $X$ , the AE signal reflecting fatigue damage is denoted as  $\bar{X} = X + \tau$ , where  $\tau$  represents the AE response to the defect that emerged in the roller. This type of signal appears as illustrated in figure 4(2–3). With the propagation of the fault, the peak amplitude of the corresponding AE burst signal clearly exceeds the noise threshold at periodic intervals, as shown in figure 4(4–5). The triple roller arrangement shown in figure 1 assumes that each point of the test roller interacts with the support rollers three times per revolution. When the test roller containing the surface (or sub-surface) faults contacts the support roller, the stress concentration in the bearing elements along the defect boundary is expected to cause an increase in the released elastic energy [2], resulting in periodic spikes in the AE waveforms.

**Table 2.** Autoencoder architecture.

Layer	Key parameters	Output size	Activation function
Input	/	$1 \times 1024$	/
C1	$1 \times 25@16$	$1 \times 1000@16$	ReLU
P1	$1 \times 5@16$	$1 \times 200@16$	/
C2	$1 \times 16@32$	$1 \times 185@32$	ReLU
P2	$1 \times 5@32$	$1 \times 37@32$	/
FC1	/	$1 \times 1184$	/
FC2	$1184 \times 100$	$1 \times 100$	ReLU
Output	/	$1 \times 1024$	Sigmoid

*Note:* C, P and FC denote the convolutional layer, pooling layer and fully connected layer, respectively. The notation ' $a \times b@c$ ' describes the kernel size and the output size, where  $a$  and  $b$  represent the row and column of the matrix, and  $c$  denotes the number of channels.

### 3.2. Implementation details

The MEW should contain information about at least one entire axel revolution. For instance, the lowest axle rotation frequency in the present work is 254 rpm, i.e. for a 2 s recording, eight complete rotations are captured. Therefore, the moving step of MEW should be 8, and the maximum value of IEPF is calculated as 3. Based on the sampling frequency used (2 MHz), the window length and the moving stride of MVW and MEW are set at 464 and 1024, respectively.

The first 60% of healthy data (325 recorded AE fragments) were used to train the neural network and calculate the fault alarm threshold. The threshold is calculated conventionally as  $\text{mean}(x) \pm 3 \times \text{std}(x)$ , where  $x$  denotes the IEPF values of training data, and mean and std stand for the mean value and the standard deviation, respectively. Since each record contains eight complete rotations, after being processed by MVW and MEW, 2471 training samples were constructed to train autoencoder. The size of the training dataset is  $2471 \times 1024$ . The rest of the records are testing data. Both the training and testing data are normalised using the maxminmap method before feeding into autoencoder. Details of the network architecture are presented in table 2.

The reconstructed data are randomly exemplified from different stages of the experiment, as shown in figure 5. Each image represents eight stacked sub-signals covered by MEW. It is hard to identify the difference between healthy and faulty data from the raw signal by the naked eye, especially at the early damage stage featured by the 0.5 mm crack length; see, for example, indices 448, 452, 561 and 805. However, the reconstructions unveil more clear features if compared to those of the raw signal. The reconstruction error (MSE) of autoencoder is shown in figure 6. As mentioned before, autoencoder is reasonably tolerant to waveform changes; i.e. if the discrepancy between signals is not very large, the neural network can still fit the data with a low reconstruction error. As shown in figure 6, the reconstruction error of the trained network is still very low, especially at the earliest crack growth stage. The drastic increase in the reconstruction error appears only with the emergence of AE bursts. Mathematically, this is because the sigmoid activation function maps the output to the range

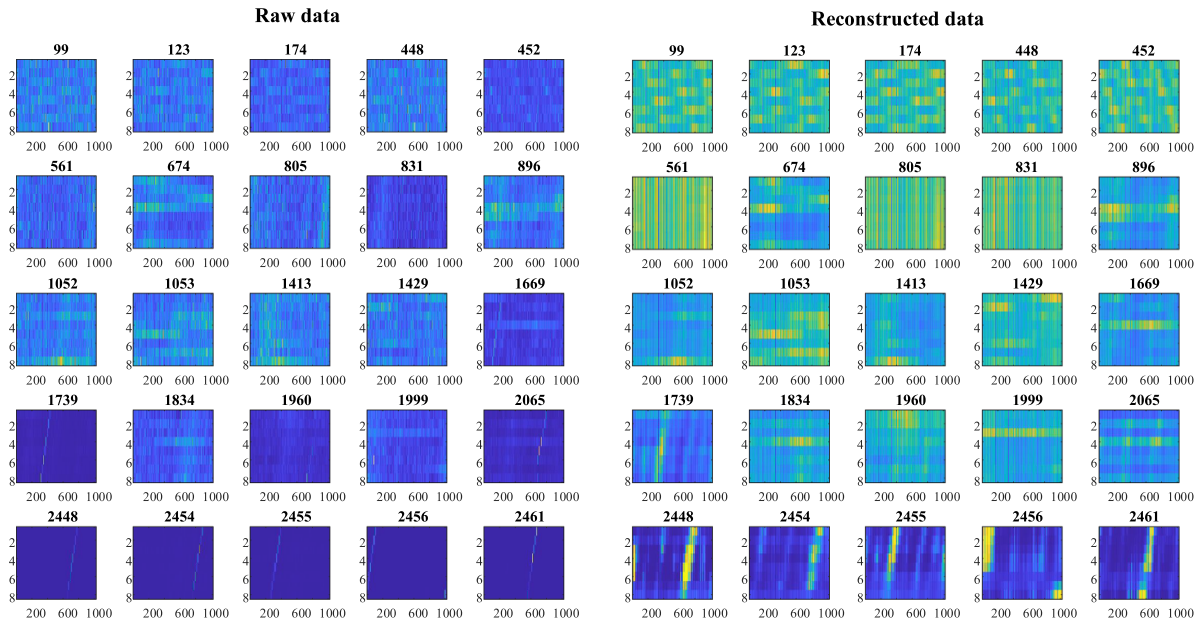


Figure 5. Visualisation of the reconstructed data from different stages of damage propagation.

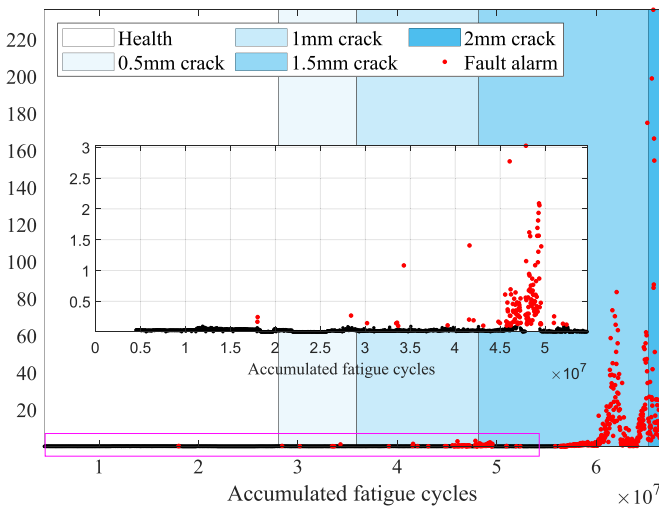


Figure 6. MSE (reconstruction error) of autoencoder; the inset shows a magnified view of MSE in the range of record indices from 0 to 2000.

of (0, 1); however, the peak of the AE burst will exceed the upper boundary of the sigmoid function without the corresponding training data. The results manifest that the MSE indicator taken alone is not sufficient to identify early faults.

### 3.3. Evaluation methods

In the following analysis, the performance of conventional statistical parameters is investigated and compared with the one proposed in this work. The quality of the probed parameters is assessed from two aspects: (1) timely and accurate detection of emerging faults and (2) better description of the AE waveform evolution. The 19 statistical parameters

Table 3. Statistical parameters extracted from time domain, frequency domain and time–frequency domain.

Domains	Parameters
Time domain	(1) Root mean square (RMS); (2) skewness; (3) kurtosis; (4) shape factor; (5) crest factor; (6) impulse factor; (7) margin factor; (8) histogram-based information entropy (IE); (9) energy entropy (EE); (10) power.
Frequency domain	(11) Mean frequency (MF); (12) RMS of frequency (RMSF); (13) standard deviation of frequency (SF).
Time–frequency domain	(14) STFT + power entropy; (15) STFT + MF entropy; (16) STFT + RMSF entropy; (17) STFT + SF entropy; (18) wavelet packet energy entropy (WPEE); (19) wavelet packet singular entropy (WPSE).

Note: (14–17) short-time Fourier transform (STFT) was implemented with a hamming window with a length equal to 4096 readings. The window slid over the original data to calculate the discrete Fourier transform of the windowed data, and the overlap of each moving step was 512. (18–19) Wavelet packet transform was applied to perform three-layer decomposition of the original AE signal using the ‘dmey’ wavelet, and results in eight decomposed frequency bands.

listed in table 3 were extracted from the time domain, frequency domain and time–frequency domain and were probed for the sake of comparison. To quantify the performance of all these indicators, the data were categorised into two classes as ‘healthy’ and ‘faulty’, and three evaluation indicators—Accuracy, Specificity and  $F_1$ -score were measured, as defined below [33]:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (12)$$

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (13)$$



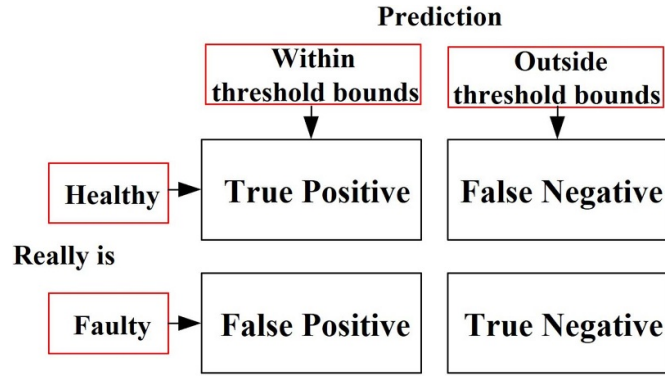


Figure 7. Definition of the confusion matrix.

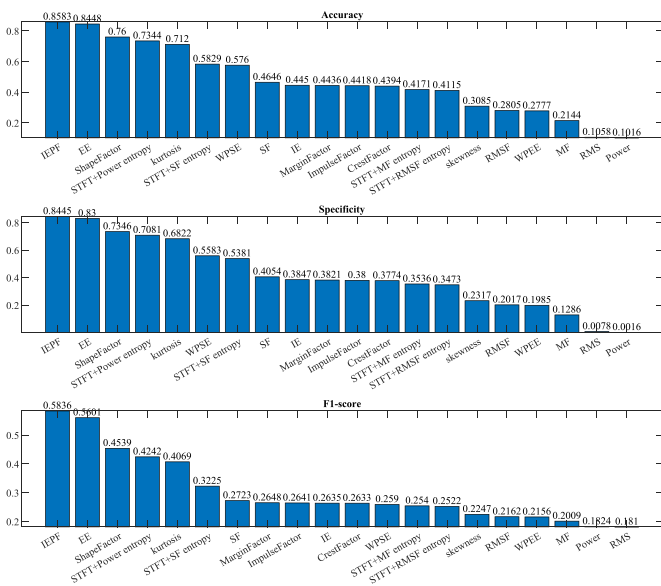


Figure 8. Accuracy, specificity and  $F_1$ -score of the probed parameters sorted in descending order.

$$F_1\text{-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (14)$$

where TP, FN, FP and TN are abbreviations of true positive, false negative, false positive and true negative, respectively, as described in figure 7. Accuracy measures all correctly classified samples. Specificity quantifies the ratio of negative class predictions of all negative samples. The  $F_1$ -score provides a single score that balances both the concerns of precision and recall. Precision and recall are defined as  $TP/TP + FP$  and  $TP/TP + FN$ , and quantify the number of correct positive results divided by all positive results and relevant samples, respectively.

Accuracy, Specificity and  $F_1$ -score of the probed parameters are compared in figure 8. Based on these three quality indicators, the proposed parameter exhibits the highest scores. IEPF generates fewer false fault alarms and more true fault alarms compared with other parameters tested. Although parameters such as RMS, skewness, crest factor, impulse factor,

etc, have been used with greater or lesser success by many researchers, in the present settings, they perform quite unsatisfactory. This prompted us to seek new reliable parameters.

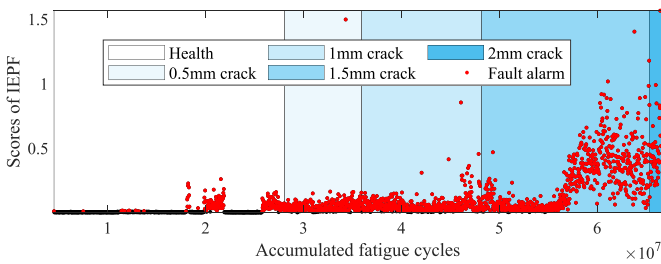
The top eight parameters performing better in Accuracy (excluding IEPF) were selected for further comparison. Table 4 presents the values of accuracy of the selected parameters at different experimental stages. The accuracy at each stage was obtained using the formula  $F/N$ , where  $N$  denotes the total number of records at a specific experimental stage, and  $F$  represents the true fault alarms at this stage. Most of the parameters indicated the AE waveform changed substantially when the crack was propagated to the mature stage with a final length of up to 2 mm. Although parameters like wavelet packet singular entropy (WPSE) and standard deviation of frequency show relatively high accuracy at the 0.5 mm crack stage, WPSE exhibits a higher rate of false alarms at the healthy stage, and standard deviation of frequency (SF) fails to detect the propagation of the fault. Since failure is an irreversible and progressively propagating process, the indicator is expected to be continuous and monotonic. Compared with other parameters, the IEPF generates notably fewer false alarms and more true fault alarms.

Scores of IEPF are plotted in figure 9 against the cumulative number of fatigue cycles. The red dots represent the events with IEPF values exceeding the threshold, which are denoted as fault alarms. One can see that IEPF transparently characterises the evolution of the recorded AE waveforms from the following aspects. First, the IEPF value corresponding to the initial healthy stage is approximately 0 (the average IEPF value of the recorded AE signal at the healthy stage is 0.0026), which indicates that the recorded AE signals present strong periodical patterns. The IEPF increases steeply in the second stage when the first 0.5 mm crack is detected. Thus, a breakpoint between the healthy and faulty stages can be easily identified. Second, the IEPF value captures the initiation of the persistent AE bursts at the intersection of the 1 mm crack and 1.5 mm crack. Additionally, the results successfully characterise the increase in the AE bursts after the 2000th record while maintaining the general trend towards higher values.

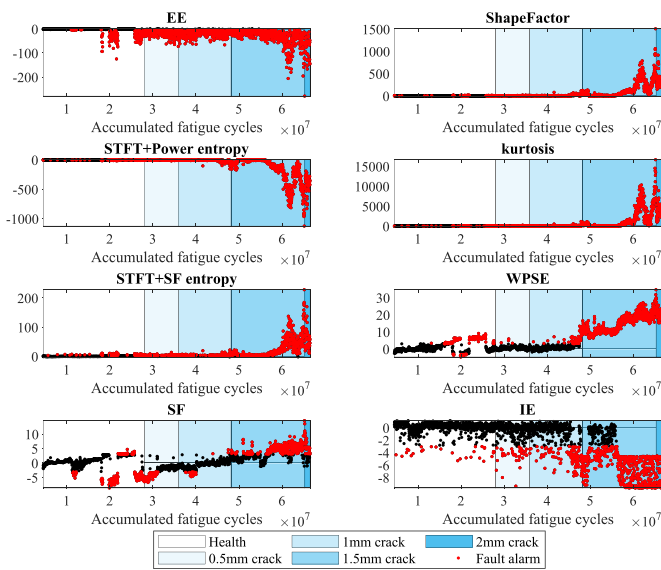
To further compare the performance of IEPF with other parameters, figure 10 shows the variation of the selected parameters from the beginning of the test to failure. Although EE

**Table 4.** The accuracy of selected parameters at different stages of damage propagation.

Parameter	Fault conditions				
	No damage	0.5 mm crack	1 mm crack	1.5 mm crack	2 mm crack
IEPF	0.9742	0.4259	0.9148	0.9791	1
EE	0.9723	0.4233	0.8938	0.9652	1
Shape factor	0.9815	0.3836	0.6642	0.9889	1
STFT + power entropy	0.9705	0.4233	0.5889	0.9819	1
Kurtosis	0.9760	0.3307	0.5691	0.9833	1
STFT + SF entropy	0.9797	0.3598	0.3370	0.8428	1
WPSE	0.8930	0.6614	0.1037	1.0000	1
SF	0.9779	0.6111	0.1938	0.5369	0.3462
IE	0.9742	0.0397	0.1889	0.7650	1



**Figure 9.** The scores of IEPF from the durability test of the roller bearing element.



**Figure 10.** The scores of selected parameters of the AE signals from the durability test of the roller bearing element.

also presents relatively high accuracy in figure 8, it fails to characterise the AE behaviour in response to the crack growth up to 1 mm and further to 1.5 mm length. Compared to other parameters, the IEPF shows excellent sensitivity to the emergence of periodical AE impulses and exhibits a clearer description of the waveform evolution corresponding to the propagation of internal fatigue cracks.

### 4. Conclusion

In this paper, a durability test of a roller bearing element was carried out to investigate the application of the AE technique to sub-surface fault detection in a roller. The experimental results show that many known parameters, such as RMS, skewness, crest factor, impulse factor etc, fail to characterise the evolution of AE signals in relation to the damage initiation and propagation. Therefore, a hybrid parameter called IEPF is proposed to assess the fault behaviour through the evolution of AE waveforms. The proposed method combines the advantages of information theory and autoencoder to achieve a high sensitivity to the periodicity and its disturbance in AE signals. Comparative tests were carried out to assess the quality of the health status indicators from two aspects: (i) timely and accurate detection of emerging faults and (ii) a more elucidative description of the AE waveform evolution in response to the emerging and propagating fatigue damage. The experimental results verify the effectiveness of the proposed data processing scheme for fault monitoring and possible diagnostics in roller bearings. The proposed methodology can be reasonably easily adapted to the CM of other rotating machines since it is driven primarily by data and does not rely on specific knowledge of the mechanical features of the system under control.

### Data availability statement

The data generated and/or analysed during the current study are not publicly available for legal/ethical reasons but are available from the corresponding author on reasonable request.

### Acknowledgments

The financial support from the Norwegian Research Council through RCN Project No. 296236 is gratefully appreciated.

### ORCID iD

Yu Wang <https://orcid.org/0000-0003-0854-6975>

## References

- [1] Wadley H and Mehrabian R 1984 Acoustic emission for materials processing: a review *Mater. Sci. Eng.* **65** 245–63
- [2] Choudhury A and Tandon N 2000 Application of acoustic emission technique for the detection of defects in rolling element bearings *Tribol. Int.* **33** 39–45
- [3] Liu Z, Yang B, Wang X and Zhang L 2021 Acoustic emission analysis for wind turbine blade bearing fault detection under time-varying low-speed and heavy blade load conditions *IEEE Trans. Ind. Appl.* **57** 2791–800
- [4] He Y and Zhang X 2012 Approximate entropy analysis of the acoustic emission from defects in rolling element bearings *J. Vib. Acoust.* **134** 061012
- [5] Hemmati F, Alqaradawi M and Gadala M S 2016 Optimized statistical parameters of acoustic emission signals for monitoring of rolling element bearings *Proc. Inst. Mech. Eng. J* **230** 897–906
- [6] Yoshioka T and Fujiwara T 1982 A new acoustic emission source locating system for the study of rolling contact fatigue *Wear* **81** 183–6
- [7] Al-Ghamd A M and Mba D 2006 A comparative experimental study on the use of acoustic emission and vibration analysis for bearing defect identification and estimation of defect size *Mech. Syst. Signal Process.* **20** 1537–71
- [8] Caso E, Fernandez-del-rincon A, Garcia P, Iglesias M and Viadero F 2020 Monitoring of misalignment in low speed geared shafts with acoustic emission sensors *Appl. Acoust.* **159** 107092
- [9] Prosvirin A, Kim J and Kim J-M 2018 Bearing fault diagnosis based on convolutional neural networks with kurtogram representation of acoustic emission signals *Advances in Computer Science and Ubiquitous Computing (Singapore)* vol 474 pp 21–6
- [10] Hasan M J, Manjurul Islam M M and Kim J-M 2019 Acoustic spectral imaging and transfer learning for reliable bearing fault diagnosis under variable speed conditions *Measurement* **138** 620–31
- [11] Shi H, Guo L, Tan S and Bai X 2019 Rolling bearing initial fault detection using long short-term memory recurrent network *IEEE Access* **7** 171559–69
- [12] Li X, Li J, He D and Qu Y 2019 Gear pitting fault diagnosis using raw acoustic emission signal based on deep learning *Eksplot. Niezawodn.* **21** 403–10
- [13] Zhiyi H, Haidong S, Lin J, Junsheng C and Yu Y 2020 Transfer fault diagnosis of bearing installed in different machines using enhanced deep auto-encoder *Measurement* **152** 107393
- [14] He Z, Shao H, Wang P, Lin J, Cheng J and Yang Y 2020 Deep transfer multi-wavelet auto-encoder for intelligent fault diagnosis of gearbox with few target training samples *Knowl.-Based Syst.* **191** 105313
- [15] Zhu H, Cheng J, Zhang C, Wu J and Shao X 2020 Stacked pruning sparse denoising autoencoder based intelligent fault diagnosis of rolling bearings *Appl. Soft Comput.* **88** 106060
- [16] Lu W, Li Y, Cheng Y, Meng D, Liang B and Zhou P 2018 Early fault detection approach with deep architectures *IEEE Trans. Instrum. Meas.* **67** 1679–89
- [17] Dwiputranto T H, Setiawan N A and Aji T B (eds) 2017 Machinery equipment early fault detection using artificial neural network based autoencoder *2017 3rd Int. Conf. on Science and Technology-Computer (ICST)* (IEEE) (<https://doi.org/10.1109/icstc.2017.8011854>)
- [18] König F, Sous C, Chaib A O and Jacobs G 2021 Machine learning based anomaly detection and classification of acoustic emission events for wear monitoring in sliding bearing systems *Tribol. Int.* **155** 106811
- [19] Caesarendra W, Kosasih B, Tieu A K, Zhu H, Moodie C A and Zhu Q 2016 Acoustic emission-based condition monitoring methods: review and application for low speed slew bearing *Mech. Syst. Signal Process.* **72** 134–59
- [20] Mano R, Yoshioka T, Miti A K and Yamamoto T 2000 Relationship between growth of rolling contact fatigue cracks and load distribution *Tribol. Trans.* **43** 367–76
- [21] Morhain A and Mba D 2003 Bearing defect diagnosis and acoustic emission *Proc. Inst. Mech. Eng. J.* **217** 257–72
- [22] Wang Z, Chegdani F, Yalamarti N, Takabi B, Tai B, El Mansori M and Bukkapatnam S 2020 Acoustic emission characterization of natural fiber reinforced plastic composite machining using a random forest machine learning model *J. Manuf. Sci. Eng.* **142** 031003
- [23] Bianchi D, Mayrhofer E, Gröschl M, Betz G and Vernes A 2015 Wavelet packet transform for detection of single events in acoustic emission signals *Mech. Syst. Signal Process.* **64** 441–51
- [24] Gao Z, Lin J, Wang X and Xu X 2017 Bearing fault detection based on empirical wavelet transform and correlated kurtosis by acoustic emission *Materials* **10** 571
- [25] Liu D, Tao J, Luo A and Wang Q (eds) 2018 An optimized kurtogram method for early fault detection of rolling element bearings using acoustic emission *2018 Int. Conf. on Information Systems and Computer Aided Education (ICISCAE)* (IEEE) (<https://doi.org/10.1109/icisca.2018.8666929>)
- [26] Liu L, Chen L, Wang Z and Liu D 2020 Early fault detection of planetary gearbox based on acoustic emission and improved variational mode decomposition *IEEE Sens. J.* **21** 1735–45
- [27] Elforjani M and Mba D 2010 Accelerated natural fault diagnosis in slow speed bearings with acoustic emission *Eng. Fract. Mech.* **77** 112–27
- [28] Amiri M, Modarres M and Droguett E L (eds) 2015 AE entropy for detection of fatigue crack initiation and growth *Austin 2015 IEEE Conf. on Prognostics and Health Management (PHM)* (IEEE) (<https://doi.org/10.1109/ICPHM.2015.7245038>)
- [29] Kahirdeh A, Sauerbrunn C and Modarres M 2016 Acoustic emission entropy as a measure of damage in materials *AIP Conf. Proc.* vol 1757 060007
- [30] Chai M, Zhang Z and Duan Q 2018 A new qualitative acoustic emission parameter based on Shannon's entropy for damage monitoring *Mech. Syst. Signal Process.* **100** 617–29
- [31] Karimian S F, Modarres M and Bruck H A 2020 A new method for detecting fatigue crack initiation in aluminum alloy using acoustic emission waveform information entropy *Eng. Fract. Mech.* **223** 106771
- [32] Shannon C E 1948 A mathematical theory of communication *Bell Syst. Tech. J.* **27** 379–423
- [33] Amin H U, Mumtaz W, Subhani A R, Saad M N M and Malik A S 2017 Classification of EEG signals based on pattern recognition approach *Front. Comput. Neurosci.* **11** 103