Entropy Measures in Machine Fault Diagnosis: Insights and Applications

Zhiqiang Huo[1], Miguel Martínez-García[2], Yu Zhang[3*], Ruqiang Yan[4], Lei Shu[5,1]
[1]School of Engineering, University of Lincoln, Lincoln, UK
[2]Wolfson School of Mechanical, Electrical and Manufacturing Engineering, Loughborough University, Loughborough, UK
[3]Department of Aeronautical and Automotive Engineering, Loughborough University, Loughborough, UK
[4]School of Mechanical Engineering, Xi'an Jiaotong University, China
[5]NAU-Lincoln Joint Research Center of Intelligent Engineering, Nanjing Agricultural University, China
Email: {zhuo, lshu}@lincoln.ac.uk, {M.Martinez-Garcia, Y.Zhang}@lboro.ac.uk, rqyan@ieee.org

***Abstract*—Entropy, as a *complexity measure*, has been widely applied for time series analysis. One preeminent example is the design of machine condition monitoring and industrial fault diagnostic systems. The occurrence of failures in a machine will typically lead to non-linear characteristics in the measurements, caused by instantaneous variations, which can increase the complexity in the system response. Entropy measures are suitable to quantify such dynamic changes in the underlying process, distinguishing between different system conditions. However, notions of entropy are defined differently in various contexts (e.g., information theory and dynamical systems theory), which may confound researchers in the applied sciences. In this paper, we have systematically reviewed the theoretical development of some fundamental entropy measures and clarified the relations among them. Then, typical entropy-based applications of machine fault diagnostic systems are summarized. Further, insights into possible applications of the entropy measures are explained, as to where and how these measures can be useful towards future data-driven fault diagnosis methodologies. Finally, potential research trends in this area are discussed, with the intent of improving online entropy estimation and expanding its applicability to a wider range of intelligent fault diagnostic systems.**

***Index Terms*—Entropy, Fault diagnosis, Rotating machinery.**

I. Introduction

Engineering machinery in modern industries is usually operated in complex, and often harsh, environments. It is of paramount importance to ensure safe and reliable system operation. As a result, fault diagnosis is essential to detect and identify potential failures as early as possible; so that necessary machine maintenance can be performed to troubleshoot faults, and performance degradation can be minimized. Commonly, the dynamic response of a system, due to a change of state, is reflected in the sensor measurements. By monitoring the consistency between these measurements and the machine operational regime, it is possible to predict the operating status of the machine and potential faults.

Given a system or process, be it natural or man-made, its evolution can be followed by a finite amount of measurements. A subject of special interest is how to analyze these measurements - such as vibration and acoustic signals - to monitor and diagnose different machine conditions in the system. In this paper, non-linear time series complexity analysis is surveyed from the perspective of entropy measures and their particular application to machine fault diagnosis.

Entropy has been a transcendental and pervasive concept in numerous disciplines, ranging from logic and physics to biology and engineering. Although entropy has been studied since the nineteenth century, it still attracts interest – due to its flexibility and applicability into different contexts, and to the multiple interpretations of its implications [1]. Entropy links the notions of disorder and uncertainty with physical states – which are interpreted as information communication channels. The topic examined here corresponds to the case when these *channels* are the sensors monitoring the responses of a system or machine.

Historically, entropy arose after the invention of the heat engine, through pioneering research towards clarifying thermodynamical processes and increasing the efficiency of such machines [2]. This research led to the formulation of the *Second Law of thermodynamics*, which reveals that entropy of an isolated system can never decrease over time. Later, Ludwig Boltzmann and Josiah Gibbs independently interpreted the definition of entropy as a measure of the number of states that a physical system can adopt from a molecular perspective, giving rise to *statistical mechanics* [3]. They observed that macrostates with a higher number of possible microstates are more likely and exhibit larger entropy. More importantly, Gibbs revealed that entropy could be described in terms of statistical quantities, such as probabilities and their logarithms – setting the path towards the usage of entropy as a tool for non-linear signal analysis, which is the broader theme of the literature reviewed in this paper.

Subsequent research by Hartley, Wiener and Shannon resulted in the introduction of a parallel entropy formulation, which lies at the center of information theory – known as *information entropy* or *Shannon entropy* (ShanEn) [4]. ShanEn was proposed to quantify the amount of information content conveyed by messages from an information source [5]. It

Corresponding author: Yu Zhang (email: Y.Zhang@lboro.ac.uk). This research is supported by EPSRC Grant EVES (EP/R029741/1).

interprets the uncertainty and randomness of the *system's events* – i.e., its behavior – from a probability viewpoint. That is, the examined system is understood as a random variable. Inspired by ShanEn, various concepts of entropy were later developed within complexity theory, particularly in the study of dynamical systems. One example is the Kolmogorov-Sinai (KS) entropy measure. Since its introduction, many studies have attempted to estimate KS entropy for practical use, among which the Eckmann-Ruelle entropy can potentially be implemented in experimental cases [6].

Motivated by Eckmann-Ruelle entropy, some other fundamental entropy measures have been developed for time-series complexity analysis. For instance, Approximate Entropy (ApEn) was constructed to be thematically similar to the KS entropy, and it is based on the Eckmann-Ruelle entropy [7]. ApEn estimates dynamical changes in time series by characterizing the underlying deterministic or stochastic components. Later, Sample Entropy (SampEn) [8] and Fuzzy Entropy (FuzzyEn) [9] were proposed as improvements of ApEn for entropy estimation. Besides, Permutation Entropy (PerEn) was put forth by Bandt and Pompe to measure symbolic dynamic changes that are encoded in ordinal patterns in time series [10].

All these measures are referred to as *single-scale* entropy measures. By contrast, *multiple-scale* entropy measures are derived from the above and consist in analyzing a time series from different time scales. The concept of multiple-scale entropy was initially introduced by Costa et al. [11]. A modified entropy definition, named *Multiscale Entropy* (MSEn), was proposed to estimate entropy over a range of scales enabled by a coarse-graining procedure [12]. Since then, various definitions of scale-extraction mechanisms were proposed leading to an increasing number of *multiple-scale* entropy measures, in which *single-scale* entropy measures provide the basis of entropy estimation under the multiple-scale framework [13], [14].

One advantage of entropy measures is that they do not rely on linear assumptions, and are suitable for distinguishing regular, chaotic and random behaviors. Complex systems with nonlinear dynamics present larger response diversity and uncertainty; thus it is sometimes easier to characterize underlying patterns in terms of dynamic changes, than to analyze the little knowledge base data available. Entropy measures can directly detect dynamic changes and quantify the degree of complexity of a system, which would be challenging to assess by traditional statistical indicators [15], [16]. Since the performance degradation of a machine will present more non-linear characteristics, the analysis of the *complexity* of the measurements has revealed that the change in the *complexity* value is related to the deterioration of the machine component [17]. With the significant advancements in sensor networks and computing systems, data-driven fault diagnosis has become increasingly attractive. Continuing advances in signal analysis and Artificial Intelligence (AI) techniques have led to a growing number of data-driven fault diagnostic systems. Such systems are based on large amounts of sensor data and knowledge mining techniques [18].

For data-driven machine fault diagnosis, extracting useful underlying knowledge – related to fault patterns – is fundamental. The underlying knowledge represents fault features capable of distinguishing between system states. Extracted features usually represent the nature of the signal and the evolution of the state of the system. Fault diagnosis can be carried out by checking the consistency between feature representations extracted from sensor readings and the values predicted from a model - constructed upon historical signal features. The most common statistical features can be either time domain (e.g., mean, standard deviation, kurtosis and skewness) or frequency domain (e.g., power spectrum).

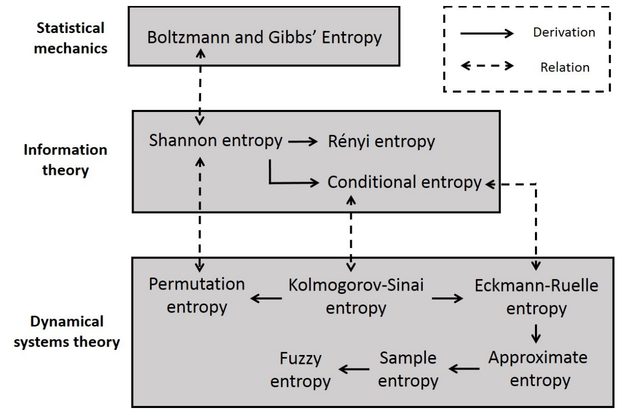


Fig. 1: Relations between the various entropy definitions found within the contexts of statistical mechanics, information theory, and dynamical systems (solid line arrows indicate direct mathematical derivations, while dashed arrows show conceptual association).

One of the difficulties with these traditional methods is that they rely on linearity and signal stationarity assumptions – which may not appropriately extract signal symptoms, especially under complex environments with interacting components (systems of systems) and strong background noise. In an actual example, the level of kurtosis was reduced as damage in the machine bearings increased; when the vibration pattern became more complex – due to the bearing damage, the kurtosis matched that of undamaged bearings [19]. In complicated industrial systems, the machine may exhibit non-linear behavior due to instantaneous variations in friction, damping, or load and speed conditions; thus quantifying dynamic changes of system responses is significant to early fault detection [20], [21].

The extensive flexibility of entropy analysis methods is advertised by their all-encompassing applicability to the analysis of complex systems, be it natural or man-made; besides the subject of monitoring industrial machines, entropy analysis has been extensively applied for studying the complexity of dynamical systems in multiple fields. Such areas of research may be far more complex than mechanical systems, including language [22], biological [23], financial [24] and other complex systems [25]–[28]. However, much less work has been reported on the comparative domain between different definitions of entropy measures and their modular usages for machine fault diagnosis.

Thus, this paper aims to arrive at an understanding of some of the most significant principles of entropy measures and to clarify their relations. Their applicability to rotary machine fault diagnosis is considered as the main illustrative example.

II. Theoretical Background of Entropy Measures

This section introduces several of the most widely used entropy measures for time series complexity analysis. The essential properties of these entropy measures are also discussed, along with their mutual associations from an information-theoretic perspective. Fig. 1 outlines the mathematical and conceptual interrelationships between different entropy definitions provided, while Table I comparatively summarizes their characteristics.

A. Shannon Entropy (ShanEn) and Related Concepts

Occupying the center stage in information theory, ShanEn measures the missing information about a random variable X. When the random variable is understood as the outcome (observation) of a system, ShanEn can be interpreted as the rate of generation of new information processed by the system [4]. According to Shannon, information and uncertainty are two sides of the same coin: the reception of a certain amount of information is equivalent to a reduction in uncertainty. Thus, the larger the entropy about a system, the more uncertainty about its response, and the more information can be gained by observing the outcomes of the corresponding random variable.

For a discrete random variable X – with $n \in \mathbb{N}$ possible outcomes – ShanEn is defined as

$$H(X) = -\sum_{i=1}^{n} p(x_i) \log_2 p(x_i), \tag{1}$$

where $p(X) = \{p(x_1), p(x_2), \cdots, p(x_n)\}$ are the assigned probabilities to the outcomes of X.

In (1), $H(X)$ coincides with the average number of bits per outcome yielded by X. $H(X)$ ranges from 0 to $\log_2 n$. When all the outcomes are equally probable – i.e., $p(X)$ is uniformly distributed – $H(X)$ reaches its maximum value. In contrast, when the outcome is certain, $H(X)$ is zero, and there is no information from the outcome.

Based on ShanEn, other related formulations were put forth in information theory. Preeminent examples are conditional entropy, mutual information, and cross-entropy. Conditional entropy can be expressed as $H(X|Y) = H(X,Y) - H(X)$, where $H(X,Y)$ is the entropy of the joint probability distribution $P(X,Y)$. It measures the missing information and uncertainty about X upon observing another measurement of Y. Mutual information is defined as $I(X,Y) = H(X) - H(X|Y)$, and it captures the amount of information that two variables X and Y share [39]. Moreover, cross-entropy is expressed as $H(p,q) = -\sum_x p(x) \log_2 q(x)$ where $p(x)$ and $q(x)$ are typically the ground-truth and estimated probability distributions, respectively. Cross-entropy minimization has been popularly used in optimization algorithms, such as model optimization in neural networks. Also, it has been proved that ShanEn is no larger than cross-entropy[1].

[1] $H(p) = -\sum_x p(x) \log_2 p(x) \leq -\sum_x p(x) \log_2 q(x) = H(p,q)$.

In addition to information theory, entropy is also a crucial notion in complexity and chaos theory. Entropy is often linked to the degree of chaos in an observed dynamical system because uncertainty can be explained as unpredictability or irregularity in a system. In dynamical systems theory, KS entropy is an interesting concept, which is a generalization of ShanEn employed in the study of seemingly random but deterministic dynamical systems (i.e., *deterministic chaotic systems*) [40]. KS entropy analyzes how the uncertainty about a system evolves from its dynamical equations. That is, it yields the rate of generation of new information by the examined system. From an information-theoretic standpoint, chaotic behaviors are described by KS entropy through a partition of the state space [41]; thus, it is equally suitable for discrete and for continuous dynamical systems. Positive values of KS entropy are interpreted as an increase in uncertainty with respect to the system's responses [41]. Hence, systems with positive KS entropy can be regarded as chaotic systems – displaying sensitive dependence on the initial conditions [42].

In the study of non-linear dynamical systems, the Lyapunov exponents are relevant indicators, suitable to quantify the topological characteristics of the dynamics and system stability. Pesin's theorem establishes a relationship between the KS entropy and Lyapunov exponents [43]. Nevertheless, when performing numerical analysis by way of experimental data, it is usually very hard to calculate Lyapunov exponents and KS entropy directly. Added difficulty results from the fact that KS entropy relies on arbitrarily fine partitions of the state space, and from its lack of robustness to noisy measurements. Typically, KS requires a large amount of measured data to achieve convergence [7].

For this reasons, various studies on entropy have led to alternative entropy formulations, which attempt to estimate time-varying dynamic changes within a system (e.g., the methods by Grassberger and Procaccia [44] and Eckmann and Ruelle [6]). Thus many entropy analysis methods populate literature, which is described in this section.

B. Rényi entropy

The discussion proceeds with a generalization of ShanEn: Rényi entropy, defined as

$$H_\alpha(p) = \frac{1}{1-\alpha} \log_2 \left(\sum_{i=1}^{n} p_i^\alpha \right), \tag{2}$$

where $\alpha \in [0, \infty)$ and $\alpha \neq 1$. Eq. 2 becomes ShanEn when $\alpha \to 1$.

Rényi entropy is characterized as a continuous family of entropy measures (H_α) by way of a *bias* parameter α [45]; α controls the degree of sensitivity of $H_\alpha(p)$ towards particular probability distribution functions [45] and makes $H_\alpha(p)$ non-negative for all α. Other special cases of Rényi entropy include collision entropy ($\alpha = 2$) and min-entropy ($\alpha \to \infty$). Collision entropy is the negative logarithm of the probability that two independent and identically distributed random variables present the same outcome (or *collide*). As more likely events are more probable to collide, these are more conspicuous under the collision entropy measure than with ShanEn.

TABLE I: List of advantages/limitations of various entropy measures for complexity analysis in dynamical systems.

Year	Entropy measures	Advantages	Limitations	Algorithmic Complexity[1]
1948	ShanEn [4]	• foundational measure to estimate the amount of information content of messages from probability viewpoint [5]	• dependence on the probabilistic model of uncertainty as present in a probabilistic event space [29] • neglect of temporal relationship between values [30]	$O(n)$
1991	ApEn [7]	• applicable to measuring the complexity change of deterministic and chaotic dynamical systems • suitable to medium-sized data [21]	• generation of more similarity than is present • lack of consistency relative to SampEn • heavily dependent on data length [8]	$O(n^{\frac{3}{2}})$
2000	SampEn [8]	• better consistency relative to ApEn • robustness to small noisy data [31]	• discontinuity and mutation at the boundary [32] • sensitive to parameter selection and data length	$O(n^{\frac{3}{2}})$
2007	FuzzyEn [9]	• better consistency relative to ApEn and SampEn • continuity at the boundary [33]	• sensitive to parameter selection • membership function needs more physical meaning	$O(n^{\frac{3}{2}})$
2002	PerEn [10]	• partition naturally derived from ordinal patterns • invariance with respect to non-linear monotonous transformations [34], [35]	• dependence on parameter selection • amplitude difference in values is neglected • cases with many equal values are not considered [36]	$O(n \log_2 n)$
2002	Multiple-scale entropy [11]	• better classification accuracy relative to single-scale entropy measure • more robust to small degree of noise • more information related to frequency characteristics	• efficiency differs depending on applied scale-extraction mechanism and selected single-scale entropy • more time consumption because of computation of entropy measures via a range of scales [13]	$O(mn)$ $\vdots$ $O(mn^{\frac{3}{2}})$

[1] The algorithmic complexity of ApEn, SampEn, and FuzzyEn refers to optimized calculation algorithms in [37], [38]. For multiple-scale entropy measures, their computational efficiency depends on mainly selected scale-extraction mechanism and single-scale entropy method for entropy estimation. Herein, n denotes the input size in units of bits needed to represent the input, and m is the number of scales in multiple-scale entropy methods.

As $\alpha \to \infty$, Rényi entropy is increasingly determined by the events of highest probability; thus, min-entropy is the negative logarithm of the probability of the most likely outcome only.

C. Approximate Entropy (ApEn) and its Variants

1) ApEn: Another complexity indicator was introduced by Pincus and known as ApEn [46]. ApEn was constructed thematically similar to KS entropy.

During the 1980s, several studies attempted to directly compute KS entropy, among which Eckmann-Ruelle entropy[2] displays the greatest potential for practical implementation [6]. Later, Pincus modified Eckmann-Ruelle entropy for the analysis of finite and noisy time series derived from experiments.

ApEn assumes that, for fixed m, if two measures describing two different systems – that have distinct marginal probabilities – it is sufficient to discriminate between these two measures to classify the different underlying processes in the two systems. In contrast to Eckmann-Ruelle entropy, ApEn requires fewer points to estimate marginal probabilities, usually allowing practical discrimination. [46]

This measure was employed initially for the study of deterministic complex dynamical models (e.g., Rossler model, logistic map, and Henon map) and later applied to the analysis of biological signals – such as heart rate recordings – incorporating both stochastic and deterministic components [48].

ApEn is computed as follows. For a time series $x_1 \dots, x_N$ and a value $m \in \mathbb{N} < N$ the vectors

$$\boldsymbol{x}^m(i) = (x_i, x_{i+1}, \dots, x_{i+m-1}) \in \mathbb{R}^m \tag{3}$$

are considered ($\boldsymbol{x}^m_k(i) = x_{i+k}$ with $k \in \{0, \dots, m-1\}$). These are referred to as *templates*. Then, given a template $\boldsymbol{x}^m(i)$, the quantity $C^m_{ij}(r)$ is defined as $1/(N-m+1)$ times the number of templates such that

$$\max_k |\boldsymbol{x}^m_k(i) - \boldsymbol{x}^m_k(j)| < r \tag{4}$$

for a given $r \in \mathbb{R}^+$, where i and j range from 1 to $N-m+1$.

Considering $\Phi^m(r)$ as the estimated probabilities of the natural logarithm of $C^m_{ij}(r)$

$$\Phi^m(r) = (N-m+1)^{-1} \sum_{i=1}^{N-m+1} \log_2 C^m_{ij}(r), \tag{5}$$

ApEn is defined:

$$\text{ApEn}(m, r, N) = \Phi^m(r) - \Phi^{m+1}(r). \tag{6}$$

ApEn can be shown to be closely related to the notion of conditional entropy. Thus, it estimates the uncertainty with respect to future observations of a time series, given the knowledge of the past observations. It is proposed that, when the behavior of the process generating the time series becomes irregular or chaotic, ApEn increases – although a nonzero ApEn value does not certify that the dynamics are chaotic [48].

Several hyperparameters must be fine-tuned for optimal performance (such as the *embedding dimension* m and the tolerance r) – although empirical values are offered in the literature; when $m = 2$, values of r ranging between 0.1 to 0.25 times the standard deviation (σ) of time series can produce reasonable results [7]. For the analysis of rotating machinery, the values $m = 2$ and $r = 0.4\sigma$ have been suggested [21]. In the same publication, it is claimed that $N = 750 - 5000$ is sufficient for achieving consistent results. Lu et al. [33] have developed an automatic r selection

[2] Eckmann-Ruelle entropy approximates the KS entropy as $\lim_{r\to 0} \lim_{m\to\infty} \lim_{N\to\infty} [\Phi^m(r) - \Phi^{m+1}(r)]$, and it is based on the work by Grassberger and Procaccia [44] and Takens [47]. A nonzero Eckmann-Ruelle entropy value assures the deterministic system is chaotic.

approach that can reduce the computational cost while fitting the hyperparameter r. Kaffashi et al. [49] have investigated the influence of hyperparameter selection on analyzing real-time series with ApEn.

Several modified ApEn algorithms have been proposed with alleged improved performance. One example is Cross-ApEn, also developed by Pincus, that measures the statistical independence of two concurrent time series, by capturing both spatial and temporal irregularity [50]. Another example is SampEn, which is discussed next.

2) Sample Entropy (SampEn): SampEn refines the ApEn algorithm through two differing aspects [51]: *i)* SampEn excludes self-matches while counting template matches (Eq. 4); *ii)* in SampEn only the first $N-m$ vectors are considered (Eq. 5) – this ensures that for $1 \leq i \leq N-m$ both $\boldsymbol{x}_k^m(i)$ and $\boldsymbol{x}_k^m(j)$ are defined [8].

To compute SampEn, first count the number of template matches obtained from Eq. 4, where i ranges from 1 to $N-m$ and $j \neq i$. Then, define $C_{ij}^m(r)$ as $1/(N-m-1)$ times the number of template matches.

Define $\Phi^m(r)$ as

$$\Phi^m(r) = (N-m)^{-1} \sum_{i=1}^{N-m} C_{ij}^m(r), \tag{7}$$

And from (7) SampEn is defined:

$$\text{SampEn}(m, r, N) = \log_2 \Big[\frac{\Phi^m(r)}{\Phi^{m+1}(r)} \Big]. \tag{8}$$

Hence in the computation of SampEn, unlike that of ApEn, the logarithm is applied after Φ^m is obtained. Because the quantities $C_{ij}^m(r)$ act as surrogates of the probabilities $p(x_i)$ in (1), ApEn is closer to the mathematical formulation of the original entropy. Nonetheless, it has been verified that SampEn reduces bias and maintains relative consistency as compared to ApEn [31]. That is, if a time series A arising from a more ordered system than time series B, then ApEn of A has been shown to be smaller than ApEn of B for all conditions tested [48]. As an example, Yentes et al. comparatively investigated the performance of ApEn and SampEn in time series analysis. They found that SampEn is less sensitive to the change of data length and shows better performance compared to ApEn when analyzing clinical data sets in pathological populations [52], [53].

There exist enhanced formulations of SampEn algorithm, reducing its algorithmic complexity. For instance, Lu et al. [54] presented a method to accelerate the computation of ApEn and SampEn by exploiting vector dissimilarity. This method omits the computation of distances between the most dissimilar vectors, which further reduces the time complexity. Additionally, Manis et al. [55] proposed three SampEn algorithms that yield identical values but are less expensive computationally (by avoiding the similarity check between points in m dimensional phase space). Moreover, Silva et al. [56] extended SampEn to two-dimensional time series analysis. This method was applied to the analysis of image data.

A potential limitation of ApEn and SampEn resides in Eq. 4: the method to select template matches consists in establishing a crisp boundary. This may generate discontinuities and implies a strong dependence on the parameter r. To address this shortcoming, new methods have been proposed that introduce the concept of *fuzziness*. These are discussed next.

3) Fuzzy Entropy (FuzzyEn): Chen et al. introduced the notion of FuzzyEn to measure time series irregularity based on SampEn [9]. In FuzzyEn, the concept of degree of 'fuzzy membership', inherited from the framework of fuzzy logic, was introduced. Fuzzy sets are characterized by 'vague boundaries', enabling continuous membership assignments through a fuzzy membership function. A Fuzzy membership function can be employed to quantify the degree of similarity between two vectors – for example, by mapping two vectors to an scalar in $[0, 1]$. Thus, these scalars can also be understood as probabilities, and are subject to entropy analysis.

FuzzyEn uses the membership function $\exp(-d^n/r)$, where r and n control the width and gradient of the boundary respectively, and d is the maximum absolute difference of the corresponding scalar components according to Eq. 4. Other membership functions have been considered in the literature, such as $\exp(-d^{\ln(\ln 2^c)/lnr}/c)$ in [57], $\exp(-\ln 2(d/r)^n)$ in [58], and $\exp(-(d/r)^p)$ in [59]. Moreover, other modified FuzzyEn approaches have been developed for improved performance: a piecewise fuzzy membership function proposed in [60] and a modified Fuzzy Entropy, which operates by increasing the number of samples during the computation of the entropy [61].

Some comparative studies have investigated the performance of FuzzyEn [62], [63], in terms of its relative consistency, dependency on parameter choice, and robustness to noise; FuzzyEn offers better consistency and is less dependent on the size of the data set compared to SampEn [63].

D. Permutation Entropy (PerEn)

PerEn, proposed by Bandit and Pompe [10], measures the underlying dynamic changes encoded in the ordinal patterns of a time series. PerEn is essentially ShanEn over the empirical probability distribution of the ordinal patterns naturally originated from the time series data:

$$\text{PerEn}(m, \lambda, N) = -\sum_{j=1}^{m!} p(\pi_j) \log_2 p(\pi_j) \tag{9}$$

where λ is time delay and $\pi_j = (j_1, j_2, \cdots, j_m)$ is one of $m!$ possible permutation patterns; the ordinal pattern π_j is obtained from m ranked data points in ascending order. $p(\pi_j)$ is the probability of a given ordinal pattern, and defined as

$$p(\pi_j) = \frac{\#(\boldsymbol{x}^m(i) \text{ has type } \pi_j)}{N-(m-1)\lambda} \tag{10}$$

where $\#$ denotes the cardinality of each permutation pattern. Accordingly, PerEn can be interpreted as a measure of the rate at which new permutation patterns are produced in the process of a system. In contrast to ShanEn, PerEn results from the symbolic dynamics of the studied system [30] thus is less likely to be affected by transients in the data.

In the analysis of dynamical systems, PerEn is related to KS entropy, when the partition is defined based on the order of a time series. More specifically, in PerEn, permutation patterns (i.e., the partitions) result from a map, by translating into a sequence of symbols. In addition, PerEn provides an upper bound for KS entropy when $m \rightarrow \infty$ [64] and is also related to the Lyapunov exponents of a dynamical system [10].

PerEn presents a few limitations, which are caused by only considering the order, and not the amplitudes in neighboring elements [30], [34]. Thus, different time series may have the same PerEn value, lowering its discriminating capacity. Also, when repeated values emerge in the sensor data, PerEn assigns their sequential order according to emergence order. This results in ambiguity in the mapping from sensor data to permutations, and may introduce bias in the empirical distribution estimates. Typically repeated values are rare, but this is not the case in quasi-stationary systems or systems in an stationary operational regime [35].

To overcome these limitations, several variants of PerEn have been proposed. Some of them take into account the amplitude difference – by using weighting coefficients such that the magnitudes of neighboring elements have different contribution to the relative frequencies of the permutation types [65]–[68]. Further, in order to tackle the problem of repeated measurements mentioned above, Bian et al. [69] presented a solution by mapping the repeated values onto the same symbol.

E. Multiple-scale Entropy Measures

Multiple-scale entropy measures are generalized entropy methods based on scale-extraction mechanisms, and the already defined single-scale entropy methods. In general, multiple-scale entropy algorithms consist of two steps: *i)* the extraction of multiple time series of different scales from the original data through a scale-extraction mechanism; *ii)* the calculation of the entropy for each extracted time series via a single-scale entropy method. Thus the performance of a multiple-scale entropy method greatly depends on that of its associated single-scale entropy measure.

One most common approach is MSEn, which consists in measuring SampEn through a coarse-graining – or *averaging* – procedure [11]. The algorithm to compute MSEn is described in the following. Given a time series $x_1 \ldots, x_N$ of length N and a scaling factor τ, the coarse-grained time series, $\mathbf{y}^{(\tau)}$, is obtained by the relation

$$y_j^{(\tau)} = \frac{1}{\tau} \sum_{i=(j-1)\tau+1}^{j\tau} x_i, \quad \text{for } 1 \leq j \leq \frac{N}{\tau}, \quad N > \tau. \quad (11)$$

When $\tau = 1$, $\mathbf{y}^{(\tau)}$ coincides with the original time series $\mathbf{x}$. From $\mathbf{y}^{(\tau)}$, MSEn is defined as:

$$\text{MSEn}(x, \tau, m, r) = \text{SampEn}(\mathbf{y}^{(\tau)}, m, r). \quad (12)$$

Compared to its single-scale counterpart SampEn, MSEn can extract coarse-grained time series representing the system dynamics over a range of multiple temporal scales. More information associated with the complexity change underlying in measurements is then characterized from the coarse-grained time series [12].

Beyond its advantages, the coarse-graining procedure has several limitations. First, from a signal processing standpoint, the coarse-graining procedure in (11) is a linear smoothing operation, which abandons high-frequency information. Secondly, the down-sampling effect will result in increasingly shorter time series, which may introduce bias when estimating the entropy through SampEn [70]–[72].

Motivated by MSEn, several modified multiple-scale entropy measures have been developed through various scale-extraction frameworks and notions of single-scale entropy algorithms [13], [14]. Multiple-scale entropy measures can be classified into three main groups according to their operational principles:

- *Improved coarse-graining procedure based entropy approaches:* modified coarse-graining procedures and variants of single-scale entropy algorithms are applied for entropy analysis. Modified coarse-graining procedures mainly focus on improving the efficacy of extracting multiple-scale coarse-grained time series. The shortcoming of generating time series with greatly decreasing data length is alleviated using improved coarse-grained techniques. Examples include composite MSEn [72], generalized MSEn [73], and refined composite MSEn [74]. Moreover, the use of different single-scale entropy algorithms can improve the performance in analyzing coarse-grained time series further, such as Multiscale Permutation Entropy (MPEn) [36], refined composite MPEn [75], Multiscale Fuzzy Entropy (MFEn) [76], and modified multiscale symbolic dynamic entropy [77].
- *Filter-inspired scale-extraction based entropy approaches:* improved scale-extraction procedures are applied for entropy analysis where both low- and high-frequency information is refined and maintained in extracted multiple-scale time series via filter-inspired operations. For instance, a hierarchical decomposition is used in [78], preserving the strength of the multiscale decomposition with additional components of higher frequency in different scales. A fine-to-coarse procedure is developed in [79], aiming to generate multiple-scale components with fine-grained low- and high-frequency information, and to yield better consistent entropy values even with high scales and strong noise.
- *Multivariate analysis based entropy approaches:* the complexity of multichannel data is assessed with multivariate extensions of MSEn where multichannel data is analyzed with a definition of multivariate single-scale entropy algorithm. Examples include multivariate MSEn [80], refined composite multivariate MFEn [81], and refined composite multivariate generalized MFEn [82].

Continuing research in entropy measures (i.e., single- and multiple-scale entropy approaches) has driven the emergence of more useful non-linear time series analysis, which can effectively distinguish the different operational regimes of the system. There exist many improved notions of single-scale entropy approaches (e.g., increment entropy [83], joint distribution entropy [84], and dispersion entropy [85]) and multiple-scale entropy approaches (e.g., composite interpolation-based MFE [86] and multiscale fluctuation-based dispersion entropy [87]) for time series complexity analysis. In the next section, entropy-based applications are surveyed and summarized for machine fault diagnosis.

III. Entropy-based Applications for data-driven Machine Fault Diagnosis

Data-driven analysis of system performance has shown that, changes in complexity are often linked to machine degradation and failure emergence. Entropy measures are suitable to detect and quantify underlying dynamic changes in system response. These changes in complexity allow for machine condition monitoring, and for distinguishing among various operational regimes. The entropy measures, discussed in Sec. II, facilitate the usage for machine health condition monitoring in industrial applications. With advanced signal analysis and AI techniques, entropy measures have assisted in enhancing maintenance strategies and increasing the reliability of machine fault diagnostic systems. With these applications in mind, entropy analysis can be further classified in three categories: entropy measure as a feature indicator, entropy criterion for parameter selection, and entropy usage in pattern recognition.

A. Entropy Measure as a Feature Indicator

In data-driven fault diagnosis, entropy measures are mostly employed as complexity indicators. Since existing faults often introduce non-linear characteristics in the measurements, changes in complexity of a system are correlated with its failure rate. Thus, entropy measures facilitate machine condition monitoring and can detect performance degradation in the machine. A schematic of the entropy-based feature extraction – towards machine fault diagnosis – is presented in Fig. 2.

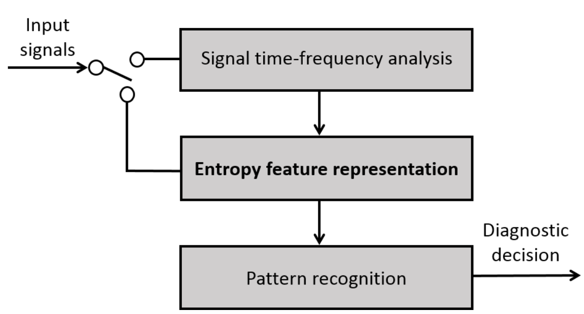


Fig. 2: Schematic of the usage of entropy measure as a feature indicator towards machine fault diagnosis.

With respect to feature extraction, ShanEn is typically used to estimate the complexity and uncertainty in the measurements, such as the analysis of vibration data [88]. The occurrence of incipient failures in the machinery will introduce coupling frequencies, resulting in changes in both, the energy and the spectrum in the measurements. Thus, there exist some entropy notions inspired by ShanEn, which enable entropy feature characterization from time- or frequency-domain sensor data.

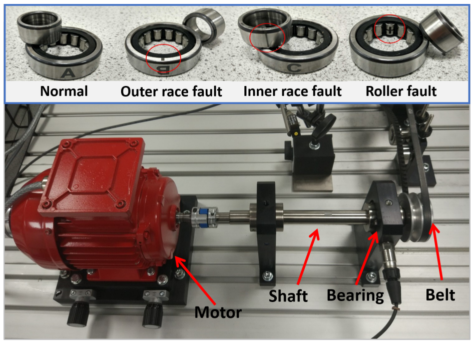
Specifically, energy entropy and spectral entropy are two useful health monitoring indicators, which characterize complexity changes from the time-domain and the frequency-domain, respectively. Both are defined as ShanEn of a given probability distribution. The probability distribution in energy entropy is associated with power energy distribution of the transformed (decomposed) components in the time domain [89]. By contrast, the probability distribution in spectral entropy is related to the power spectrum distribution of the transformed components in the frequency domain [90]. These two entropy measures are practical for distinguishing machine health conditions, usually in combination with signal time-frequency analysis techniques, such as wavelet analysis [89], Fourier analysis [91], and Hilbert transform [92].

As non-linear complexity indicators, ApEn, SampEn, and FuzzyEn have been examined in detecting structural defects in mechanical systems [93]–[95]. Some studies have analyzed their performance in entropy analysis using various parameters. For instance, Yan [21] studied the effects of data length, embedding dimension, and tolerance in the calculation of ApEn for the analysis of bearing vibration signals. Moreover, Sampaio et al. [96] studied the effectiveness of ApEn for detecting rotating shaft deterioration; it was reported that ApEn is applicable for detecting crack defects in rotating shafts – when the crack depth is larger than 5% of the shaft diameter. Further, Kedadouche [32] verified that ApEn and SampEn enable detecting structural damage in gearboxes, suggesting that $m = 2$ and $r = 0.5\sigma$ are suitable for the calculation of ApEn and SampEn values.

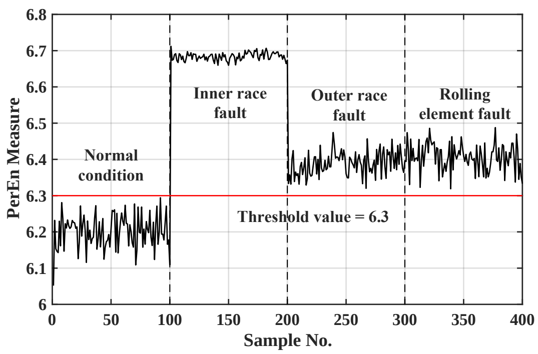
With respect to contrasting performance, FuzzyEn was compared with ApEn and SampEn in [57], for the particular case of rolling bearing fault diagnosis. The discriminatory capability of these three methods was evaluated and their multiple-scale entropy methods based on the coarse-graining procedure were also studied. The results indicated that FuzzyEn – and its multiple-scale counterpart – outperform ApEn and SampEn in improved classification accuracy, and can yield smoother entropy estimations [57].

PerEn applies to distinguish machine health conditions. One example is the study in Yan [15], where a comparative study was performed on the usage of PerEn in bearing diagnosis. In their study, the efficiency of PerEn with different parameters was investigated, such as data length, embedding dimension, time delay, and computational efficiency. The authors concluded that $m = 6$ and time delay $\lambda = 3$ could give reasonable PerEn values for practical bearing diagnosis. Moreover, PerEn values extracted from a healthy machine can be employed as threshold indicators for anomaly detection in the operation of machine [97].

We now provide an example of the usage of PerEn for detecting early faults in roller bearings in a test rig. This example is typical as the most common machine failures are linked to structural damage, such as wear-out and corrosion. The overall reliability of the machinery is highly dependent upon the health state of the bearing, which accounts for approximate 45% to 55% of the total number of failures [98]. Fig. 3 (a) shows a PT 500 series bearing test rig benchmark, composed of a motor, a shaft, bearing, and belt drive [99]. Four



(a) PT 500 test rig



(b) PerEn measurement under different conditions.

Fig. 3: Comparison of PerEn values between from four types of bearing vibration data (with embedding dimension $m = 6$ and time delay $t = 1$). The red line shows a threshold empirically tuned at PerEn $= 6.3$.

bearing states are considered, including normal bearing and faulty bearings with damage in the inner race, outer race, and roller element. Vibration data was collected with an operation at speed 2000 r.p.m and with sampling frequency of 8 kHz. PerEn values are calculated from vibration signals with data length of 1024. PerEn results are presented in Fig. 3 (b); it demonstrates that machine faults can lead to higher complexity within the system. Also, entropy indicators apply to performance degradation detection and anomaly detection.

With respect to multiple-scale entropy measures, many studies have explored their capability as fault indicators. The majority of these studies has mainly focused on improved single-scale entropy approaches and on enhanced scale-extraction procedures. For instance, some works investigate the performance of entropy measures for detecting failures in the machinery, where different single-scale entropy values are calculated under a multiple-scale framework [100]–[102]. Further, some modified entropy measures have been proposed based on enhanced scale-extraction mechanisms. Related studies include generalized composite MPE [103], hierarchical entropy [104], modified hierarchical PerEn [105], and fine-to-coarse MPE [79]. In general, these methods earn higher consistency and reduced bias in time series complexity analysis, as compared with single-scale methods. As a result, multiple-scale entropy measures with improved scale-extraction framework usually present higher fault classification accuracy, specially when more machine conditions and variable working condition are considered (e.g., various rotating speeds, signal-to-noise ratios, and loads).

In summary, entropy measures display an extensive application prospect in monitoring machine health states. Through system complexity analysis, it is possible to distinguish different underlying processes in the system, therefore detecting potential failures in the machine.

B. Entropy Criterion for Parameter Selection

Entropy measures bring up the possibility of specifying desired parameters to characterize time-frequency representations in signal processing techniques. In the machinery, the occurrence of defects in rotating components will excite characteristic amplitudes and frequencies in both time- and the frequency-domain. Usually, signal time-frequency analysis methods are used to transform raw signals into time-frequency representations, and then crucial fault symptoms of interest are characterized with statistical indicators from the obtained components. Nonetheless, not all components are directly associated with fault symptoms, and some components contain redundant information. Thus, the selection of prominent time-frequency components is necessary. As larger entropy values usually indicate more irregularity, entropy measures can help to select salient components whose complexity degree may increase – due to the existence of defects. Moreover, instead of specifying parameters according to prior knowledge, entropy measures facilitate the choice of the optimal parameters. Fig. 4 shows a schematic of entropy-criterion for parameter selection in fault diagnostic systems.

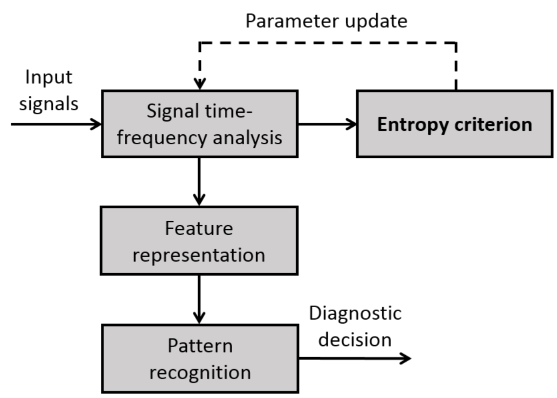
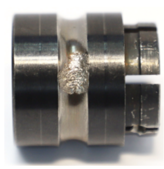


Fig. 4: Schematic of the entropy-based criterion for selecting parameters in signal time-frequency analysis towards machine fault diagnosis.

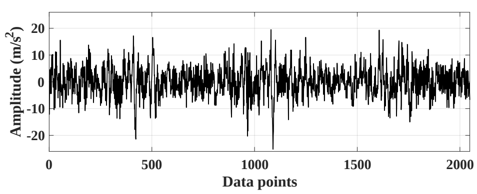
One of such entropy methods for parameter selection is *wavelet analysis*, that has been extensively applied for fault diagnosis – by transforming signals into wavelet coefficients in the time-scale domain. Examples of the studied wavelet analysis methods are: Continuous Wavelet Transform (CWT), Discrete Wavelet Transform (DWT), and Wavelet Packet Transform (WPT) [106], [107]. In wavelet analysis, the selection of appropriate mother wavelet and decomposition scale is the key to capture crucial features from signals; however, it usually requires prior knowledge to fine-tune these wavelet parameters

TABLE II: Entropy-based criteria for optimal parameter selection in wavelet analysis.

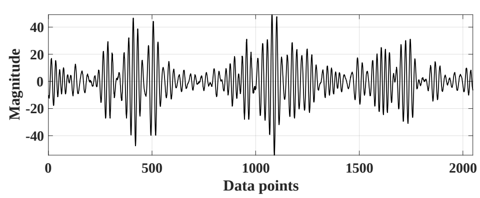
Criterion	Description	Application
Minimum ShanEn	Energy content of a few wavelet coefficients is high with the occurrence of characteristic frequency components, resulting in decreased entropy values.	Optimal coefficient selection, suitable for CWT [108]
Minimum-entropy	A node is decomposed if and only if entropy of its two child nodes is no larger than that of their father node.	Optimal tree selection, suitable for DWT and WPT [109]
Maximum energy to ShanEn ratio	Desired wavelet usually extracts maximum amount of energy while minimizing the ShanEn of corresponding wavelet coefficients.	Optimal coefficient selection, suitable for CWT [110]



(a) Bearing with wear damage on inner race



(b) Raw vibration signal



(c) Extracted wavelet coefficients

Fig. 5: (a) Faulty bearing vibration presenting inner race wear [116], (b) resulting raw vibration signal, and (c) obtained wavelet coefficients through CWT analysis using maximum energy to ShanEn ratio criterion.

for any signal. The most common criteria include minimum ShanEn criterion [108], minimum-entropy criterion [109], and maximum energy to ShanEn ratio criterion [110]. In other related works [111]–[115], wavelet analysis methods – with entropy-based parameters selection – are investigated. Table II summarizes the description and applicability of three typical ShanEn-based criteria for wavelet analysis.

An example of CWT analysis for extracting fault features from the fault-deduced transient vibration signals is now presented (Fig. 5). In this case, an appropriate mother wavelet was selected using the maximum energy to ShanEn ratio criterion. A bearing with wear damage on the inner race was studied with vibration data contributed by the Xi'an Jiaotong University [116]. Fig. 5 (b) shows the vibration signal of the bearing; it is apparent that fault features of the raw signal are difficult to be identified, due to instantaneous variations and background noise. For this purpose, CWT is then applied for the identification of underlying fault symptoms in the bearing signal. The maximum energy to ShanEn ratio values are calculated based on the wavelet coefficients to select an appropriate mother wavelet that can best match the shape of the bearing signal. For this purpose, the vibration sensor data is decomposed into 64 sub-signals using five different mother wavelets: Meyer, Morlet, Mexican, Daubechies 4, and Haar. It is observed that the coefficients using Morlet wavelet at scale 18 achieve the highest maximum energy to ShanEn ratio value – their waveform is shown in Fig. 5 (c). The figure suggests characteristic fault symptoms that are related to the successive periodic pulses, caused by fundamental frequency in the bearing with inner race fault.

In summary, several entropy-based criteria are available for specifying appropriate parameters in multi-resolution signal analysis. Through maximizing the total amount of extracted information, fault detection is enhanced via optimal transformation of raw signals – and the extraction of characteristic fault features.

C. Entropy Usage in Pattern Recognition

Various entropy based methods can be employed for pattern classification and model optimization. In pattern recognition, designing reliable and optimized data-driven models [117] is the key to guarantee accurate diagnostic decision-making. As ShanEn evaluates the uncertainty in the variables of a system, based on an empirical probability distribution, it can be used to describe the closeness of two probability distributions - the ground-truth and prediction probability distribution. This is done via a generalization of ShanEn known as cross-entropy (Sec. II-A). Smaller cross-entropy values indicate that the probability distribution of a model is closer to the empirical distribution in the data. Fig. 6 shows a schematic of entropy-based pattern recognition techniques towards machine fault diagnosis.

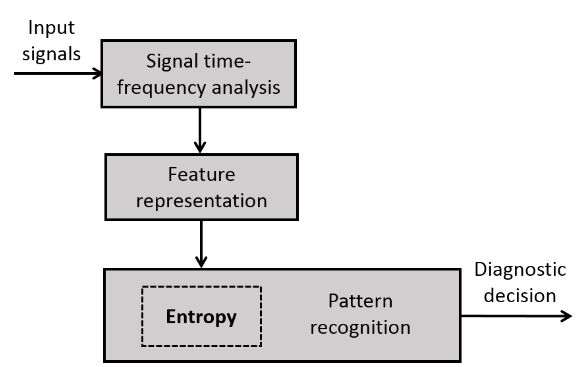
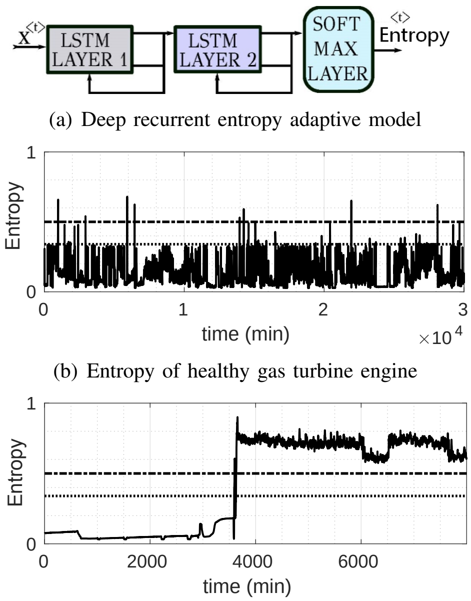


Fig. 6: Schematic of entropy-based model optimization in pattern recognition towards machine fault diagnosis.

For these reasons, cross-entropy is among the most commonly used loss functions for training and evaluating the performance of artificial neural network classifiers [118].

Regarding probabilistic classification, the estimation of the effectiveness of the acquired models is usually required [119], by which hyper-parameters are fine-tuned through minimizing the cross-entropy over a development and a test set – not used during the training phase. The cross-entropy function [120] is expected to perform better at improving the efficacy of training models, compared with traditional square error objective functions. Related works where cross-entropy is used for the construction of *deep learning models* refer to [121]–[125].



(a) Deep recurrent entropy adaptive model

(b) Entropy of healthy gas turbine engine

(c) Entropy of gas turbine engine with compressor fault

Fig. 7: Illustration of ShanEn in neural networks for industrial gas turbine fault diagnosis [126]. The dashed lines indicate entropy-based thresholds for warning and a faulty system. LSTM: Long Short-Term Memory.

Because ShanEn is always smaller or equal than cross-entropy (Sec. II-A), minimizing cross-entropy can be understood as estimating ShanEn. The usage of cross-entropy for industrial gas turbine compressor fault diagnosis is explored in [126]. A regressor *recurrent neural network* model was converted into a classifier by bucketing the outputs. An example of this model is presented in Fig. 7. The model consists of two *long short-term memory* layers, incorporating a gating mechanism to control the memory retention operation. The classifier – once trained through a cross-entropy approach – yields ShanEn estimates, indicating the degree of uncertainty in the system. After that, the entropy adaptive model is capable of distinguishing between typical dynamics, corresponding to healthy engines, and anomalous behaviour from faulty engines. Also, it was shown that different changes of the uncertainty values correspond to typical faults in industrial gas turbine systems.

IV. Conclusions and Future Directions

In this paper, we have provided a systematic overview of many known entropy measures, highlighting their applicability to machine fault diagnosis. The underlying principles of the fundamental definitions are reviewed, including Shannon entropy, approximate entropy, sample entropy, fuzzy entropy, permutation entropy, and the multiple-scale entropy measures. Their potential usages and roles in fault detection and diagnosis are summarized into three categories: entropy measure as a feature or health indicator, entropy criterion for wavelet parameter selection, and the usage of entropy in pattern recognition. These practices are complemented with case studies. The literature has shown that the entropy measures and their extensions are an effective and *low-cost* method for machine health monitoring and fault diagnosis, requiring little to none domain knowledge.

Although the entropy measures are indicative for machine condition monitoring, they only provide information on the uncertainties of the system, and therefore are normally accompanied with other machine learning techniques, specialized in fault classification. Entropy techniques are intuitive in nature and cost-effective in computation, as compared with deep learning techniques for instance. Therefore, they are suitable for *early stage* anomaly detection in an industrial system, but may not be sensitive enough for classifying specific types of fault.

Subsequently, potential future work in this research area is proposed as follows:

1) The development of entropy algorithms (i.e., single-scale and multiple-scale entropy measures) to further enhance performance in the complexity analysis of time series, such as improved reliability and robustness under noisy environmental conditions – while maintaining the computational efficiency of entropy analysis;
2) The investigation of parameter selection in entropy estimation procedures: more studies are needed to clarify parameter interaction (e.g., embedding dimension and time delay) in different entropy measures, and their effect in the algorithm performance towards assessing machine fault types and fault severity levels;
3) The extension of entropy usage from one-dimensional time series to two or more-dimensional data. There are interesting prospects on the usage of entropy techniques for image and video analysis (e.g., infrared thermal imaging) for fault diagnosis [118];
4) The application of entropy measures as non-linear feature indicators for fault severity assessment and unit remaining life estimation, by investigating the relation between the component defect progression and the entropy values;
5) The development of entropy-based criterion for specifying hyper-parameters in signal time-frequency analysis and selecting appropriate time-frequency components that contain crucial fault information;
6) The exploitation of multivariate entropy measures in machine health monitoring, as multichannel sensor data incorporate richer fault information in monitoring the system state simultaneously compared to a single sensor;
7) The application of entropy feature fusion from the feature-level or the decision-level, to yield more comprehensive feature indicators that incorporate less redundant information and achieve better diagnosis performance – in fault type identification towards machine fault diagnosis.

ABBREVIATIONS

λ	Time delay
τ	Scale factor
m	Embedding dimention
r	Tolerance
AI	Artificial Intelligence
ApEn	Approximate Entropy
CWT	Continuous Wavelet Transform
DWT	Discrete Wavelet Transform
FuzzyEn	Fuzzy Entropy
KS	Kolmogorov-Sinai
MFEn	Multiscale Fuzzy Entropy
MPEn	Multiscale Permutation Entropy
MSEn	Multiscale Entropy
PerEn	Permutation Entropy
SampEn	Sample Entropy
ShanEn	Shannon Entropy
WPT	Wavelet Packet Transform

REFERENCES

[1] A. Ben-Naim, *Information, Entropy, Life and the Universe: What We Know and What We Do Not Know*. World Scientific Publishing Co., Inc., 2015.

[2] E. T. Jaynes, "Information theory and statistical mechanics," *Physical review*, vol. 106, no. 4, p. 620, 1957.

[3] J. V. Stone, *Information theory: a tutorial introduction*. Sebtel Press, 2015.

[4] C. E. Shannon, "A mathematical theory of communication," *The Bell System Technical Journal*, vol. 27, no. 3, pp. 379–423, 1948.

[5] R. M. Gray, *Entropy and information theory*. Springer Science & Business Media, 2011.

[6] J.-P. Eckmann and D. Ruelle, "Ergodic theory of chaos and strange attractors," in *The Theory of Chaotic Attractors*. Springer, 1985, pp. 273–312.

[7] S. Pincus, "Approximate entropy as a measure of system complexity." *Proceedings of the National Academy of Sciences*, vol. 88, no. 6, pp. 2297–2301, 1991.

[8] J. S. Richman and J. R. Moorman, "Physiological time-series analysis using approximate entropy and sample entropy," *American Journal of Physiology-Heart and Circulatory Physiology*, vol. 278, no. 6, pp. H2039–H2049, 2000.

[9] W. Chen, Z. Wang, H. Xie, and W. Yu, "Characterization of surface EMG signal based on fuzzy entropy," *IEEE Transactions on neural systems and rehabilitation engineering*, vol. 15, no. 2, pp. 266–272, 2007.

[10] C. Bandt and B. Pompe, "Permutation entropy: a natural complexity measure for time series," *Phys. Rev. Lett.*, vol. 88, no. 17, p. 174102, 2002.

[11] M. Costa, A. L. Goldberger, and C.-K. Peng, "Multiscale entropy analysis of complex physiologic time series," *Phys. Rev. Lett.*, vol. 89, no. 6, p. 068102, 2002.

[12] M. Costa, A. L. Goldberger, and C.-K. Peng, "Multiscale entropy analysis of biological signals," *Physical review E*, vol. 71, no. 2, p. 021906, 2005.

[13] A. Humeau-Heurtier, "The multiscale entropy algorithm and its variants: a review," *Entropy*, vol. 17, no. 5, pp. 3110–3123, 2015.

[14] M. Hu and H. Liang, *Multiscale Entropy: Recent Advances*. Cham: Springer International Publishing, 2017, pp. 115–138.

[15] R. Yan, Y. Liu, and R. X. Gao, "Permutation entropy: a nonlinear statistical measure for status characterization of rotary machines," *Mechanical Systems and Signal Processing*, vol. 29, pp. 474–484, 2012.

[16] A. K. Jardine, D. Lin, and D. Banjevic, "A review on machinery diagnostics and prognostics implementing condition-based maintenance," *Mechanical systems and signal processing*, vol. 20, no. 7, pp. 1483–1510, 2006.

[17] Y. Li, X. Wang, Z. Liu, X. Liang, and S. Si, "The entropy algorithm and its variants in the fault diagnosis of rotating machinery: A review," *IEEE Access*, vol. 6, pp. 66 723–66 741, 2018.

[18] R. Liu, B. Yang, E. Zio, and X. Chen, "Artificial intelligence for fault diagnosis of rotating machinery: A review," *Mechanical Systems and Signal Processing*, vol. 108, pp. 33–47, 2018.

[19] I. Howard, "A review of rolling element bearing vibration 'detection, diagnosis and prognosis'," DTIC Document, Tech. Rep., 1994.

[20] R. Yan and R. X. Gao, "Complexity as a measure for machine health evaluation," *IEEE Transactions on instrumentation and measurement*, vol. 53, no. 4, pp. 1327–1334, 2004.

[21] R. Yan and R. X. Gao, "Approximate entropy as a diagnostic tool for machine health monitoring," *Mechanical Systems and Signal Processing*, vol. 21, no. 2, pp. 824–839, 2007.

[22] A. L. Berger, V. J. D. Pietra, and S. A. D. Pietra, "A maximum entropy approach to natural language processing," *Computational linguistics*, vol. 22, no. 1, pp. 39–71, 1996.

[23] M. Borowska, "Entropy-based algorithms in the analysis of biomedical signals," *Studies in Logic, Grammar and Rhetoric*, vol. 43, no. 1, pp. 21–32, 2015.

[24] R. Zhou, R. Cai, and G. Tong, "Applications of entropy in finance: A review," *Entropy*, vol. 15, no. 11, pp. 4909–4931, 2013.

[25] N. Georgescu-Roegen, "The entropy law and the economic problem," *Valuing the earth: Economics, ecology, ethics*, pp. 75–88, 1993.

[26] D. R. Brooks, E. O. Wiley, and D. Brooks, *Evolution as entropy*. University of Chicago Press Chicago, 1988.

[27] M. Martínez-García and T. Gordon, "A new model of human steering using far-point error perception and multiplicative control," in *2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE, 2018, pp. 1245–1250.

[28] M. Martínez-García, Y. Zhang, and T. Gordon, "Memory pattern identification for feedback tracking control in human–machine systems," *Human factors*, 2019.

[29] L. Warren, "A new interpretation of the Shannon entropy measure," Defence Science and Technology Group Edinburgh (Australia), Tech. Rep., 2015.

[30] M. Zanin, L. Zunino, O. A. Rosso, and D. Papo, "Permutation entropy and its main biomedical and econophysics applications: a review," *Entropy*, vol. 14, no. 8, pp. 1553–1577, 2012.

[31] R. Alcaraz and J. J. Rieta, "A review on sample entropy applications for the non-invasive analysis of atrial fibrillation electrocardiograms," *Biomedical Signal Processing and Control*, vol. 5, no. 1, pp. 1–14, 2010.

[32] M. Kedadouche, M. Thomas, A. Tahan, and R. Guilbault, "Nonlinear parameters for monitoring gear: comparison between Lempel-Ziv, approximate entropy, and sample entropy complexity," *Shock and Vibration*, vol. 2015, 2015.

[33] S. Lu, X. Chen, J. K. Kanters, I. C. Solomon, and K. H. Chon, "Automatic selection of the threshold value r for approximate entropy," *IEEE Transactions on Biomedical Engineering*, vol. 55, no. 8, pp. 1966–1972, 2008.

[34] M. Riedl, A. Müller, and N. Wessel, "Practical considerations of permutation entropy," *The European Physical Journal Special Topics*, vol. 222, no. 2, pp. 249–262, 2013.

[35] K. Keller, T. Mangold, I. Stolz, and J. Werner, "Permutation entropy: New ideas and challenges," *Entropy*, vol. 19, no. 3, p. 134, 2017.

[36] W. Aziz and M. Arif, "Multiscale permutation entropy of physiological time series," in *9th International Multitopic Conference, IEEE INMIC 2005*. IEEE, 2005, pp. 1–6.

[37] Y.-H. Pan, Y.-H. Wang, S.-F. Liang, and K.-T. Lee, "Fast computation of sample entropy and approximate entropy in biomedicine," *Computer methods and programs in biomedicine*, vol. 104, no. 3, pp. 382–396, 2011.

[38] K. Keller, A. Unakafov, and V. Unakafova, "Ordinal patterns, entropy, and EEG," *Entropy*, vol. 16, no. 12, pp. 6212–6239, 2014.

[39] G. Doquire and M. Verleysen, "Mutual information-based feature selection for multilabel classification," *Neurocomputing*, vol. 122, pp. 148–155, 2013.

[40] G. Sinai Ya, "On the concept of entropy for dynamical systems (in Russian) Dokl," *Akad. Nauk SSSR*, vol. 124, pp. 768–771, 1959.

[41] R. Frigg, "In what sense is the Kolmogorov-Sinai entropy a measure for chaotic behaviour? – bridging the gap between dynamical systems theory and communication theory," *The British journal for the philosophy of science*, vol. 55, no. 3, pp. 411–434, 2004.

[42] C. Beisbart and S. Hartmann, *Probabilities in physics*. Oxford University Press, 2011.

[43] Y. B. Pesin, "Characteristic Lyapunov exponents and smooth ergodic theory," *Uspekhi Matematicheskikh Nauk*, vol. 32, no. 4, pp. 55–112, 1977.

[44] P. Grassberger and I. Procaccia, "Estimation of the Kolmogorov entropy from a chaotic signal," *Physical review A*, vol. 28, no. 4, p. 2591, 1983.

[45] A. Rényi, "On measures of entropy and information," in: Fourth Berkeley Symposium on Mathematics, Statistics and Probability, Tech. Rep., 1960.

[46] S. M. Pincus and A. L. Goldberger, "Physiological time-series analysis: what does regularity quantify?" *American Journal of Physiology-Heart and Circulatory Physiology*, vol. 266, no. 4, pp. H1643–H1656, 1994.

[47] F. Takens, "Detecting strange attractors in turbulence," in *Dynamical systems and turbulence, Warwick 1980*. Springer, 1981, pp. 366–381.

[48] S. Pincus, "Approximate entropy (ApEn) as a complexity measure," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 5, no. 1, pp. 110–117, 1995.

[49] F. Kaffashi, R. Foglyano, C. Wilson, and K. Loparo, "The effect of time delay on approximate & sample entropy calculations," *Physica D: Nonlinear Phenomena*, vol. 237, no. 23, pp. 3069–3074, 2008.

[50] S. M. Pincus, T. Mulligan, A. Iranmanesh, S. Gheorghiu, M. Godschalk, and J. D. Veldhuis, "Older males secrete luteinizing hormone and testosterone more irregularly, and jointly more asynchronously, than younger males," *Proceedings of the National Academy of Sciences*, vol. 93, no. 24, pp. 14 100–14 105, 1996.

[51] J. S. Richman, D. E. Lake, and J. R. Moorman, "Sample entropy," in *Methods in enzymology*. Elsevier, 2004, vol. 384, pp. 172–184.

[52] J. Yentes, N. Hunt, K. Schmid, J. Kaipust, D. McGrath, and N. Stergiou, "The appropriate use of approximate entropy and sample entropy with short data sets," *Annals of biomedical engineering*, vol. 41, no. 2, pp. 349–365, 2013.

[53] J. Yentes, W. Denton, J. McCamley, P. C. Raffalt, and K. Schmid, "Effect of parameter selection on entropy calculation for long walking trials," *Gait & posture*, vol. 60, pp. 128–134, 2018.

[54] Y. Lu, M. Wang, R. Peng, and Q. Zhang, "Accelerating the computation of entropy measures by exploiting vectors with dissimilarity," *Entropy*, vol. 19, no. 11, p. 598, 2017.

[55] G. Manis, M. Aktaruzzaman, and R. Sassi, "Low computational cost for sample entropy," *Entropy*, vol. 20, no. 1, p. 61, 2018.

[56] L. Silva, A. Senra Filho, V. Fazan, J. Felipe, and L. Junior, "Two-dimensional sample entropy: assessing image texture through irregularity," *Biomedical Physics & Engineering Express*, vol. 2, no. 4, p. 045002, 2016.

[57] G. Xiong, L. Zhang, H. Liu, H. Zou, and W. Guo, "A comparative study on ApEn, SampEn and their fuzzy counterparts in a multiscale framework for feature extraction," *Journal of Zhejiang University SCIENCE A*, vol. 11, no. 4, pp. 270–279, Apr 2010.

[58] J. Zheng, J. Cheng, Y. Yang, and S. Luo, "A rolling bearing fault diagnosis method based on multi-scale fuzzy entropy and variable predictive model-based class discrimination," *Mechanism and Machine Theory*, vol. 78, pp. 187–200, 2014.

[59] L. Ji, P. Li, K. Li, X. Wang, and C. Liu, "Analysis of short-term heart rate and diastolic period variability using a refined fuzzy entropy method," *Biomedical engineering online*, vol. 14, no. 1, p. 64, 2015.

[60] J.-M. Girault and A. Humeau-Heurtier, "Centered and averaged fuzzy entropy to improve fuzzy entropy precision," *Entropy*, vol. 20, no. 4, p. 287, 2018.

[61] T. Zhang, W. Chen, and M. Li, "Fuzzy distribution entropy and its application in automated seizure detection technique," *Biomedical Signal Processing and Control*, vol. 39, pp. 360–377, 2018.

[62] W. Chen, J. Zhuang, W. Yu, and Z. Wang, "Measuring complexity using FuzzyEn, ApEn, and SampEn," *Medical Engineering and Physics*, vol. 31, no. 1, pp. 61–68, 2009.

[63] S. Simons, P. Espino, and D. Abásolo, "Fuzzy entropy analysis of the electroencephalogram in patients with Alzheimer's disease: Is the method superior to sample entropy?" *Entropy*, vol. 20, no. 1, p. 21, 2018.

[64] K. Keller and M. Sinn, "Kolmogorov–Sinai entropy from the ordinal viewpoint," *Physica D: Nonlinear Phenomena*, vol. 239, no. 12, pp. 997–1000, 2010.

[65] X. Liu and Y. Wang, "Fine-grained permutation entropy as a measure of natural complexity for time series," *Chinese Physics B*, vol. 18, no. 7, p. 2690, 2009.

[66] B. Fadlallah, B. Chen, A. Keil, and J. Príncipe, "Weighted-permutation entropy: A complexity measure for time series incorporating amplitude information," *Phys. Rev. E*, vol. 87, p. 022911, 2013.

[67] H. Azami and J. Escudero, "Amplitude-aware permutation entropy: Illustration in spike detection and signal segmentation," *Computer methods and programs in biomedicine*, vol. 128, pp. 40–51, 2016.

[68] Z. Huo, Y. Zhang, L. Shu, and X. Liao, "Edge permutation entropy: An improved entropy measure for time-series analysis," in *IECON 2019 - 45th Annual Conference of the IEEE Industrial Electronics Society*, vol. 1, Oct 2019, pp. 5998–6003.

[69] C. Bian, C. Qin, Q. Ma, and Q. Shen, "Modified permutation-entropy analysis of heartbeat dynamics," *Phys. Rev. E*, vol. 85, p. 021906, 2012.

[70] J. F. Valencia, A. Porta, M. Vallverdu, F. Claria, R. Baranowski, E. Orlowska-Baranowska, and P. Caminal, "Refined multiscale entropy: Application to 24-h holter recordings of heart period variability in healthy and aortic stenosis subjects," *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 9, pp. 2202–2213, 2009.

[71] V. V. Nikulin and T. Brismar, "Comment on "multiscale entropy analysis of complex physiologic time series"," *Phys. Rev. Lett.*, vol. 92, p. 089803, 2004.

[72] S.-D. Wu, C.-W. Wu, S.-G. Lin, C.-C. Wang, and K.-Y. Lee, "Time series analysis using composite multiscale entropy," *Entropy*, vol. 15, no. 3, pp. 1069–1084, 2013.

[73] M. D. Costa and A. L. Goldberger, "Generalized multiscale entropy analysis: application to quantifying the complex volatility of human heartbeat time series," *Entropy*, vol. 17, no. 3, pp. 1197–1203, 2015.

[74] S.-D. Wu, C.-W. Wu, S.-G. Lin, K.-Y. Lee, and C.-K. Peng, "Analysis of complex time series using refined composite multiscale entropy," *Physics Letters A*, vol. 378, no. 20, pp. 1369–1374, 2014.

[75] A. Humeau-Heurtier, C. Wu, and S. Wu, "Refined composite multiscale permutation entropy to overcome multiscale permutation entropy length dependence," *IEEE Signal Processing Letters*, vol. 22, no. 12, pp. 2364–2367, Dec 2015.

[76] H. Zhao, M. Sun, W. Deng, and X. Yang, "A new feature extraction method based on EEMD and multi-scale fuzzy entropy for motor bearing," *Entropy*, vol. 19, no. 1, p. 14, 2016.

[77] Y. Li, Y. Yang, G. Li, M. Xu, and W. Huang, "A fault diagnosis scheme for planetary gearboxes using modified multi-scale symbolic dynamic entropy and mRMR feature selection," *Mechanical Systems and Signal Processing*, vol. 91, pp. 295–312, 2017.

[78] Y. Jiang, C.-K. Peng, and Y. Xu, "Hierarchical entropy analysis for biological signals," *Journal of Computational and Applied Mathematics*, vol. 236, no. 5, pp. 728–742, 2011.

[79] Z. Huo, Y. Zhang, L. Shu, and M. Gallimore, "A new bearing fault diagnosis method based on fine-to-coarse multiscale permutation entropy, Laplacian score and SVM," *IEEE Access*, vol. 7, pp. 17 050–17 066, 2019.

[80] M. U. Ahmed and D. P. Mandic, "Multivariate multiscale entropy analysis," *IEEE Signal Processing Letters*, vol. 19, no. 2, pp. 91–94, 2011.

[81] J. Zheng, D. Tu, H. Pan, X. Hu, T. Liu, and Q. Liu, "A refined composite multivariate multiscale fuzzy entropy and Laplacian score-based fault diagnosis method for rolling bearings," *Entropy*, vol. 19, no. 11, p. 585, 2017.

[82] H. Azami and J. Escudero, "Refined composite multivariate generalized multiscale fuzzy entropy: A tool for complexity analysis of multichannel signals," *Physica A: Statistical Mechanics and its Applications*, vol. 465, pp. 261–276, 2017.

[83] X. Liu, A. Jiang, N. Xu, and J. Xue, "Increment entropy as a measure of complexity for time series," *Entropy*, vol. 18, no. 1, p. 22, 2016.

[84] P. Li, K. Li, C. Liu, D. Zheng, Z. Li, and C. Liu, "Detection of coupling in short physiological series by a joint distribution entropy method," *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 11, pp. 2231–2242, 2016.

[85] M. Rostaghi and H. Azami, "Dispersion entropy: A measure for time-series analysis," *IEEE Signal Processing Letters*, vol. 23, no. 5, pp. 610–614, 2016.

[86] Q. Liu, H. Pan, J. Zheng, J. Tong, and J. Bao, "Composite interpolation-based multiscale fuzzy entropy and its application to fault diagnosis of rolling bearing," *Entropy*, vol. 21, no. 3, p. 292, 2019.

[87] H. Azami, S. E. Arnold, S. Sanei, Z. Chang, G. Sapiro, J. Escudero, and A. S. Gupta, "Multiscale fluctuation-based dispersion entropy and its applications to neurological diseases," *IEEE Access*, vol. 7, pp. 68 718–68 733, 2019.

[88] V. Tran, B. Yang, M. Oh, and A. Tan, "Fault diagnosis of induction motor based on decision trees and adaptive neuro-fuzzy inference," *Expert Systems with Applications*, vol. 36, no. 2, Part 1, pp. 1840 – 1849, 2009.

[89] J. Ma, J. Wu, and X. Wang, "Fault diagnosis method based on wavelet packet-energy entropy and fuzzy kernel extreme learning machine," *Advances in Mechanical Engineering*, vol. 10, no. 1, p. 1687814017751446, 2018.

[90] Z. Zheng and G. Xin, "Fault feature extraction of hydraulic pumps based on symplectic geometry mode decomposition and power spectral entropy," *Entropy*, vol. 21, no. 5, p. 476, 2019.
[91] Y. Wang, B. Tang, Y. Qin, and T. Huang, "Rolling bearing fault detection of civil aircraft engine based on adaptive estimation of instantaneous angular speed," *IEEE Transactions on Industrial Informatics*, 2019.
[92] M. S. Hoseinzadeh, S. E. Khadem, and M. S. Sadooghi, "Modifying the Hilbert-Huang transform using the nonlinear entropy-based features for early fault detection of ball bearings," *Applied Acoustics*, vol. 150, pp. 313–324, 2019.
[93] F. Wang, C. Liu, W. Su, Z. Xue, Q. Han, and H. Li, "Combined failure diagnosis of slewing bearings based on MCKD-CEEMD-ApEn," *Shock and Vibration*, vol. 2018, 2018.
[94] J. Chen, D. Zhou, C. Lyu, and C. Lu, "An integrated method based on CEEMD-SampEn and the correlation analysis algorithm for the fault diagnosis of a gearbox under different working conditions," *Mechanical Systems and Signal Processing*, vol. 113, pp. 102–111, 2018.
[95] J. Zheng, J. Cheng, and Y. Yang, "A rolling bearing fault diagnosis approach based on LCD and fuzzy entropy," *Mechanism and Machine Theory*, vol. 70, pp. 441–453, 2013.
[96] D. Sampaio and R. Nicoletti, "Detection of cracks in shafts with the approximated entropy algorithm," *Mechanical Systems and Signal Processing*, vol. 72, pp. 286–302, 2016.
[97] S. Zhou, S. Qian, W. Chang, Y. Xiao, and Y. Cheng, "A novel bearing multi-fault diagnosis approach based on weighted permutation entropy and an improved SVM ensemble classifier," *Sensors*, vol. 18, no. 6, p. 1934, 2018.
[98] S. Nandi, H. A. Toliyat, and X. Li, "Condition monitoring and fault diagnosis of electrical motors - A review," *IEEE transactions on energy conversion*, vol. 20, no. 4, pp. 719–729, 2005.
[99] Z. Huo, Y. Zhang, P. Francq, L. Shu, and J. Huang, "Incipient fault diagnosis of roller bearing using optimized wavelet transform based multi-speed vibration signatures," *IEEE Access*, vol. 5, pp. 19 442–19 456, 2017.
[100] Y. Li, X. Wang, S. Si, and S. Huang, "Entropy based fault classification using the Case Western Reserve University data: A benchmark study," *IEEE Transactions on Reliability*, 2019.
[101] Y. Yang, H. Zheng, J. Yin, M. Xu, and Y. Chen, "Refined composite multivariate multiscale symbolic dynamic entropy and its application to fault diagnosis of rotating machine," *Measurement*, vol. 151, p. 107233, 2020.
[102] Y. Ye, Y. Zhang, Q. Wang, Z. Wang, Z. Teng, and H. Zhang, "Fault diagnosis of high-speed train suspension systems using multiscale permutation entropy and linear local tangent space alignment," *Mechanical Systems and Signal Processing*, vol. 138, p. 106565, 2020.
[103] J. Zheng, H. Pan, S. Yang, and J. Cheng, "Generalized composite multiscale permutation entropy and Laplacian score based rolling bearing fault diagnosis," *Mechanical Systems and Signal Processing*, vol. 99, pp. 229–243, 2018.
[104] K. Zhu, X. Song, and D. Xue, "A roller bearing fault diagnosis method based on hierarchical entropy and support vector machine with particle swarm optimization algorithm," *Measurement*, vol. 47, pp. 669–675, 2014.
[105] Y. Li, G. Li, Y. Yang, X. Liang, and M. Xu, "A fault diagnosis scheme for planetary gearboxes using adaptive multi-scale morphology filter and modified hierarchical permutation entropy," *Mechanical Systems and Signal Processing*, vol. 105, pp. 319–337, 2018.
[106] R. X. Gao and R. Yan, *Wavelets: Theory and applications for manufacturing*. Springer Science & Business Media, 2010.
[107] R. Yan, R. X. Gao, and X. Chen, "Wavelets for fault diagnosis of rotary machines: A review with applications," *Signal Processing*, vol. 96, pp. 1–15, 2014.
[108] R. Yan, "Base wavelet selection criteria for non-stationary vibration analysis in bearing health diagnosis," *Electronic Doctoral Dissertations for UMass Amherst*, 2007.
[109] R. Coifman and M. Wickerhauser, "Entropy-based algorithms for best basis selection," *IEEE Transactions on Information Theory*, vol. 38, no. 2, pp. 713–718, 1992.
[110] P. Kankar, S. Sharma, and S. P. Harsha, "Fault diagnosis of ball bearings using continuous wavelet transform," *Applied Soft Computing*, vol. 11, no. 2, pp. 2300–2312, 2011.
[111] S. Singh and N. Kumar, "Detection of bearing faults in mechanical systems using stator current monitoring," *IEEE Transactions on Industrial Informatics*, vol. 13, no. 3, pp. 1341–1349, 2016.
[112] K. Deák, T. Mankovits, and I. Kocsis, "Optimal wavelet selection for the size estimation of manufacturing defects of tapered roller bearings with vibration measurement using Shannon entropy criteria," *Strojniski Vestnik/Journal of Mechanical Engineering*, vol. 63, no. 1, 2017.
[113] S. Wan, X. Zhang, and L. Dou, "Shannon entropy of binary wavelet packet subbands and its application in bearing fault extraction," *Entropy*, vol. 20, no. 4, p. 260, 2018.
[114] R. Xiao, Q. Hu, and J. Li, "Leak detection of gas pipelines using acoustic signals based on wavelet transform and support vector machine," *Measurement*, vol. 146, pp. 479–489, 2019.
[115] A. Darji, P. Darji, and D. Pandya, "Fault diagnosis of ball bearing with WPT and supervised machine learning techniques," in *Machine Intelligence and Signal Analysis*. Springer, 2019, pp. 291–301.
[116] B. Wang, Y. Lei, N. Li, and N. Li, "A hybrid prognostics approach for estimating remaining useful life of rolling element bearings," *IEEE Transactions on Reliability*, 2018.
[117] R. Zhao, R. Yan, Z. Chen, K. Mao, P. Wang, and R. X. Gao, "Deep learning and its applications to machine health monitoring: A survey," *arXiv preprint arXiv:1612.07640*, 2016.
[118] M. Martinez-Garcia, Y. Zhang, J. Wan, and J. Mcginty, "Visually interpretable profile extraction with an autoencoder for health monitoring of industrial systems," in *2019 IEEE 4th International Conference on Advanced Robotics and Mechatronics (ICARM)*. IEEE, 2019, pp. 649–654.
[119] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. MIT press, 2016.
[120] D. M. Kline and V. L. Berardi, "Revisiting squared-error and cross-entropy functions for training neural network classifiers," *Neural Computing & Applications*, vol. 14, no. 4, pp. 310–318, 2005.
[121] S. Shao, R. Yan, Y. Lu, P. Wang, and R. Gao, "DCNN-based multi-signal induction motor fault diagnosis," *IEEE Transactions on Instrumentation and Measurement*, 2019.
[122] X. Li, W. Zhang, N.-X. Xu, and Q. Ding, "Deep learning-based machinery fault diagnostics with domain adaptation across sensors at different places," *IEEE Transactions on Industrial Electronics*, 2019.
[123] M. Zhao, S. Zhong, X. Fu, B. Tang, and M. Pecht, "Deep residual shrinkage networks for fault diagnosis," *IEEE Transactions on Industrial Informatics*, 2019.
[124] B. Yang, Y. Lei, F. Jia, and S. Xing, "An intelligent fault diagnosis approach based on transfer learning from laboratory bearings to locomotive bearings," *Mechanical Systems and Signal Processing*, vol. 122, pp. 692–706, 2019.
[125] R. Yan, F. Shen, C. Sun, and X. Chen, "Knowledge transfer for rotary machine fault diagnosis," *IEEE Sensors Journal*, 2019.
[126] M. Martínez-García, Y. Zhang, K. Suzuki, and Z. Yudong, "Measuring system entropy with a deep recurrent neural network model," in *Industrial Informatics (INDIN), 2019 IEEE International Conference on*. Helsinki-Espoo, Finland: IEEE, 2019.