A systematic review of fuzzy formalisms for bearing fault diagnosis

Chuan Li, José Valente de Oliveira, Mariela Cerrada, Diego Cabrera, René Vinicio Sánchez, and Grover Zurita

***Abstract*—Bearings are fundamental mechanical components in rotary machines (engines, gearboxes, generators, radars, turbines, etc) that have been identified as one of the primary causes of failure in these machines. This makes bearing fault diagnosis (detection, classification and prognosis) an economic very relevant topic, as well as a technically challenging one as evaluated by the extensive research literature on the subject. This paper employs a systematic methodology to identify, summarize, analyze, and interpret the primary literature on fuzzy formalisms for bearing fault diagnosis from 2000 to 2017 (March). The main contribution is an updated, unbiased and (to a higher extend) repeatable search, review and analysis (summary, classification, and critique) of the available approaches resorting to fuzzy formalisms in this trendy topic. A discussion on a new promising future research direction is provided. A comprehensive list of references is also included.**

***Index Terms*—Bearing, fault detection, fault diagnosis, fault classification, fault prognosis, fuzzy entropy, fuzzy rules, fuzzy clustering.**

I. Introduction

A bearing is a mechanical component used to reduce friction between other moving parts and thus one of the most common components of rotary machinery. To recognize the importance of such components, it is stressed that most of world energy production, consumption, and transformation relies on rotating machinery such as alternators, compressors, motors, or (wind) turbines. Bearings have been identified as the primary cause of failure in some of such machinery [1], [2]. Bearings are especially prone to faults due to diverse causes such as excessive load, lubricant failure, corrosion or fatigue. Downtime for replacement is the least problem of such faults. Should the engine bearing supporting a crankshaft break and the whole engine can disintegrate. In general, faults result in abrasion due to steady friction of mechanical parts that, in turn, can have severe consequences for the overall system where the bearing is working in. Therefore, it is apparent the need for early diagnosis of such faults.

By fault diagnosis we mean one of the following tasks: i) fault detection, ii) fault classification, or iii) fault prognosis. Informally, fault detection refers to the real-time signal processing required to know whether or not a bearing is in its healthy (normal operating) state. Fault classification refers to the determination of the type of faults an unhealthy bearing is suffering from and is a pattern recognition task. Fault prognosis refers to the forecast of the remaining useful life of a bearing and is based on dynamic modeling.

Bearing fault diagnosis involves the following typical data pipeline: data acquisition and conditioning, feature extraction, feature selection, and a final detection/classification/forecast stage (Fig. 1). Data acquisition can resort to different types of

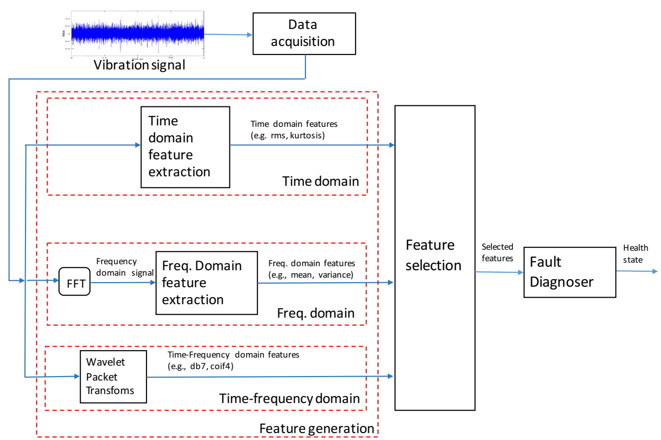

Fig. 1: A typical pipeline for vibration based bearing fault diagnosis.

sensors including acoustic, electric, oil (debris) spectrometers, thermal imaging, or radial accelerations sensors (accelerometers) for vibration monitoring [3], [4]. Comparatively, vibration measurement and analysis is a well-known, cost-effective, and widely applied technique in bearing fault diagnosis [5]. Features are extracted from collected vibration signals. Typically, time, frequency, and time-frequency features are computed. In the time domain, statistics such as root mean squares, kurtosis, skewness, or the crest factor are computed. Fourier transforms are usually employed for converting time-domain signals to frequency domain where further processing occurs. This includes the partition of the signals in frequency bands and computation of features, e.g., mean, standard deviation, for each one of the bands. For the time-frequency domain, wavelets analysis is the preferred method [6].

Similar to most pattern recognition problems, feature selection is a critical step for optimizing efficiency, accuracy and for mitigating overtraining. Feature selection can be accomplished by experts with or without the help of feature selection methods. These include the employment of genetic algorithms, e.g., [7], correlation-based methods such principal component analysis, e.g., [8], [9], [10], fuzzy measures, e.g.,

C. Li is with National Research Base of Intelligent Manufacturing Service, Chongqing Technology and Business University, China

J.V. de Oliveira is visiting the National Research Base of Intelligent Manufacturing Service, Chongqing Technology and Business University, China, on Sabbatical leave from CEOT, Universidade do Algarve, Portugal

M. Cerrada, D. Cabrera, and R.V. Sánchez are with the GIDTEC, Universidad Politécnica Salesiana, Cuenca, Ecuador

G. Zurita is with Universidad Privada Boliviana, Cochabamba, Bolivia

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2018.2878200, IEEE Transactions on Fuzzy Systems

[11], rough sets, e.g., [12], entropy based criteria like those used for growing decision trees, e.g., [13], or orthogonal fuzzy neighborhood discriminant analysis [14], [15]. The literature on bearing fault detection and diagnosis is extensive. This paper focuses on the employment of fuzzy formalisms in the different phases of bearing fault diagnosis (Fig. 1). No restrictions are applied on the type of bearings nor on the type and location of the faults. The first works seem to have appeared as early as 1994 [16] and since then many real world applications have been reported including fuzzy fault diagnosis of induction motors [17], [18], [19], [20], machine tool spindle head ball bearing [21], three-phase brush less DC motor [22] and motor bearing shield [23], also motor roller bearing [24], crank bearing of emulsion pumps [25], bearings of railway locomotives [26], the wearing of a thrust bearing of an industrial multishaft centrifugal compressor located in an air separation unit of an integrated gasification and combined cycle section of a refinement plant [27], and crankshaft bearings of Diesel engines [28].

The main motivation for using fuzzy formalisms in this domain concerns their ability to deal with the inherent uncertainty of the feature space. Actually, i) the number of faulty samples available is typically much smaller than the number of healthy samples; ii) there is no guarantee that all relevant features are fully observable; iii) the interference between different faults is not easily identifiable; and iv) the measurement is non-stationary, noisy, and often redundant. Therefore, the knowledge that a fault diagnoser has about the system is necessarily incomplete and ambiguous. However, another important advantage of fuzzy formalisms when compared with other nonlinear diagnosis techniques such as artificial neural networks, is that fuzzy models can provide an insight on the linguistic relationship between the variables [29]; an issue that is often forgotten in bearing fault literature with some notable exceptions [30], [31].

Our research questions can be stated as follows: i) what are the main fuzzy formalisms employed in bearing fault diagnosis? and ii) in which tasks of the faults diagnosis process are these formalisms being used?

The main contributions of this work are: i) a first systematic literature review on this matter and ii) a proposal of a new research direction. Our motivation is threefold:

i) To summarize, categorize, and interpret the extensive primary literature on this trendy topic;
ii) To provide a sound background for new studies on the subject;
iii) and to identify gaps in the current literature and suggest new directions of research.

The paper follows the typical layout of a systematic review paper and also includes a self-contained summary of the identified fuzzy formalisms. Section II briefly describes the type of bearings, their anatomy, and their faults. Section III presents the used sources and the search methodology. Section IV offers an overview of the selected formalisms. Whenever a formalism is used in more than one activity of the fault diagnosis pipeline (Fig. 1) a summary of its applications is also provided. Section V presents the motivations underlying the used fuzzy formalism(s) for each fault diagnosis activity, how a formalism can be used for a given activity, and a discussion on the comparison with other applicable techniques, when available in the respective literature. Section VI proposes a new promising research direction. Finally, Section VII ends the paper with the main conclusions.

II. On bearings and their faults

There are several types of bearings ranging from the material that are made of, i.e., metal, polymers, ceramics, and composites, to the type of components, e.g., rolling balls, cylinders, tapered, needles, or barrel rollers. Bearings are also quite diverse in terms of size, outer diameters ranging from 1 mm up to 4 meters.

Metal ball rolling bearings are by far the most common type of bearings found in the reviewed literature and thus here we concentrate only on these. This type of bearings consists of an inner, an outer race or ring, inside which a set of balls rotate. In some models, a cage holds and keeps the balls equally spaced preventing internal collisions; In ball bearings, a fault can be classified according to the location where it occurs: at the inner race, outer race, or at the rolling body, cage included. Cage failure is normally secondary, i.e., it is due to the failure of the other 3 main bearing components and, as such, cage faults are not normally studied in bearing diagnosis.

Fig. 2: A single-point fault in a bearing outer race.

Also, a fault can be classified according to its type: it can be i) a single point (Fig. 2), ii) multi-point, or iii) a generalized roughness (distributed) fault. A single point defect is defined by a unique, well-localized defect on an otherwise intact bearing surface. Examples of such defects include cracks, pits, holes, or spalls. Often, generalized faults originate from contamination, lack or loss of lubricant, shaft currents, misalignment, or more frequently simply by the detrioration of single point faults. Different faults can and do occur simultaneously [32]. Most of the reviewed literature concerns the diagnosis of single point faults.

III. Sources and search methodology

In order to attempt an unbiased and, as reproducible as possible, literature review, in this section we apply a systematic literature review of studies on fuzzy formalism based bearing fault diagnosis. The reviewed works are from 2000 to 2017 (March). The following scientific index engines and search strings were used:

- ACM [1]: (Title:fuzzy) and (Title:fault or Title:bearing) and (Keywords:fuzzy OR Keywords:bearing OR Keywords:fault)

[1] http://dl.acm.org/

1063-6706 (c) 2018 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2018.2878200, IEEE Transactions on Fuzzy Systems

- ieeexplore [2]: ((fuzzy) AND bearing fault) and refined by Year: 2000-2017
- Science direct [3]: pub-date > 1999 and TITLE-ABSTR-KEY(fuzzy) and TITLE-ABSTR-KEY(bearing fault)
- Scopus [4]: TITLE-ABS-KEY (fuzzy AND fault bearing) AND PUBYEAR > 1999

To be included in this review, a work must meet cumulatively the following selection criteria: i) it must deal with a fuzzy formalism based bearing fault diagnosis. No restriction is applied on the type of fuzzy formalism employed, on the type of bearings, nor on the type of faults; ii) it must be an English peer reviewed workshop, conference, edited volume contribution, or journal paper. The selection was performed by reading the abstract and, in some cases, the sections Introduction and Conclusions for improving the selection accuracy. A first observation is that, most of the selected results from ACM, ieeexplore, and Science Direct results are also included in the selected results from Scopus. Another observation is that in the last decades, there was a clear increasing tendency in the number of published works on this topic (Fig. 3) revealing an increased interest on the subject.

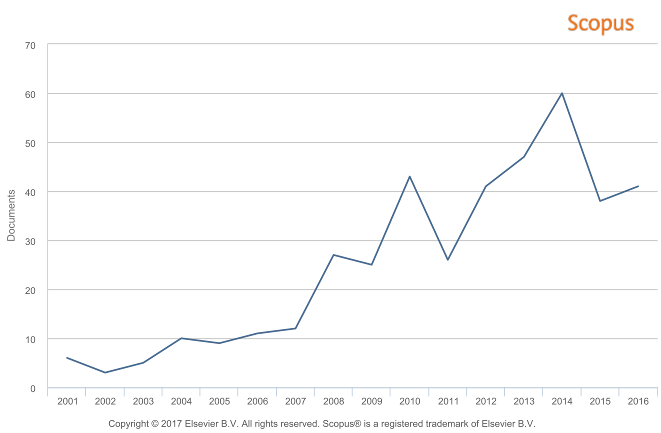

Fig. 3: Evolution of the number of published documents on fuzzy formalism based bearing fault diagnosis as retrieved by the search string used in Scopus. A clear long-term increasing tendency is observed.

IV. The Formalisms

We identify sixteen main categories of fuzzy formalisms applied to bearing fault diagnosis (Table I). The neuro-fuzzy approach is the most applied approach appearing in 32.1% of the papers. The *Adaptive Neuro Fuzzy Inference System* (ANFIS) appears in 38.9% of the studies within this category. Studies on ANFIS applications to this domain seem to have appeared as early as 1994 [16] and are still very popular nowadays, e.g. [33], [34], [35], [36]. Less common within this category is the fuzzy ARTMAP network, e.g., [37] and the fuzzy lattice neurocomputing [38]. Clustering is also among the most applied formalisms appearing in 23.8% of the selected works. The well-known fuzzy c-means (FCM) algorithm is used in 72.5% of the works in this category. The application of fuzzy entropy has also been extensively studied representing 10.2% of the selected studies. In this approach, fuzzy entropy is used as a feature (measure of the complexity) of vibrational signals. The employment of explicit fuzzy if-then rules and fuzzy inference systems, is found in 9.0% of the selected works. Fuzzy support vector machines [39], a fuzzy extension of the well-known maximal margin classifier, SVM are investigated in 3.6% of the selected papers. Together with fuzzy support vector data description represent 4.8% of the selected works. Possibility and Dempster-Shafer evidence theory together were applied in 6.0% of the selected works. Fuzzy relation equations, an alternative inference formalism to fuzzy rule based systems, rank 7th with 3.0% of share. Fuzzy numbers rank 8th *ex aequo* with (fuzzy) rough sets both with 2.4% of the selected works. Others, less pursued and thus more original approaches, include fuzzy k-nearest neighbor, the Sugeno integral, fuzzy grey-optimization methods, and probabilistic fuzzy systems. The following sections discuss each one of the main categories, following the ranking presented in Table I.

TABLE I: Most employed fuzzy formalisms in bearing fault diagnosis from the year 2000 onward.

Rank	Formalism	No papers	Share
1	Neuro-fuzzy	54	32.1%
2	Fuzzy clustering	40	23.8%
3	Fuzzy entropy	17	10.1%
4	Fuzzy rule based	15	8.9%
5	Fuzzy SVM and SVDD	8	4.8%
6	Possibility theory	7	4.2%
7	Fuzzy relations	5	3.0%
8	(Fuzzy-)rough sets	4	2.4%
8	Fuzzy numbers	4	2.4%
10	Evidence theory	3	1.8%
10	Fuzzy discriminant analysis	3	1.8%
10	Type-2 fuzzy sets	3	1.8%
13	Fuzzy k-nearest neighbor	2	1.2%
14	Sugeno integral	1	0.6%
14	Fuzzy-grey models	1	0.6%
14	Probabilistic fuzzy systems	1	0.6%

A. The neuro-fuzzy approach

In this approach a fuzzy model is represented as a neural network and typically the learning rules used in neurocomputing, like the classical backpropagation algorithm, are used to parameter estimation of the fuzzy model [17], [19], [40], [41], [42], [43], [44], [45], [46], [47], [48], [49]. Instead of neurocomputing, least mean squares [50] and evolutionary computation [51] have been used for parameter estimation.

Neuro-fuzzy formalims have been used for fault detection, classification, and prognosis. In [52], [53] a combination of the fuzzy min-max neural network and the classification and regression tree (CART) is proposed and applied to an electrical motor bearing faults detection. In [44], [54], [55] the application of fuzzy B-spline neural networks is studied. In [56] evidence theory is used for time domain feature fusion and a fuzzy neural network with possibility measures

[2] http://www.ieeexplore.ieee.org
[3] http://www.sciencedirect.com
[4] http://www.scopus.com

1063-6706 (c) 2018 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2018.2878200, IEEE Transactions on Fuzzy Systems

based inference is used in a multi-stage (sequential) fault classification process. The works [30], [31] are particularly relevant as are the first ones to take into account interpretability issues in the design of neuro-fuzzy system for bearing fault diagnosis. The fuzzy model can be of the Mamdani type, Sugeno type (Section IV-E) or relational type (Section IV-G). Further studies in the neuro-fuzzy category are included in the subcategories ANFIS, Fuzzy ARTMAP, and fuzzy lattice, presented below.

1) ANFIS: The *Adaptive Neuro Fuzzy Inference System* is characterized by a five layer network representing a Sugeno model, and a learning algorithm combining gradient descent and least mean squares. ANFIS has been used for fault detection, classification, and prognosis [4], [7], [22], [23], [35], [36], [52], [57], [58], [59], [60], [61], [62], [63], [64], [65], [66], [67], [68], [69], [70]. In [7], [59] multi ANFIS are trained and their results are combined by genetic algorithms. See [71] for a multiple output ANFIS application to fault classification.

2) Fuzzy ARTMAP: The fuzzy ARTMAP (FAM for short) is another type of fuzzy-network used for fault classification, where adaptive resonance theory (ART) is combined with fuzzy set theory [72]. ARTMAP is a neural network model for on-line incremental supervised pattern recognition but it can deal with binary inputs only. FAM extends ARTMAP by replacing the ART1 module with a fuzzy ART module where fuzzy set operations are used for modeling the module operations. This allows FAM to cope with both binary and real-valued inputs. Additionally, FAM employs a new learning rule that aims at maximizing predictive generalization. FAM has been used both as a direct fault classifier [8], [10], [37], [73], [74], [75] and as a base classifier in an ensemble [76], [77], [78], [79]. In [76] a homogeneous ensemble based on a simplified four-fields FAM is proposed (Section V-E5).

3) Fuzzy lattice neurocomputing: Originated from ART, fuzzy lattice neurocomputing [80] is a fundamentally different formalism when compared with the previous ones. Fuzzy lattice neurocomputing processes (d-dimensional) hyper-boxes in the input space and is able to process real-valued, interval, as well as incomplete (e.g., missing values), ambiguous (e.g., don't care values), and linguistic data (e.g., fuzzy sets). Furthermore, a learning algorithm such as Algorithm 1 requires only a single pass through the data set. The formalism can be also used for mining interpretable rules from data observations. However, some research is still required for attempting to keep the number of extracted rules to a minimum. Fuzzy lattice neurocomputing has found practical applications in different domains but in what concerns bearing diagnosis is only found in [38]. This is unfortunate given its just mentioned merits. For easy reference this type of neurocomputing is briefly surveyed next.

Fuzzy lattice neuro-classification can be thought as taking place in a three layers network where the input layer receives data, the middle layer computes an inclusion degree in the lattice hyper-boxes, and the decision layer selects the class that maximizes the inclusion degrees previously computed. Hyper-box processing is based on the notion of fuzzy lattice [38], [80]:

Definition 1. *A partially ordered set P is a set in which a binary relation $\leq$ is defined s.t. $\forall_{x,y,z\in P}$ the following properties hold: i) reflexivity, i.e., $x \leq x$; ii) anti-symmetry, i.e., if $x \leq y$ and $y \leq x$ then $x = y$; and iii) transitivity, i.e., if $x \leq y$ and $y \leq z$ then $x \leq z$.*

A lattice L is a partially ordered set s.t. $\forall x, y \in \mathsf{L}$ there is a greatest lower bound, hereafter denoted by $x \wedge y$ and a least upper bound, denoted by $x \vee y$. The dual of L, denoted as L^∂ is a lattice with the same underlying partially ordered set and with an ordering relation $\leq_\partial$ which is the converse of the ordering relation $\leq$ in L, i.e., $a \leq_\partial b$ in L^∂ iff $b \leq_\partial a$ in L.

Definition 2. *L is a complete lattice if L is a partial ordered set and $\forall_{X\subseteq\mathsf{L}}$, X has a minimum supremum denoted as $\vee X$ and a maximum infimum denoted as $\wedge X$.*

Definition 3. *A fuzzy lattice is a pair $\langle \mathsf{L}, \mu\rangle$ where L is a conventional lattice and $(\mathsf{L}\times\mathsf{L}, \mu)$ is a fuzzy set with membership function $\mu : \mathsf{L}\times\mathsf{L} \to [0, 1]$, s.t., $\mu(x, y) = 1$ iff $x \leq y$.*

Definition 4. *An inclusion measure is a real valued function $\sigma : \mathsf{L}\times\mathsf{L} \to [0, 1]$ s.t., $\forall x, y, u, w \in \mathsf{L}$, L being a complete lattice, the following properties are verified: i) $\sigma(x, 0) = 0, x \neq 0$; ii) $\sigma(x, x) = 1$; iii) Consistency, i.e., if $u \leq w$ then $\sigma(x, u) \leq \sigma(x, w)$; and iv) if $x \vee y < x$ then $\sigma(x, y) < 1$.*

Notice that an inclusion $\sigma(x, y)$ can be interpreted as the degree to which $x \leq y$. Also, it can be shown [80]:

Proposition 1. *If $\sigma : \mathsf{L}\times\mathsf{L} \to [0, 1]$ is an inclusion measure on lattice L then $\langle\mathsf{L}, \sigma\rangle$ is a fuzzy lattice.*

The inclusion measure used in [38] is $\sigma(x, y) = \frac{\mathrm{v}(y)}{\mathrm{v}(x\vee y)}$ where v is a valuation function, i.e., $\mathrm{v} : \mathsf{L} \to \mathbb{R}$ and $\forall x, y \in \mathsf{L}$, $\mathrm{v}(x \vee y) = \mathrm{v}(x) + \mathrm{v}(y) - \mathrm{v}(x \wedge y)$. Moreover, a positive valuation function is a valuation function for which if $x < y$ then $\mathrm{v}(x) < \mathrm{v}(y)$. Useful for the fuzzy lattice classifier specification is the observation that the Cartesian product of n lattices is itself a lattice, i.e., $\mathsf{L}^d = \mathsf{L}_1 \times \ldots \times \mathsf{L}_d$ and [80]:

Proposition 2. *If $\mathrm{v}_1, \ldots, \mathrm{v}_d$ are valuation functions on $\mathsf{L}_1, \ldots, \mathsf{L}_d$, respectively, then $\mathrm{v} = \mathrm{v}_1 + \ldots + \mathrm{v}_d$ is a valuation function on $\mathsf{L}^d = \mathsf{L}_1 \times \ldots \times \mathsf{L}_d$.*

Let $\mathsf{L}_l (l = 1, \ldots, d)$ be a totally-ordered lattice. The set of generalized intervals $\tau(\mathsf{L}_l) = \{[a, b] : a, b \in \mathsf{L}_l\}$ is also a lattice. The corresponding order relation $[a, b] \leq [c, d]$ in $\tau(\mathsf{L}_l)$ is equivalent to $c \leq a$ and $b \leq d$. Also, $[a, b] \wedge [c, d] = [a \vee c, b \wedge d]$ and $[a, b] \vee [c, d] = [a \wedge c, b \wedge d]$.

Proposition 3. *Let L_l be a totally-ordered lattice, $\mathrm{v} : \mathsf{L}_l \to \mathbb{R}$ a positive valuation function, and $\theta : \mathsf{L}_l^\partial \to \mathsf{L}_l$ an isomorphic function in L_l. Then $\forall_{[a,b]\in\tau(\mathsf{L}_l)}$ a positive valuation function $\mathrm{v} : \tau(\mathsf{L}_l) \to \mathbb{R}$ is given by $\mathrm{v}([a, b]) = \mathrm{v}(\theta(a)) + \mathrm{v}(b)$,*

Consider $\mathsf{L}^d = \mathsf{L}_1 \times \ldots \times \mathsf{L}_d$, $[\boldsymbol{a}, \boldsymbol{b}] \in \tau(\mathsf{L}^d)$ is called a hyper-box, $\boldsymbol{a}$ is its lower vertex, $\boldsymbol{b}$ its upper vertex, and its valuation function is

$$\begin{aligned}\mathrm{v}([a_1, b_1], \ldots, [a_d, b_d]) = &\,\mathrm{v}(\theta(a_1)) + \mathrm{v}(b_1) + \cdots \\ &+ \mathrm{v}(\theta(a_d)) + \mathrm{v}(b_d)\end{aligned} \quad (1)$$

1063-6706 (c) 2018 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2018.2878200, IEEE Transactions on Fuzzy Systems

An element $\boldsymbol{x} \in \mathsf{L}^d$ is contained in the hyper-box $[\boldsymbol{a}, \boldsymbol{b}] \in \tau(\mathsf{L}^d)$ iff

$$\boldsymbol{a} \leq \boldsymbol{x} \leq \boldsymbol{b} \Leftrightarrow a_1 \leq x_1, a_d \leq x_d, b_1 \geq x_1 \text{ and } b_d \geq x_d \quad (2)$$

In [38] only unit hyper-boxes are considered, i.e., hyper-boxes for which $\wedge \mathsf{L}_l = 0$ and $\vee \mathsf{L}_l = 1$ which allows to use positive valuation and isomorphic functions as simple as:

$$\mathrm{v}_l(x) = x; \;\; \theta_l(x) = 1 - x; \;\; x \in \mathsf{L}_l \quad (3)$$

To obtain unit hyper-boxes it is sufficient to normalize the data set features. The final required notion is the notion of size (or diagonal) of a hyper-box.

Definition 5. *Let $[\boldsymbol{a}, \boldsymbol{b}] \in \tau(\mathsf{L}^n)$ be a hyper-box. The diagonal of $[\boldsymbol{a}, \boldsymbol{b}]$ is a non-negative real function $diag_p : \tau(\mathsf{L}^d) \to \mathbb{R}_0^+$ given by the Minkowski metric $\mathsf{d}_p \in \mathsf{L}^d$, i.e.,*

$$\begin{aligned} diag_p([\boldsymbol{a}, \boldsymbol{b}]) &= \mathsf{d}_p((a_1, \ldots, a_d), (b_1, \ldots, b_d)) \\ &= [\mathsf{d}_1(a_1, b_1)^p + \cdots + \mathsf{d}_d(a_d, b_d)^p]^{1/p} \end{aligned}$$

with $p \in \{1, 2, \ldots\}$ and $\mathsf{d}_l(a_l, y_l) = \mathrm{v}_l(a_1 \vee b_l) - \mathrm{v}_l(a_1 \wedge b_l)$, $a_l, b_l \in \mathsf{L}_l, l = 1, \ldots d$.

After trained, this formalism classifies a hitherto unknown $\boldsymbol{x}$ by computing the inclusion measure (Def. 4) of $\boldsymbol{x}$ in all the families of sets $W^i (i = 1, \ldots, c)$ and by selecting the class corresponding to the family that maximizes the measure.

B. Fuzzy clustering

Clustering is a tool for exploratory data analysis aiming at segmenting a finite, unlabeled, multivariate data set into a set of homogeneous groups, categories or clusters. Alternatively, clustering can be seen as the process of identifying groups in data so that data in one group are similar to each other, and are as different as possible from data in other groups.

Clustering is the second most used approach in bearing fault diagnosis. Notably, the well-known fuzzy c-means (FCM) algorithm is the single most used formalism of all (17.3%): [9], [26], [27], [28], [81], [82], [83], [84], [85], [86], [87], [88], [89], [90], [91], [92], [93], [94], [95], [96], [97]. FCM aims at minimizing the objective function (4) for a specified number of cluster c and a given set of observations $\mathbf{X} = \{\boldsymbol{x}_j | j = 1, \ldots, n\}$ with $\boldsymbol{x}_j \in \mathbb{R}^d$,

$$J = \sum_{i=1}^{c} \sum_{j=1}^{n} \mu_{ij}^m ||\boldsymbol{x}_j - \boldsymbol{v}_i||^2 \quad (4)$$

under the constraints $\mu_{ij} \in [0, 1]$, $\sum_{j=1}^n \mu_{ij} > 0$, and $\sum_{i=1}^c \mu_{ij} = 1$, where μ_{ij} represents the membership of observation $\boldsymbol{x}_j (j = 1, \ldots, n)$ in the i-th cluster $(i = 1, \ldots, c)$, $\boldsymbol{v}_i$ refers to the centroid of the i-th cluster, $||.||$ stands for the Euclidean norm, $m > 1$ being the so-called fuzziness parameter. Increasing m increases the overlapping among the clusters. On the other hand, when $m \to 1$ FCM degenerates into k-means. Experimental evidence reported in the above studies clearly shows the effectiveness of FCM in bearing diagnosis, especially when contrasted with more classical methods such as k-means. Beyond FCM, specific variants of the algorithm have been considered. This is the case of [98] in which a specific cost-functional is proposed, and the works in [84], [99], [100] that use kernel-based FCM. A kernel performs a non-linear mapping from the feature space to a higher dimensional space where clusters can be better defined. In [101] a cluster shape free fuzzy-neighborhood density-based FN-DBSCAN clustering algorithm is applied to the bearing fault classification of an induction motor. In [102] a clustering method combining adaptive resonance theory (Section IV-A2) with a Yu's norm based similarity measure is proposed for bearing diagnosis. More recently, the FCM algorithm was extended with an observer point that, when manipulated, allows for the generation of partitions with different granularity in different regions of the feature space [13]. Moreover, in [103] the Geth-Geva clustering algorithm is used.

Fuzzy clustering has been used as an unsupervised fault classification tool, e.g., [13], [91], [103], [104], [105] (Section V-E7); to estimate data membership values in fuzzy support vector machines, e.g., [106]; for identifying the centers of

Algorithm 1: Fuzzy lattice fault classifier: learning.

Input : A finite (training) set $\mathbf{X} = \{\boldsymbol{x}_k | k = 1, \ldots, n\} \subset \mathsf{L}^d$ of normalized features. Associated with each feature vector $\boldsymbol{x}_k$ there is a label $y_k \in \{f_1, \ldots, f_c\}$ representing the fault type (or class) of the k-th input.
Output: c families of sets $W^1, \ldots, W^i, \ldots, W^c$, where the family W^i contains the set of all hyper-boxes induced for fault f_i.

for i = *1* **to** c **do**
  Let $X = \{\boldsymbol{x}_k | y_k == f_i; k = 1, \ldots, n\}$; $X' = \mathbf{X} \backslash X$;
  Let $\boldsymbol{w}_1 = [\boldsymbol{x}_1, \boldsymbol{x}_1]$; $W^i = \{\boldsymbol{w}_1\}$;
  for k = *2* **to** $|X|$ *(cardinality of X)* **do**
    if ($\boldsymbol{x}_k \notin \boldsymbol{w}_j$ *for all* $\boldsymbol{w}_j \in W^i; j = 1, \ldots, |W^i|$) **then**
      Dilate all hyper-boxes in W^i, i.e., $\boldsymbol{v}_j = \boldsymbol{w}_j \vee \boldsymbol{x}_k; j = 1, \ldots, |W^i|$;
      Let $V = \{\boldsymbol{v}_j | j = 1, \ldots, |W^i|\}$;
      Let F be the set of the hyper-boxes in V containing no elements in X' ;
      if ($F == \emptyset$) **then**
        Let $\boldsymbol{w}_k = [\boldsymbol{x}_k, \boldsymbol{x}_k]$;
        $W^i = W^i \cup \{\boldsymbol{w}_k\}$;
      end
      else
        if ($|F| == 1$) **then**
          Replace the original hyper-box $\boldsymbol{w}_j$ by its dilated version $\boldsymbol{v}_j$;
        end
        else
          Select the smallest hyper-box $\boldsymbol{w}_j$ and replace it by its dilated version $\boldsymbol{v}_j$;
        end
      end
    end
  end
end

1063-6706 (c) 2018 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2018.2878200, IEEE Transactions on Fuzzy Systems

radial basis functions neural networks acting as bearing fault classifiers [9], [107], and for identifying the antecedent of the rule in fault classifier probablistic fuzzy systems [108] (Section IV-N). In [90] FCM was used as the first step in a sparse component analysis procedure for feature extraction in a fault detection problem.

C. Fuzzy measures and the Sugeno integral

A fuzzy measure generalizes the notion of measure in the sense that the weaker property of monotonicity is considered instead of the usual additive property. Formally:

Definition 6. *Let X be a nonempty set, $\mathcal{P}$ the power set of X, and $g : \mathcal{P} \to [0, 1]$ be a set function. The set function g is a (normal) fuzzy measure iff i) $g(\emptyset) = 0$ ($g(X) = 1$) and ii) if $A \subset B \subset X$ implies that $g(A) \leq g(B)$.*

Definition 7. *Let $K \subset X$, a 2-additive fuzzy measure is defined as*

$$g(K) = \sum_{\{i,j\} \subseteq K} g_{ij} - (|K| - 2) \sum_{i \in K} g_i;\ |K| \geq 2 \tag{5}$$

This type of measures requires $k(k + 1)/2$ coefficients g_i, g_{ij} where $k = |K|$.

Fuzzy measures include plausibility, belief, possibility, necessity, and probability measures. Plausability and belief measures are basilar in the Dempster-Shafer evidence theory that can be viewed as a generalization of Bayesian subjective probabilities and offers a framework for combining evidence from different sources. Possibility theory has been used in a series of studies [24], [109], [110], [111], [112], [113], [114] especially in the scope of membership function tuning. Possibility measures resort to fuzzy sets for providing a qualitative framework for handling incomplete information. The numeric-to-linguistic conversion (otherwise known as *fuzzification*) of a feature, say x, can be accomplished by computing a possibility distribution of x relatively to a class of reference fuzzy sets representing the linguistic terms of a linguistic variable.

Definition 8. *The Sugeno g_λ-fuzzy measure is a fuzzy measure that satisfies the following additional property: $\forall_{A,B \subset X | A \cap B = \emptyset}, g_\lambda(A \cup B) = g_\lambda(A) + g_\lambda(B) + \lambda g_\lambda(A) g_\lambda(B)$ for some $\lambda > -1$.*

Based on this the Sugeno fuzzy integral can be given by:

Definition 9. *Let X be a finite fuzzy set with membership function $\mu : X \to [0, 1]$. The fuzzy integral of μ with respect to a fuzzy measure g_λ is defined by*

$$\begin{aligned} \int_X \mu(x) g_\lambda &= \max_{E \subseteq X} [\min(\min_{x \in E} \mu(x), g_\lambda(E))] \\ &= \max_{\alpha \in [0,1]} [\min(\alpha, g_\lambda(X_\alpha))] \end{aligned} \tag{6}$$

where $X_\alpha = \{x | \mu(x) \geq \alpha\}$.

The integral can be computed efficiently by pre-rearranging X so that $\mu(x_1) \geq \mu(x_2) \geq \ldots \geq \mu(x_n)$ and noting that $\int_X \mu(x) g_\lambda = \sum_{i=1}^{n} \min(\mu(x_i), g_\lambda(A_i))$ where $A_i = \{x_1, \ldots, x_i\}$ and

$$g_\lambda(A_i) = \begin{cases} \gamma^i + g_\lambda(A_{i-1}) + \lambda \gamma^i g_\lambda(A_{i-1}) & \text{for } i = 2, \ldots, n \\ g_\lambda(\{x_1\}) = \gamma^1 & \text{for } i = 1 \end{cases}$$

Moreover λ can be obtained by solving

$$\lambda + 1 = \prod_{i=1}^{n} (1 + \lambda \gamma^i);\ \lambda > -1 \text{ and } \lambda \neq 0$$

where $\gamma^i (i = 1, \ldots, n)$ is the so-called fuzzy density, a user-defined subjective importance value assigned to feature x_i. Fuzzy measures and the Sugeno integral can be used to assess the importance of features, and to represent synergetic, inhibitory, or non-existent interactions among them. These properties make these formalisms suitable for feature selection (Section V-B) and data fusion (Section V-C).

D. Fuzzy entropy

Fuzzy entropy can be used to measure the amount of fuzzy information in a fuzzy set and can be defined in several different ways. The following studies employ fuzzy entropy based features: [97], [115], [116], [117], [118], [119], [120], [121], [122], [123], [124], [125], [126], [127], [128], [129]. Experimental studies show that fuzzy entropy based features outperfom their classical counterparts in fault classification problems using SVM classifiers [117] (Section V-A).

E. Rule-based models

The fundamental motivation behind rule-based models is that they provide a convenient framework for expressing existing *a priori* expert knowledge on a problem. Two classical approaches are the Mamdani model and the Takagi-Sugeno-Khan model (or Sugeno model for short). In the former, both input and output variables are characterized by linguistic terms, as

$$r^{(i)} := \text{if } \boldsymbol{x} \text{ is } \boldsymbol{A}^{(i)} \text{ then } y^{(i)} \text{ is } B^{(i)} \tag{7}$$

where $r^{(i)}$ is the identifier of the i-th rule, $\boldsymbol{x} = [x_1, \cdots, x_n]'$ stands for the n real-valued inputs, $\boldsymbol{A}^{(i)} = [A^{(i)}, \cdots, A_n^{(i)}]'$ are the corresponding linguistic terms represented by membership functions; $\hat{y}^{(i)}$ is the rule output, and $B^{(i)}$ being the corresponding linguistic term.

In Sugeno models the consequent of each rule is a real-valued function of the inputs, f, typically a linear in the parameters function. This type of consequent has some advantages over a fuzzy consequent: i) it gives extrapolation ability to the model; ii) defuzzification is simplified; iii) the accuracy of the approximation tends to be improved by local consequents; and iv) optimal output parameters can be easily obtained by least squares techniques.

First generation fuzzy rule systems were designed by domain experts, using an expert-intensive time consuming design approach. Work along this line includes [20], [130], [131], [132]. In [20], [131], [132], [133], [134], [135], [136] Mamdani type of classifiers are considered with only a single output variable. The linguistic terms of this output variable

1063-6706 (c) 2018 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2018.2878200, IEEE Transactions on Fuzzy Systems

represent the several type of faults. In [137] a Mamdani type is proceeded by morphological operators and frequency domain based feature extraction for classification of incipient faults.

A second generation appeared aiming at proposing automatic and systematic design approaches with a minimum of human intervention. This second generation includes [138], [139], [140], [141]. In [138], [141] a decision tree is used for both feature selection and rule identification.

Rule-based models have been used essentially for fault classification. However, applications to fault detection are found in, e.g., [21], [142], [143].

F. Fuzzy support vector machines and fuzzy support vector data description

A fuzzy support vector machine (FSVM) [39] is a generalization of the well-known maximum margin classifier SVM. In this generalization each element of the data set is allowed to belong partially to each one of the existing classes. More concretely, the data set for a linearly separable binary classification problem is assumed to be defined as $\{(\boldsymbol{x}_i, y_i, \mu_i)|i = 1, \ldots, n\}$ where $\boldsymbol{x}_i \in \mathbb{R}^d$ is the i-th datum, $y_i \in \{-1, 1\}$ the corresponding label, and $\mu_i \in [0, 1]$ its membership value. Usually the membership values are computed from data observations $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$ by a clustering algorithm (see Section V-E4). The problem of estimating the decision boundary hyper-plane $\hat{y}(\boldsymbol{x}) = \boldsymbol{w}'\boldsymbol{x}+b$ ($\boldsymbol{w} \in \mathbb{R}^d; b \in \mathbb{R}$) yielding the max margin between the two classes is reformulated as the following quadratic optimization problem:

$$\begin{aligned} &\underset{\boldsymbol{w},b,\xi_i}{\text{minimize}} && ||\boldsymbol{w}||^2 + C\sum_{i=1}^{n}\mu_i\xi_i \\ &\text{subject to} && y_i(\boldsymbol{w}'\boldsymbol{x}_i + b) \geq 1 - \xi_i; \quad \xi_i \geq 0; \end{aligned} \tag{8}$$

which can be solved resorting to a Lagrangian and to a dual problem as in the original SVM. Similarly kernels can be also considered. Notice that i) C is a positive user-defined regularization parameter that controls how much the constraints influence the solution, i.e., the larger C the larger this influence; ii) ξ_i is the slack variable associated with the i-th training element that allows for margin violation or misclassification of $\boldsymbol{x}_i$, and iii) the membership value μ_i accounts for the contribution of i-th training element for the solution. For example if there is uncertainty on whether or not $\boldsymbol{x}_i$ is corrupted by noise, this could be reflected in a small μ_i that in turn makes the solution less dependent of this particular datum, independently of any eventual margin violation.

Related to the SVM formalism is the support vector data description and its fuzzy generalization (FSVDD) [144]. The problem here is, given a single class multivariate data set, estimate the minimum radius of a hyper-sphere that includes the data set. More concretely, consider the data set $\mathbf{X} = \{(\boldsymbol{x}_i, \mu_i)|i = 1, \ldots, n\}$ where where $\boldsymbol{x}_i \in \mathbb{R}^d$, and $\mu_i \in [0, 1]$ is a membership value. The problem of estimating the hyper-sphere with center $\boldsymbol{c} \in \mathbb{R}^d$ and minimum radius $r \in \mathbb{R}^+$ can be formulated as the following quadratic optimization problem:

$$\begin{aligned} &\underset{\boldsymbol{c},r,\xi_i}{\text{minimize}} && r^2 + C\sum_{i=1}^{n}\mu_i\xi_i \\ &\text{subject to} && ||\boldsymbol{x}_i - \boldsymbol{c}||^2 \leq r^2 + \xi_i; \quad \xi_i \geq 0; \end{aligned} \tag{9}$$

where as before, C is a positive user-defined regularization parameter controlling constraint influence, ξ_i, and μ_i are the slack variable and the membership function associated with $\boldsymbol{x}_i$, respectively. Being a quadratic optimization problem it can be solved using the optimization techniques used in the (F)SVM optimization problem above. Problem (9) can be further refined considering kernels, the center and radius as fuzzy sets, or all these. In [145] a bearing health degradation index based on FSVDD was proposed.

G. Fuzzy relation equations

A multi-input single output fuzzy relational model can be described by:

$$Y = X_1 \square X_2 \square \cdots \square X_k \square R \tag{10}$$

where $\square$ denotes s-t composition; $X_i : \mathbf{X} \rightarrow [0, 1] (i = 1, \ldots, k)$ and $Y : \mathbf{Y} \rightarrow [0, 1]$ represents the fuzzy descriptions (fuzzy sets) of the k inputs and the output, respectively; $\mathbf{X}$ and $\mathbf{Y}$ being the input and output domains, respectively. The relation matrix R being such that $R : \mathbf{X}^k \times \mathbf{Y} \rightarrow [0, 1]$.

Given the inputs X_i (linguistic description of features) and the output Y (linguistic descriptions of faults) it is possible to computed the relation matrix R algebraically, or resorting to numerical optimization techniques, including the learning rules of neural networks [146]. This type of formalism is an alternative inference formalism to fuzzy rule based systems and is used in [147], [148], [149], [150], [151] for fault classification.

H. Rough sets and fuzzy rough sets

Rough sets and fuzzy sets are two distinct formalisms that can be combined for taking advantage of their merits, i.e., indiscernibility and vagueness handling, respectively.

Informally, a rough set X is defined by a pair of *lower* and *upper* approximations, where the lower approximation is the conventional subset of elements that are certain to belong to X while the upper approximation is the conventional subset of elements that possibly belong to X. In the context of rough sets theory, an *information system*, $\mathcal{S}$, is defined as $\mathcal{S} = (\mathbb{U}, \mathbb{A}, V, f)$ where $\mathbb{U}$ is a non-empty finite set of objects (Universe of Discourse); $\mathbb{A}$ is a non-empty set of features (or attributes in Rough Set terminology), s.t., $\forall_{a\in\mathbb{A}} a : \mathbb{U} \rightarrow V_a$, V_a being the domain of attribute a, $V = \cup_{a\in\mathbb{A}} V_a$; and $f : \mathbb{U} \times \mathbb{A} \rightarrow V$ is the total decision (or information) function, s.t., for $\forall_{x\in\mathbb{U},a\in\mathbb{A}} f(x, a) \in V_a$. Consider a subset of attributes $B \subseteq \mathbb{A}$, then an indiscernibility relation, $Ind(B)$, can be defined as

$$Ind(B) = \{(x, y) \in \mathbb{U} \times \mathbb{U} : \forall_{b\in B} f(x, b) = f(y, b)\} \tag{11}$$

Objects x, y satisfying the relation $Ind(B)$ are indiscernible by attributes in B. An attribute $b \in B$ is said dispensable in B if $Ind(B) = Ind(B \backslash \{b\})$; it is indispensable in B otherwise.

Again in this context, a *decision table*, $\mathcal{D}$, is a special case of an information system, in the sense that it can be defined as the quadruple

$$\mathcal{D} = (\mathbb{U}, \mathbb{C} \cup \mathbb{D}, V, f) \tag{12}$$

1063-6706 (c) 2018 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2018.2878200, IEEE Transactions on Fuzzy Systems

where $\mathbb{C}$ and $\mathbb{D}$ are the sets of condition and decision attributes, respectively and $\mathbb{C} \cup \mathbb{D} = \mathbb{A}$. Table II shows an example of a decision table. In this example $Ind(B = \{a, b\}) = \{\{x_1\}, \{x_2\}, \{x_3, x_4\}\}$. Given (12) and $B \subseteq \mathbb{A}$, $X \subseteq \mathbb{U}$,

TABLE II: Example of a *decision table* $\mathcal{D} = (\mathbb{U}, \mathbb{C} \cup \mathbb{D}, V, f)$ with $\mathbb{U} = \{x_1, x_2, x_3, x_4\}$, $\mathbb{C} = \{a, b, c\}$, $\mathbb{D} = \{d\}$, $V = \{0, 1, 2\}$, and f given below. The task is to find the smallest subsets of $\mathbb{C}$ such that the reduced data are still consistent with $\mathbb{D}$.

	$\mathbb{C}$			$\mathbb{D}$
$\mathbb{U}$	a	b	c	d
x_1	0	1	2	0
x_2	1	0	1	1
x_3	2	2	1	0
x_4	0	0	2	2

the B-lower, and B-upper approximation of X are defined respectively by:

$$\underline{B}X = \bigcup \{Y \in Ind(B) : Y \subseteq X\} \tag{13}$$

$$\overline{B}X = \bigcup \{Y \in Ind(B) : Y \cap X \neq \emptyset\} \tag{14}$$

The set $\underline{B}X$ contains all the elements of $\mathbb{U}$ that are with certainty elements of X according to the attributes in B. The set $\overline{B}X$ contains the elements of $\mathbb{U}$ that possibly belong to X according to the attributes in B. From Table II with $B = \{a, b\}$, $X = \{x_1, x_3\}$ it comes $\underline{B}X = \{x_1\}$ and $\overline{B}X = \{x_1, x_3, x_4\}$. Given (12) and B, $B \subseteq \mathbb{C}$, the B-positive region $Pos_B(\mathbb{D})$ contains all the elements of $\mathbb{U}$ that can be correctly classified in the different classes defined by $Ind(\mathbb{D})$ using the attributes in B, and is given by

$$Pos_B(\mathbb{D}) = \bigcup_{X \in Ind(\mathbb{D})} \{\underline{B}X\} \tag{15}$$

For the example in Table II, $Pos_{B=\{a,b\}}(\mathbb{D}) = \{x_1, x_2\} \cup \{x_2\} \cup \emptyset = \{x_1, x_2\}$. Based on this a degree of dependency between attributes B and $\mathbb{D}$, can be defined as

$$\gamma_B(\mathbb{D}) = \frac{|Pos_B(\mathbb{D})|}{|\,\mathbb{U}\,|} \tag{16}$$

where $0 \leq \gamma_B(\mathbb{D}) \leq 1$ and can be interpreted as the degree to which $\mathbb{D}$ completely depends on the set of attributes B. If $\gamma_B(\mathbb{D}) = 1$ then all the attributes in $\mathbb{D}$ are uniquely defined by B. In our illustrative example, $\gamma_{B=\{a,b\}}(\mathbb{D}) = \frac{|\{x_1,x_2\}|}{|\{x_1,x_2,x_3,x_4\}|} = 0.5$.

In a decision system such as (12) a *reduct* is defined as a subset R of the conditional attribute set $\mathbb{C}$ s.t. $\gamma_R(\mathbb{D}) = \gamma_{\mathbb{C}}(\mathbb{D})$. The set of all reducts, Red, is $Red = \{R : R \subset \mathbb{C}$ and $\gamma_R(\mathbb{D}) = \gamma_{\mathbb{C}}(\mathbb{D})\}$ The mininum reduct set is $Red_{\text{min}} = \{R \in Red : \forall_{S \in Red} |R| \leq |S|\}$ For Table II, $Red_{\text{min}} = \{\{a, c\}, \{b, c\}\}$. Clearly, all the elements of $\mathbb{U}$ can be classified as in $\mathbb{D}$ using either the subset of attributes $\{a, c\}$ or $\{b, c\}$. Finding a minimal reduct is an NP-hard problem. In [12] a pruning algorithm based on an information degree that evaluates the significance of a given attribute is used for determining a reduct.

In bearing fault diagnosis features are real-valued. However, rough sets deal only with discrete values. To mitigate this, usually an application dependent discretization of features domain is performed. This type of approach was followed in [12], [67], [152], [153], [154]. An alternative approach is to resort to fuzzy rough sets [155]. In this case, $X \subseteq \mathbb{U}$ is viewed as a fuzzy set with membership function μ_X, both the conditional and the decision attributes can be also characterized by fuzzy sets. In addition, the indiscernibility relation (11) is defined as a fuzzy equivalence relation S with membership function μ_S and with the three basic properties: i) reflexivity, $\mu_S(x, x) = 1$, ii) symmetry, $\mu_s(x, y) = \mu_s(y, x)$, and iii) transitivity, $\mu_S(x, z) \geq \mu_S(x, y) \wedge \mu_S(y, z)$ where $\wedge$ is modeled by a t-norm. Under this framework, (13) and (14) for lower and upper approximations are now given in terms of membership functions, respectively

$$\mu_{\underline{B}X}(F) = \inf_{x \in \mathbb{U}} \max(1 - \mu_F(x), \mu_X(x)) \tag{17}$$

$$\mu_{\overline{B}X}(F) = \sup_{x \in \mathbb{U}} \min(\mu_F(x), \mu_X(x)) \tag{18}$$

where F is the fuzzy equivalence class for the objects similar to x, i.e., $\mu_F(x') = \mu_S(x, x')$. Eq. (16) still applies as long as $|.|$ is now viewed as cardinality of a fuzzy set, and $\mu_{Pos_B(\mathbb{D})} = \sup_X \mu_{\underline{B}X}$

I. Fuzzy numbers

Fuzzy numbers can be viewed as an uncertainty tolerant generalization of real numbers for modeling notions such as meant by the quotidian expression ‘about zero”. Fuzzy numbers are represented by fuzzy sets with normal, convex membership functions defined over a real domain, and form the basis of fuzzy arithmetic. Work in [17], [139] used fuzzy number in a fuzzy neural net for fault prognosis (Section V-F). In [156], [157] a model was developed representing the normal operating state of a spherical double-row roller bearing of type SKF 24124 CC/W33. The model takes as inputs the complex derivatives of acceleration signals. The time varying nature of the bearing state is captured by time-varying fuzzy number valued parameters. The model is used for fault classification and prognosis.

J. The Dempster-Shafer evidence theory

The Dempster-Shafer evidence theory is a general purpose framework for dealing with epistemic uncertainty, which can be viewed as a generalization of the Bayesian subjective probability theory. Evidence theory defines two dual non-additive fuzzy measures, belief and plausibility, based on which it combines evidence from different sources [158]. Given a set of propositions (or events) A, the associated probability $\Pr(A)$, is such that $\Pr(A) \in [Bel(A); Pl(A)]$, where the lower and upper limits of the interval are the belief and plausibility of A, respectively. $Bel(A)$ accounts for available evidence in favor of A, $Bel(A) \in [0, 1]$ with 0 meaning no evidence, and 1 certainty. On the other hand, $Pl(A)$ accounts for the extend to which the available evidence favoring the contrary (complement) of A still allows A to hold; $Pl(A) \in [0, 1]$. Formally:

1063-6706 (c) 2018 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2018.2878200, IEEE Transactions on Fuzzy Systems

Definition 10. *Let $\mathbb{U}$ be a finite universe of discourse and $\mathcal{P}$ its power set, a basic belief assignment (or mass) m, is a function $m : \mathcal{P} \to [0, 1]$, satisfying two properties: i) $m(\emptyset) = 0$ and ii) $\sum_{A \subseteq \mathcal{P}} m(A) = 1$.*

The mass $m(A)$ accounts for all the available evidence supporting A alone; not for any subset of A. Therefore,

$$Bel(A) = \sum_{B|B \subseteq A} m(B) \tag{19}$$

$$Pl(A) = \sum_{B|B \cap A \neq \emptyset} m(B) \tag{20}$$

hence $Pl(A) = 1 - Bel(\overline{A})$, with $\overline{A}$ being the complement of A. Given two masses m_1, and m_2 representing different sources of evidence, Dempster's rule combines these by computing the join mass $m_{1,2}$ as

$$m_{1,2}(A) = \frac{1}{1-k} \sum_{B \cap C = A \neq \emptyset} m_1(B) m_2(C) \tag{21}$$

where $k = \sum_{B \cap C = \emptyset} m_1(B) m_2(C)$ can be interpreted as a degree of conflict between the two sources, and $1 - k$ being a normalizing factor.

K. Fuzzy discriminant analysis

In fuzzy discriminant analysis, one starts with a data set where each element has a fuzzy multivariate attribute and want to determine a linear transformation of data into a lower dimensional space that yields the maximum separation between the existent fuzzy groups, i.e., that maximizes the ratio among between-group and within-group fuzzy variances. More concretely, consider the data set

$$\mathbf{X} = \{(\boldsymbol{x}_j, \mu_{ij}) | j = 1, \dots, n\} \tag{22}$$

where as before $\boldsymbol{x}_j \in \mathbb{R}^d$, $\mu_{ij} \in [0, 1]$ stands for the membership of the observation $\boldsymbol{x}_j$ in the i-th class ($i = 1, \dots c$), c being the number of classes. Typically, the membership values μ_{ij} are computed from data observations by a clustering algorithm (see Section V-E4). The problem of the orthogonal fuzzy neighborhood discriminant analysis is to find a matrix G that maximizes:

$$J(G) = \text{trace}(\frac{G' S_B G}{G' S_W G}) \tag{23}$$

with S_W and S_B being the within-class and the between-class scatter matrices, respectively. One way of specifying such matrices resorts to the mean values of each class, $\boldsymbol{v}_i$, and to the mean of data observations $\bar{\boldsymbol{x}}$, i.e.,

$$\boldsymbol{v}_i = \frac{\sum_{j=1}^{n} \mu_{ij} \boldsymbol{x}_j}{\sum_{j=1}^{n} \mu_{ij}} \tag{24}$$

$$\bar{\boldsymbol{x}} = \frac{\sum_{i=1}^{c} \sum_{j=1}^{n} \mu_{ij} \boldsymbol{x}_j}{\sum_{i=1}^{c} \sum_{j=1}^{n} \mu_{ij}} \tag{25}$$

Using these, S_W and S_B can be defined as:

$$S_W = \sum_{i=1}^{c} \sum_{j=1}^{n} \mu_{ij} (\boldsymbol{x}_j - \boldsymbol{v}_i)(\boldsymbol{x}_j - \boldsymbol{v}_i)' \tag{26}$$

and

$$S_B = \sum_{i=1}^{c} \sum_{j=1}^{n} \mu_{ij} (\boldsymbol{v}_i - \bar{\boldsymbol{x}})(\boldsymbol{v}_i - \bar{\boldsymbol{x}})' \tag{27}$$

L. Type-2 fuzzy sets

A type-2 fuzzy set considers that there is uncertainty in the (primary) membership function. A type-2 fuzzy set is then characterized by a primary and a secondary membership function where the latter defines a possibility distribution over the former. In this way, a type-2 fuzzy set generalizes the notion of fuzzy set as it tolerates uncertainty in the specification of the primary membership function. Inference systems based on type 2-fuzzy systems are able to accommodate higher levels of uncertainty and have been shown to outperform the corresponding type-1 systems [159]. Up to recently, and for efficiency reasons, only interval type-2 fuzzy systems have been considered in which the secondary membership function takes values in $\{0, 1\}$. This is the case of [159] where an interval type-2 fuzzy neural network is proposed and shown to outperform ANFIS in a bearing condition prediction task. Also, in [160] a type-2 based consensus is proposed to combine the results of an ensemble of SVM for fault classification (Section V-E5). With the recently proposed α-*planes* [161], and *zSlices* representation theorems [162], general type-2 fuzzy sets, in which the secondary membership function takes values in the unit interval, have become feasible for bearing fault diagnosis with clear advantages not only in terms of the performance but also in conceptual terms [163].

M. Fuzzy k-nearest neighbor

The fuzzy k-nearest neighbor (FKNN) algorithm [164] is a fuzzy extension of the simple, popular, and non-parametric k-nearest neighbor algorithm. The latter simply assigns a new previously unseen observation $\boldsymbol{x}$ to the most frequent class among the k-nearest existent labeled observations. In this algorithm k is a user-defined hyper-parameter and all the k neighbors are equally weighted in the assignment process. FKNN assigns a possibility distribution to $\boldsymbol{x}$ that depends both on the membership values of each of the k neighbors and on the distance of $\boldsymbol{x}$ to each one of these neighbors. More concretely, given (22) the membership value of $\boldsymbol{x}$ in the i-th class is:

$$\mu_i(\boldsymbol{x}) = \frac{\sum_{j=1}^{k} \mu_{ij} \mathrm{D}_j^{-2/(m-1)}}{\sum_{j=1}^{k} \mathrm{D}_j^{-2/(m-1)}} \tag{28}$$

where m is a user-defined hyper-parameter accounting for the effect of the distance D_j between the $\boldsymbol{x}$ and the j-th nearest observation.

N. Probabilistic fuzzy systems

A probabilistic fuzzy system [165] is composed by a set of rules combining linguistic information in the antecedents with probabilities in the consequents. Each rule can be viewed as describing a fuzzy region in the feature space where the

1063-6706 (c) 2018 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2018.2878200, IEEE Transactions on Fuzzy Systems

probabilities in the consequents are valid. More formally, the i-th rule, $r^{(i)}$ can be specified as:

$$\begin{aligned} r^{(i)} \quad := \quad & \text{if } x_1 \text{ is } A_1^{(i)} \text{ and } \ldots \text{ and } x_n \text{ is } A_n^{(i)} \\ & \text{then } \hat{y}^{(i)} = c_1 \text{ with } P(c_1|r^{(i)}), \cdots, \\ & \hat{y}^{(i)} = c_C \text{ with } P(c_C|r^{(i)}) \ [w^{(i)}] \end{aligned}$$

where $\boldsymbol{x} = [x_1, \cdots, x_n]'$ stands for the n real-valued inputs, $\boldsymbol{A}^{(i)} = [A^{(i)}, \cdots, A_n^{(i)}]'$ are the corresponding linguistic terms represented by membership functions; $\hat{y}^{(i)}$ is the rule output, $c_1, \ldots, c_C$ are the different faults (or classes) and $w^{(i)}$ is the certainty factor of the rule representing a belief in the accuracy of the rule. Notice that certainty factors can be associated with any type of (fuzzy) rules including those mentioned in Section IV-E and not only to probabilistic fuzzy rules. However, that practice was not found in bearing fault diagnosis.

Parameter estimation can be performed by either the two steps sequential conditional probability method, e.g., [166] or globally resorting to the maximization of the likelihood, e.g., [167].

V. A SYNTHESIS

As seen, a given formalism can be used in more than one activity of the fault diagnosis process and similarly a given activity can be based on more than one fuzzy formalism. In this section, the activity sequence: feature extraction, feature selection, fault detection, classification, and forecast (Fig. 1) is adopted for structuring the synthesis below. This includes the motivations underlying the used fuzzy formalism(s), a survey on how a formalism can be used for the current activity, and a discussion on the comparison with other applicable techniques, when available in the respective literature.

A. Feature generation

In feature generation (or extraction) the problem is: given a time domain signal (a time series) compute informative statistics and measures that characterize the signal. How to effectively address this problem is still a research issue specially due to the non-stationary non-Gaussian nature of bearing diagnosis signals.

Fuzzy versions of both approximate entropy and sample entropy can be used as time domain features of vibration signals. Both of these can be used as measures of a time series regularity and are computed by evaluating the probability of two time series becoming dissimilar when a new value is appended to them. A large entropy reveal a low time series regularity. Sample entropy can be viewed as an improved version of approximate entropy, therefore only the former will be reviewed here. Let the vibration signal be the time series $X = \{x(i)|i = 1, \ldots, n\}$. Fuzzy sample entropy $\mathcal{S}$ can be estimated by

$$\mathcal{S}(X; m, l, w) = \ln \phi^m(X; l, w) - \ln \phi^{m+1}(X; l, w) \quad (29)$$

where $\phi^m(X; l, w)$ is viewed as the probability of two subseries of X, with fix length m, be similar to each other. The similarity is characterized by a continuous convex membership function μ with parameters l and w, l being the location of the membership function while w is the width controlling the similarity tolerance. Moreover, the subseries of X, X_i^m $(i = 1, 2, \ldots, m)$, have the general form:

$$X_i^m = \{x(i), x(i+1), \ldots, x(n-m+i-1)\} \quad (30)$$

The probabilities being $(r = 0, 1)$:

$$\phi^{m+r}(X; l, w) = \frac{1}{n-m} \sum_{i=1}^{n-m} \Big(\frac{1}{n-m-1} \sum_{\substack{j=1 \\ j \neq i}}^{n-m} D_{ij}^{m+l}\Big)$$

where D_{ij}^m refers to the similarity of X_i^m, and X_j^m, i.e., typically $D_{ij}^m = \mu(d_{ij}^{(m)}; l, w) = \exp[-\ln_2(d_{ij}^{(m)}/w)^l]$ with $(i, j = 1, \ldots, n-m; i \neq j)$ and

$$d_{ij}^{(m)} = d(X_i^m, X_j^m)) = \max_{k \in \{0, \ldots, m-1\}} (|x(i+k) - x(j+k)|)$$

By resorting to a membership function, a similarity degree is obtained when comparing two subseries, rather than simply a 0 or 1 value as considered in the conventional sample entropy. Fuzzy sample entropy can be applied directly to the raw vibration signal. However, often it is applied to a coarse-grained version of the vibration signal with scale τ. This coarse-grained version, $X'_j(\tau)$, is obtained from the raw signal X by taking the arithmetic mean value of τ original neighboring points, i.e., $(1 \leq j \leq n/\tau)$

$$X'_j(\tau) = \{x'(j); x'(j) = \frac{1}{\tau} \sum_{i=1+(j-1)\tau}^{j\tau} x(i)\} \quad (31)$$

e.g., $X'_1(\tau = 2) = \{\frac{x(1)+x(2)}{2}, \frac{x(3)+x(4)}{2}, \ldots, \frac{x(N-1)+x(N)}{2}\}$. When fuzzy entropy is applied to a sequence of these coarse-grained versions it is called multi-scale fuzzy entropy. Naturally that other decomposition techniques for non-linear non-stationary signals can be used before applying fuzzy sample entropy. The work [119] applied fuzzy sample entropy to a local characteristic-scale decomposition of the original signal. Also, [97] presented two degradation indexes based on the entropy of time-frequency domain vibration signals. The study is particularly interesting as it shows how the proposed indexes evolve during run to failure tests. The concept of hierarchical fuzzy entropy is exploited in [123], [126]. This concept extends fuzzy sample entropy in the sense that it considers not only the lower frequency component produced by averaging, but it considers also a higher frequency component by taking the difference of two consecutive scales. Experimental results show that hierarchical fuzzy entropy outperforms multiscale fuzzy entropy over the Case Western Reserve University (CWRU) benchmark data sets [168].

The CWRU bearing data has become a benchmark for bearing fault diagnosis in general not only for fuzzy formalism based approaches. Data was collected from an apparatus consisting of a 2hp electric motor connected to a dynamometer through a shaft. The shaft is supported by deep groove ball bearings, a 6202-2RS JEM SKF at the fan end side and a 6205-2RS JEM SKF at the drive side. Vibration signals are acquired by accelerometers placed at 12 o'clock on the bearings housing at four successive rotation speeds: 1730, 1750, 1772, and 1797 rpm. Four health conditions were observed: 0,1778 single fault

1063-6706 (c) 2018 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2018.2878200, IEEE Transactions on Fuzzy Systems

in i) inner race, ii) outer race, iii) ball, and iv) no fault. For each of the above operating conditions, 20 data acquisition experiments were performed. Three data set are available: two for the drive end bearing (sampled at 48K samples/s) and one for the fan-end side (sampled at 12K samples/s). The latter comprises a vibration signal data set with 320 samples of 200 points each.

Another simple yet interesting concept for feature extraction is that of composite multiscale fuzzy entropy [129]. The idea is to use the average of all multi-scale fuzzy entropy for each scale factor τ. The resulting measure is more reliable than multi-scale fuzzy entropy especially for short time-series. Like any other feature, fuzzy entropy based features can be used in either fuzzy or non-fuzzy diagnosers. This type of features have been used as inputs of SVM, multi-layer perceptrons [115], [116], [117], ANFIS [119], and a predictive model based class discriminator [120] for fault classification. In general SVM has shown the best results, specially for small data sets. This last study is the only one to perform feature selection over the vector of multi-scale fuzzy entropy using the classic Laplacian Score.

B. Feature selection

The initial number of features in bearing fault diagnosis can be in the order of thousands [13], [154]. Some of them can be simply redundant, irrelevant or poorly informative. If these are not pruned, they will increase the computational burden and, what is worst, will decrease accuracy and increase overfitting. Informally, the feature selection (dimensionality reduction) problem is to select from the set of generated features those that are informative for characterizing the health state of a bearing; the remaining features being discarded with minimal information loss. Three types of formalisms have been used for feature selection: Fuzzy measures, (fuzzy) rough sets, and orthogonal fuzzy neighborhood discriminant analysis. These are surveyed below.

1) Fuzzy measures based feature selection: Let X be a set of features, the Shapley value of feature i, $i \in X$, denoted by v_i can be used to assess the importance of i with respect to a fuzzy measure g and is defined as

$$v_i = \sum_{k=0}^{n-1} \gamma_k \sum_{\substack{K \subset X \setminus i \\ |K|=k}} (g(K \cup \{i\}) - g(K)) \tag{32}$$

where $\gamma_k = \frac{k!(n-k-1)!}{n!}$ and $\sum_{j \in X} v_j = 1$. If $v_i > v_j$ then feature i is more informative than feature j. Another possibility for comparing features is to resort to the interaction index I_{ij} where

$$I_{ij} = \sum_{k=0}^{n-2} \xi_k \sum_{\substack{K \subset X \setminus \{i,j\} \\ |K|=k}} (g(K \cup \{i,j\}) - (g(K \cup \{i\}) - (g(K \cup \{j\}) + g(K))$$

with $\xi_k = \frac{k!(n-k-2)!}{(n-1)!}$. Accordingly, features $i, j \in X$ are i) redundant if $I_{ij} < 0$, ii) independent if $I_{ij} = 0$, and iii) complementary if $I_{ij} > 0$.

In [11] g is viewed as a 2-additive fuzzy measure (5) whose coefficients were estimated from data using FCM. The method is demonstrated in a SpectraQuest machinery fault simulator. Although interesting as it resorts only to fuzzy formalisms, this feature selection method does not scale well what can be a reason for the lack of follow up work so far.

2) (Fuzzy-)rough set based feature selection: Rough sets offer a framework for feature selection that requires no additional domain information besides that already available in the data set and no data transformation, thus preserving data semantics [155].

In rough set theory objects redundancy criteria as those presented in Section IV-H can be used to detect redundant features. In this case, data (features) are collectively viewed as a decision table in the sense of (12). In [12] it is reported that rough sets yields less features than the features selected by a decision tree for the same accuracy. In [152] rough set based feature selection for a neural network bearing fault diagnosis classifier was considered. The original rough set theory consider only discrete domains. When applied to real-valued features some quantization error can occur. To mitigate this effect, in [169] fuzzy rough sets were employed in a bearing fault classification problem.

3) Orthogonal fuzzy neighborhood discriminant analysis: Fuzzy discriminant analysis is adopted for feature reduction in [14], [15], [170]. In [14], [15] a variant of FCM is used to compute the required membership function degrees in (22). It was possible to reduce from 10 to 3 time-frequency domain features from stator current and lateral vibration measurements in a fault classification of WM1 ball bearing (6000 ZZ) of a brushless 1.2 kW, 4000 rpm, 50 Hz three phase BLDC motor. In [170] a similar technique is used to reduced the amount of discrete wavelet transform based features in fault classification of ball bearings under different corrosion defects.

C. Fuzzy fusion of data

Data fusion aims at integrating data from multiple sources for addressing issues such as data incompleteness, uncertainty, or high dimensionality. Data fusion can be considered at different levels, including at the signal level and at feature level. At signal level, data fusion combines signals from a set of different sensors. At the feature level, data fusion integrates features from different domains, e.g., time-domain features and frequency-domain features (Fig. 4). This is an alternative to stack different domain features into a compound vector that is treated as if features come all from the same domain. Besides that, fusion can be taken at the decision making level by finding a consensus between several independent diagnosers sharing the same or a partial view of the available input data. Such configuration is better known as an ensemble (Section V-E5). Two main type of formalisms have been used in this respect: The Dempster-Shafer evidence theory and the Sugeno integral (Section IV-J).

A fuzzy fusion module can be viewed also as a multi-input single output inference system such as those presented in Section IV-E where the inputs are the several criteria and the output is an index in $[0, 1]$ with, e.g., the healthiest condition

1063-6706 (c) 2018 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

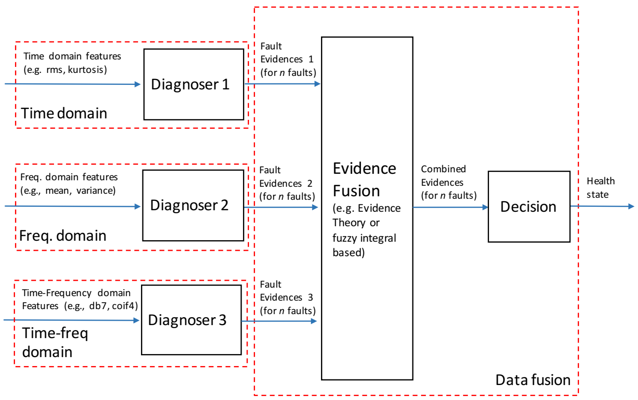

Fig. 4: A typical feature level data fusion scheme that can also be viewed a certain type of ensemble.

corresponding to 1 and a complete fault to 0. The employment of a fuzzy fusion module of multiple criteria was pursued in [149], [171].

1) Dempster-Shafer evidence theory based fusion: One practical issue in the application of the evidence theory to data fusion in bearing fault diagnosis is the determination of the basic belief assignments (mass) for applying the combination rule (21). In [172], a feature-level fusion was used where the universe of discourse propositions (Def. 10) are the inputs corresponding to each of the faulty conditions. A multi-layered perceptron per feature domain was used to estimate the mass provided by a given input. The obtained results were compared with the results obtained using Sugeno integral revealing that evidence theory outperforms the fuzzy integral approach both in simplicity (easier to tune) and performance. In [173], [174] information fusion using evidence theory is used. The work in [173] includes a variant for (21) which is applied to the fault classification of a single bearing using 3 accelerometers, each one of them installed in the X, Y, and Z directions of the bearing axis. In [172] an empiric comparison between evidence theory and fuzzy integral as formalisms for evidence fusion is provided revealing a clear advantage for the former.

2) Sugeno integral based fusion: The Sugeno integral (Section IV-C) can be used for combining n evidence sources x_i by letting $\gamma(x_i)$ in (6) be interpreted as the degree of certainty about a decision in the presence of x_i, and $g_\lambda(\{x_i\})$ as the degree of importance of the evidence towards the final evaluation. In this context, the Sugeno integral can be interpreted as the maximum grade of agreement between the objective evidence and the expectation. Sugeno integral has been used for combining the results of 3 diagnosers for a rotor-bearing system [172]. Each diagnoser was trained for its own feature domain.

D. Fault detection

Fault detection refers to the problem of identifying an incipient fault (no matter what) as early as possible. One of first fuzzy system based approaches to bearing fault diagnosis is due to [142] in which a Sugeno model is used to infer an unit interval-valued health degradation index (0 corresponding to normal health and 1 to failure) from four time-domain features (power condition indicator, standard deviation, power correlation factor, and standard deviation correlation factor). Experimental results using the CWRU data set have shown that the approach is sensitive to different types of faults (e.g., inner and outer races faults) and to different damage levels. In [61] a comparison study using three ANFIS with different type of input signals is performed. The main conclusion is that the ANFIS trained with vibration, current, and temperature signals performs only slightly better in terms of detection accuracy than the ANFIS trained with vibration signals only. The ANFIS trained with current signals only had the worst performance. In [21] a fault detection comparison between a Mamdani and Sugeno model is reported. The study uses vibration from the bearings of a heavy-duty milling machine and claims that no performance difference between the two models exists. Unfortunately details on the challenge of comparing an expert-based model with a data-driven based model or on how this comparison was performed are absent in this reference. In [143] a rule-based Sugeno model implements a proportional integral observer to fault detection of DC motor bearing.

When a bearing element rolls over a damaged surface, an impulsive force is produced that excites resonances in the mechanical system. However, the fault-induced impulsive features are buried in background noise and interference. Vibration demodulation is an effective technique to recover weak impulsive features from contaminated and modulated raw signals [5]. The technique comprises three main steps: i) determining the informative frequency band (IFB) around the resonance frequency; ii) applying a bandpass filter for IFB; and iii) generating a demodulated envelop spectrum that allows to identify the fault characteristic frequency and its harmonics, and thus the bearing health state. While steps ii) and iii) can be done using standard signal processing techniques, step i) is more challenging as IFB can depend on the type of fault and its severity. In [175], an IFB specific fuzzy clustering method

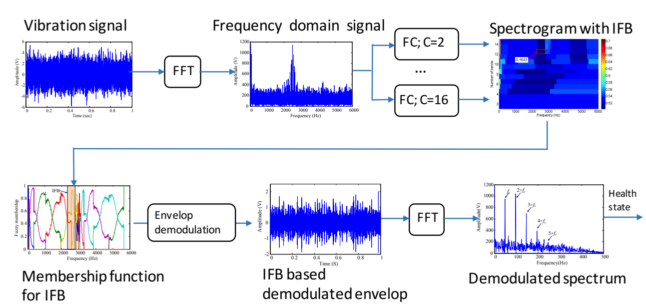

Fig. 5: The successive steps for bearing fault detection based on the informative frequency band (IFB) selection process proposed in [175]. FFT stands for Fast Fourier Transform, and FC; C=c refers to the IFB specific fuzzy clustering method with c clusters.

is proposed for the partitioning of the frequency spectrum into meaningful unequal bands. The algorithm is applied to the frequency spectrum of a vibration signal for a number of clusters $c = 2, 3, \ldots, c_{\max}$ where each cluster represents a frequency band. The band that, among all the partitioned

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2018.2878200, IEEE Transactions on Fuzzy Systems

bands, minimizes a proposed criterion (selector) is selected as IFB (Fig. 5). More concretely, as FCM performs poorly in this domain, the proposed clustering method replaces the Euclidean metric in (4) with the following metric:

$$\mathrm{d}_{ij} = \mathcal{N}(H_{ij}) + \mathcal{N}(S_{ij}) \tag{33}$$

where $\mathcal{N}$ is a normalizing function, H_{ij} measures the homogeneity between the i-th example and the j-th centroid, and S_{ij} is the selector cost that aggregates three frequency domain features (kurtosis, smoothness, and crest factor) for the i-th example and the j-th band. The method was evaluated using both simulated and experimental data and compared with the conventional techniques (kurtogram and hard partitioning) revealing superior performance.

E. Fault classification

From all the activities of fault diagnosis fault classification is by far the activity where fuzzy formalisms have been applied the most. These can be characterized as supervised, semi-supervised and unsupervised formalisms depending on the available information. Supervised formalisms assume that a labeled training set is available and the problem is a classifier design problem, i.e., to select and tune a classifier that using the available data set is able to accurately classify faults in test input samples specially for those not presented in the training set. Supervised formalisms include fuzzy rule based, fuzzy relation equations, neuro-fuzzy, fuzzy SVM, fuzzy K-NN, and probabilistic fuzzy systems.

1) Fuzzy rule based: The underlying assumption here is that an expert on a particular rotary machine will be able to detect a fault by sensing it, e.g., hearing or touching it. As described in Section IV-E this formalism allows for a heuristic and linguistic description of the human classification process. One possible fault classification rule would be

r := **if** input feature in region $\boldsymbol{A}$ of the feature space **then** health state is B

where $\boldsymbol{A}$ and B are fuzzy sets. The possibility of representing this type of vague or uncertain information in such a transparent way is an advantage relatively to non-fuzzy classification approaches. Another advantage is that no precise mathematical model of the system under diagnosis is required for designing the classifier. This early approach has been widely explored with acceptable results [20], [130], [131], [132], [133], [137], [138], [139], [140], [141], [176]. In [3] a single output Sugeno type of model was employed as roughness fault classifier. Features were extracted from vibration, current, and acoustic signals. In [177] both a Mamdani and a Sugeno type of fuzzy classifiers were tested revealing similar performance results.

2) Neuro-fuzzy based: The main motivation for using the neuro-fuzzy approach to fault classification is that it can build fuzzy rule based like models without resorting to the expensive process of eliciting rules from expert knowledge. Most of the early work on neuro-fuzzy was focused on improving the efficiency of the learning process, e.g., [40], [51]. Neuro-fuzzy based fault classification has been widely studied, e.g., [50], [67], [68], [69], [74]. These works include different type of signals (e.g., oil, temperature, current, vibration), different signal domains (time, frequency, time-frequency), type and severity of faults, and type of fuzzy neural net. Within this approach, the following studies are stressed for their originality and relevance. In [18] a comparison between a feedforward neural network, an Elman network, a radial basis function, and an ANFIS revealed that ANFIS is one of less accurate under the conditions of the study. The feedforward network was trained with 4 outputs, one for each fault, while ANFIS was trained using a simple output. This could be one reason for the poor performance of ANFIS in such study. In [50] a neuro-fuzzy classifier was shown to perform similarly to a neural-network under the same experimental conditions. However, the study stresses that neuro-fuzzy classifier codifies a set of interpretable rules that are not available in the neural-network. In [178] a multi-stage decision-tree like decision architecture for fault type classification and fault severity classification was proposed. Two instances of this architecture are analyzed. In one instance, a neuro-fuzzy network is used in each decision node while in the other instance, a conventional neural network is used. The reported results show that for the first decision levels (e.g., fault or no fault) no accuracy difference is observed. However, for deeper levels where a decision on fault severity is made (e.g., moderate or heavy fault) neuro-fuzzy appears to be advantageous. A similar approach, denoted as sequential conditional diagnosis, can be found in [56]. However, in the latter, feature fusion based on evidence theory (Section V-C1) is used together with possibility theory based inference.

a) Fuzzy lattice neurocomputing based: In [38], a simplified learning method for fuzzy lattice classifier (see Algorithm 1) was proposed and applied to the three CWRU benchmark data sets using time and frequency domain features. The simplified learning method does not require computing the values of the inclusion measure function during training. Roughly speaking, the fuzzy lattice classifier processes (d-dimensional) hyper-boxes in a normalized feature space. Typically, a hyper-box has a class label associated with it, say class c. If a feature space sample, $\boldsymbol{x}$, is inside the hyper-box, then $\boldsymbol{x}$ is from class c. Training consists in updating the vertexes of the hyper-boxes.

The empirical comparative analysis against conventional multi-layer perceptrons, SVM, and decision trees reports competitive results in terms of accuracy while it exhibits a simpler and significantly faster train (Section V-E2a). Also, no hyper-parameter tuning is required, except for the valuation and isomorphic functions.

3) Fuzzy k-nearest neighbor based: In [179] a multi-fault classifier based on vibration wavelets, selected using PCA (principal component analisys) and classified using the fuzzy k-nearest neighbor algorithm was proposed. Simplicity and efficiency are its main advantages. In [180] FKNN is applied to both gears and bearing fault classification of a gearbox. Several feature selection methods are considered and the main conclusion being that improving feature selection leads to an improvement of FKNN performance.

4) FSVM and FSVDD based: As SVM, fuzzy support vector machines are also grounded on the principles of structural risk minimization that lead to stable classifiers under

1063-6706 (c) 2018 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2018.2878200, IEEE Transactions on Fuzzy Systems

perturbation of the inputs, even in the presence of little and high dimensional training data [39], [181] (Section IV-F). This has been the main motivation for considering FSVM as a fault classifier [25], [182], [183], [184], [185].

Besides the bearing application considered, these studies range on the type of features used, on the feature selection method, and on the method used for estimation data membership values. This last issue is crucial in the design of an FSVM as its performance is particularly sensitive to it. Fuzzy clustering is often used for this purpose. In [182], [183], FSVM is used in a multi-fault classification problem. The setup includes vibration signal morphology analysis and random forest based feature selection. In [106] a weighted fuzzy clustering based FSVM design was proposed for fault classification. When compared with conventional SVM, FSVM revealed superior performance at the expenses of an extra design cost for membership grade tuning.

In [100] a multi-fault classifier based on a fuzzy support vector data description (Section IV-F) whose membership values are estimated by a kernel possibilistic c-means clustering was proposed. The proposed method outperformed both the standard SVDD and a sphere-based classifier, under equal experimental conditions.

5) Ensemble of fuzzy classifiers based: An ensemble combines the (base) classifications generated by a set of independent classifiers into a single final classification. Ensembles rely mainly on i) the diversity of the individual base classifications, and ii) on the consensus function that maps these individual classifications into a final one. Diversity can be generated in different ways, e.g., using different algorithms, different parametrization of the same algorithm, different data subsets, or different features (Fig. 4). The design of the consensus functions is arguably the most challenging issue in an ensemble setup as this function should be able to generate a better solution than the best base classification available in the ensemble. Fuzzy formalisms employed in this last stage of the ensemble were discussed in Section V-C. Ensembles of fuzzy classifiers for fault classification were studied in [76], [77], [78]. In [76] a homogeneous ensemble of 6 simplified FAM is proposed. The input features are equal to all the base classifiers. However, these were trained offline using randomized sampling of the inputs. Since the performance of the simplified FAM depends on the ordering sequence of inputs, the randomized sampling ensures the required performance diversity of the base classifiers in the ensemble. The ensemble output is computed using a weighted voting mechanism. In [77], [78], [79] ensembles of FAM classifiers are reported. More concretely, FAM classifiers are used as Diagnosers in a setup similar to Fig. 4. In [77], [78] each FAM is fed with its own set of domain specific features, i.e., time domain, frequency domain, wavelet grey moment, wavelet energy spectral, and auto-regression coefficients. The fusion method is based on the Bayesian belief method (averaging-estimated posterior probabilities). Results using the CWRU data set show that the proposed ensemble outperforms an individual FAM. In [79] a correlation degree between the individual FAM classifiers is used to determine the final classification decision.

In [160] a type-2 fuzzy set inference based consensus is proposed for an ensemble of three SVM. The reported classification results show that the proposed ensemble outperforms a standalone SVM, and the latter outperforms ANFIS for the considered experimental conditions. A similar consensus mapping using general type-2 fuzzy set inference is proposed in [163].

6) Probabilistic fuzzy systems based: In rule based bearing fault diagnosis typically Mamdami or Sugeno models are used. Considering that i) the application requires a multi-input multi-output (MIMO) model with typically one output for each fault, ii) that the mentioned models do not scale sufficiently well for the MIMO case, iii) that the larger the model the more difficult is to interpret it, and iv) the more prone it becomes to overfitting, in [108] a different approach is proposed. In [108] a parsimonious type of fuzzy rule based model is considered where each rule can diagnosis a set of faults each one with an associated probability, i.e., a probabilistic fuzzy system is considered (Section IV-N). Results show that it is possible to obtain a trade-off between the expressiveness, interpretability and accuracy of the probabilistic fuzzy system. Statistical tests of hypotheses show that it is possible to obtain parsimonious probabilistic fuzzy systems with indistinguishable accuracy when compared to other data-driven models including neural networks, and SVM.

7) Unsupervised formalisms: It is clear that with enough labeled data supervised learning methods are more effective in terms of accuracy. However, when the available data for certain types of faults are scarce or when the type of fault is unknown, unsupervised approaches such as clustering become relevant. Another reason for using clustering is for membership functions elicitation of linguistic variables. This is a common approach to automatically tune fuzzy rule-based, and fuzzy relational models. In [99] feature weights are incorporated into the FCM formulation and used to fault classification. A weight is associated with the feature relevance which is estimated by a class separability index. In [98] a specific cost-functional based partitioning clustering algorithm is derived for bispectra of vibration signals; an alternative to Fourier analysis. Moreover, in [104] the coefficients of an autoregressive model estimated from higher-order cumulants are used as features and analyzed by a fuzzy equivalence based clustering algorithm to identify the different types of faults and their stages. Using also a signal processing approach, in [83] FCM is applied to the first, second, and third order amplitudes of the characteristic frequencies for the outer and inner faults extracted from time-delayed correlation demodulation of the vibration signals. Also, in [85], [86] fault classification is accomplished by applying fuzzy clustering to binary images obtained from high-order cumulants bispectra of the vibration signals. In bispectrum analysis each fault category has associated a unique bi-dimensional pattern. In [84] kernel FCM is used for fault classification. In [27] FCM and Mahalanobis distance are used to diagnose the real breaking of a thrust bearing in an industrial multishaft centrifugal compressor located in an air separation unit. In [89] FCM classifies faults based on time-frequency features and on the ensemble local mean decomposition (ELMD) method. Similar studies are found in [88], [92],

1063-6706 (c) 2018 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2018.2878200, IEEE Transactions on Fuzzy Systems

[95], [103], [105]. In [94] the Kullback-Leibler divergence is used in FCM as a metric of acoustic signals for bearing fault classification. In [102] an ART and Yu's norm based clustering method outperforms in bearing fault classification accuracy both fuzzy ART and Self-Organizing maps. In [96] an iterative procedure is proposed for the automatic determination of the type of faults in fault classification using time domain features. This is viewed as the equivalent to identify the number of clusters in FCM. In [13] the FCMFP algorithm [186] is applied to fault classification of bearings. In FCMFP a regularization term based on a focal point $\mathbf{P} \in \mathbb{R}^d$ representing the position of an observer is embodied into the cost functional of FCM (4) as follows:

$$J = \sum_{i=1}^{c} \sum_{j=1}^{n} u_{ij}^{m} ||\boldsymbol{x}_j - \boldsymbol{v}_i||^2 + \zeta \sum_{i=1}^{c} ||\mathbf{P} - \boldsymbol{v}_i||^2 \qquad (34)$$

The regularization coefficient ζ allows one to adjust between the unbiased algorithm (FCM) for $\zeta = 0$ and a biased one for is $\zeta > 0$. Also, no restrictions apply to the position of $\mathbf{P}$ [186]. By changing these two hyper-parameters, different regions of the feature space are analyzed with different levels of detail.

8) Semi-supervised methods: When health states are unknown *a priori* but there are still some historical data available for training, resorting to semi-supervised methods is one possible approach [176], [187]. An example is when large amount of data from the healthy state are available but very few data exist (if any) for certain faulty states. In [176] FCM is first used to identify candidates to centers of the regions in the feature space that represent different healthy states. In a subsequent step, the candidates are evaluated against known data. In [187] a new semi-supervised weighted kernel clustering algorithm is proposed. Using the obtained clustering results, the semi-supervised classification of labeled and unlabeled data is formulated as the following optimization problem:

$$\underset{\boldsymbol{w}_l, \boldsymbol{w}_u, \boldsymbol{v}, \boldsymbol{w}, \sigma}{\text{minimize}} \quad \boldsymbol{w}_l \mathrm{FCR}(\boldsymbol{v}, \boldsymbol{w}, \sigma) + \boldsymbol{w}_u \mathrm{XBK}(\boldsymbol{v}, \boldsymbol{w}, \sigma) \qquad (35)$$

where $\boldsymbol{w}_l$, and $\boldsymbol{w}_u$ are weight vectors corresponding to labeled and unlabeled observations, $\boldsymbol{v}$ stands for the set of all cluster centroids, $\boldsymbol{w}$ and σ are the hyper-parameters of the kernel function. Moreover, FCR is the false classification rate for the clustering results over the labeled observations, and XBK is a kernel extension of the well-known Xie-Beni index for internal validation of clustering results measured over unlabeled observations. Solution instances of (35) are obtained by an evolutionary optimization method. The experimental results show a clear superior accuracy relatively to FCM and kernel FCM, and comparable accuracy relatively to SVM and radial-basis function neural networks for labeled data while the semi-supervised method is also able to deal effectively with both labeled and unlabeled data.

F. Fault prognosis

In fault prognosis the problem is to assess the bearing health degradation (or damage) over time and predict its remaining useful life (RUL). The health degradation of a bearing is a continuous irreversible process. Once a new bearing is first installed it should enjoy a long-term healthy working period. Eventually, minor incipient faults start to appear that gradually at first and then acceleratedly grow with operation time leading to a complete failure. Therefore, the bearing life cycle can be viewed as a transition between three main stages: normal, degeneration and failure. The prediction of RUL can

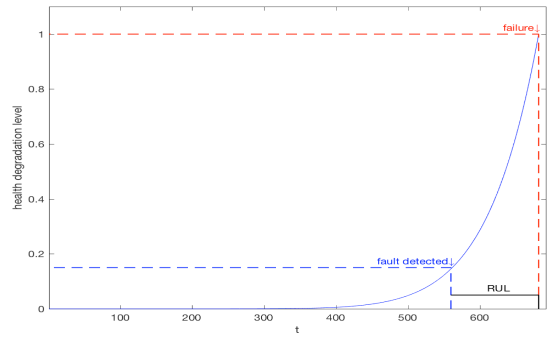

Fig. 6: The remaining useful life (RUL) as the difference between the time of fault detection and the time of failure, as estimated by a health degradation (or fault severity) index represented by the solid curve.

be accomplished using a health degradation (or fault severity) index by taking the difference between the fault detection time and the failure time as predicted by the index (Fig. 6). Notice that optimistic (pessimistic) health assessments lead to late (early) failure predictions. One of the first works to adopt fuzzy formalisms for fault prognosis is [17] where two fuzzy neural networks trained with backpropagation were used. One network estimates the current health condition of the bearing while the second network estimates the RUL using historical data. The reported results were promising but the size of the data set used was rather small. In [43] a neuro-fuzzy predictor of bearing fault trend was shown to outperfom the prediction accuracy of a radial-basis-function neural network. Similarly, in [47] an one step ahead neuro-fuzzy predictor trained with particle filtering was shown to predict better than conventional gradient descent trained recurrent neural-networks and recurrent neuro-fuzzy network [49], [65]. A case-based data driven prognosis framework resorting to ANFIS as cases models is proposed in [70]. Results show that the proposed framework outperforms conventional ANFIS based predictors. In [188] a membership function characterization for each element of a finite set of fault severity levels is proposed. The membership functions are estimated using frequency domain features. A similar membership function characterization is found in [189]. In [145], an FSVDD based bearing health monotone degradation index, d, is proposed. Time domain features are collected as a training set while the bearing is operating in normal conditions. Afterwards, whenever a testing sample, characterized by the feature vector $\boldsymbol{x}(t)$ is present at time t its monitoring coefficient, $\bar{\epsilon}(t)$, is assessed against the hyper-sphere center $\boldsymbol{c}$, and radius $\bar{r}$, previously identified, $\bar{\epsilon}(t) = \frac{r(t)-\bar{r}}{\bar{r}}$ where $r(t) = ||\boldsymbol{x}(t) - \boldsymbol{c}||$. Notice that if $\bar{\epsilon}(t) > 0$ then $\boldsymbol{x}(t)$ is considered an outlier, i.e., the bearing is no longer in a normal condition. The monotone index $\mathrm{d}(t)$ is then $\mathrm{d}(t) = \max(\frac{t}{t-1}\,\mathrm{d}(t-1), \bar{\epsilon}(t));\ t > 1$ with initial conditions

1063-6706 (c) 2018 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2018.2878200, IEEE Transactions on Fuzzy Systems

$\mathrm{d}(t \leq 1) = 0$.

Another approach to bearing health forecasting is based on a fuzzy-grey model for short term fault prediction [190]. In brief, GM(n, m) denotes a grey model of order n and m variables and it can be used to predict the H steps ahead values of a partially unknown system (charaterized by a time series) using a relatively small amount of the system historical data [191]. In [190] the popular GM(1,1) is considered thus only this model is reviewed. Consider a non-negative valued primitive time-series $X^{(0)} = (x^{(0)}(1), \ldots, x^{(0)}(k), \ldots, x^{(0)}(n))$; $n \geq 4$. For smoothing, the Accumulating Generation Operator (AGO) is applied to the time series, yielding:

$$\begin{aligned} X^{(1)} &= \mathrm{AGO}(X^{(0)}); \\ &= (x^{(1)}(1), \ldots, x^{(1)}(k), \ldots, x^{(1)}(n)) \end{aligned} \quad (36)$$

where $x^{(1)}(k) = \sum_{i=1}^{k} x^{(0)}(i)$, $k = 1, 2, \ldots n$ The GM(1,1) corresponding differential equation is

$$\frac{dx^{(1)}(t)}{dt} + ax^{(1)}(t) = b; \quad (37)$$

where $a \neq 0$ and b are updated by least square estimation each time a new value for the primitive series $X^{(0)}$ is available. Using the Inverse Accumulating Generation Operator a solution for the H steps ahead predicted value, $\hat{x}^{(0)}(k+H)$, can be obtained from (37):

$$\hat{x}^{(0)}(k+H) = (x^{(0)}(1) - \frac{b}{a})(1 - \mathrm{e}^{a})\mathrm{e}^{-a(k+H-1)} \quad (38)$$

As the horizon H increases the quality of the prediction descreases. To account for this, in [190] each predictive value has an associated membership degree. The resulting model is used to predict both time and frequency feature values, and thus faults, of an oil-line pump 3G2226E bearing.

G. Discussion

The above synthesis shows that fuzzy formalisms have been applied in each one of the activities of the fault diagnosis process, from feature extraction to fault prognosis (Table III). Fault classification is the activity where a greater number of formalisms have been studied. Neuro-fuzzy networks and fuzzy clustering are the two fuzzy formalisms with the larger range of applications, i.e., fault detection, classification and prognosis. Under fuzzy clustering, an interesting number of clustering algorithms has been studied. However it is not clear which one is more suitable (if any) for each type of activity where they were applied. The reviewed works allow us to conclude that for fault detection and fault classification fuzzy formalisms exhibit a comparable or superior performance relatively to other non-fuzzy data driven formalisms such as SVM or neural-networks. However fuzzy formalisms have the potential for providing an insight on the relationships among the model variables. Some of the revised literature presents a comparative analysis between a fuzzy formalism and the corresponding non-fuzzy technique (e.g., Fuzzy SVM versus SVM) where, in general, the fuzzy extension outperforms the conventional method.

So far fuzzy entropy based features have been used only as standalone features, i.e., no other time-domain, frequency domain, or time-frequencies domain features have been used together with them. The claim in the reviewed studies is that fuzzy entropy is sufficiently informative to identify the faults and their severity in the considered experimental data, i.e., CWRU data [168]. However, other studies using more demanding data and comparing with other type of features rather than only basic time-domain features such as root mean squares, skewness, kurtosis, and peak value may indicate otherwise.

In data fusion, evidence theory is preferred relatively to the Sugeno integral. However, it is not clear how evidence theory compares to other methods of data fusion. The same applies to feature extraction. In this activity, multi-scale fuzzy sample entropy outperforms multi-scale sample entropy. However, no comparison exists with other feature extraction methods. In general, there is also a shortage of comparisons between fuzzy and non-fuzzy feature selection methods. Currently there is no study showing which of the three formalisms, i.e., fuzzy measure based, fuzzy-rough sets based, or orthogonal fuzzy neighborhood discriminant analysis, allows for a greater reduction of dimensionality, or is more robust to outliers, if any. Also, there are no studies comparing the feature selection formalism with conventional formalisms.

In general, it would be useful to have a comparison on the performance of the several methods available for each activity of the fault diagnosis process. Unfortunately, this type of comparison cannot be performed from the reviewed literature. The variety of experimental setups, metrics, and evaluation methods used prevents one from producing such comparison from the literature. Thus an immediate follow up research question resulting from our work is: in which working conditions a given formalism is more suitable than other for a given fault diagnosis activity? In an attempt to start answering this question, recently the authors performed an empirical comparison under controlled conditions of fuzzy clustering algorithms for unsupervised fault classification. Both the benchmark CWRU data and a more realistic setup were used revealing that for the usual number of selected features (dozen) FCMFP outperforms FCM, GK, and fuzzy-DBSCAN [192], [193].

VI. A PROMISING RESEARCH DIRECTION

Besides the literature gaps just discussed that are worth pursuing, we elect multi-objective optimization of fuzzy models as a promising research direction. One of the main advantages of fuzzy models is that when compared with other nonlinear modeling and detection techniques such as artificial neural networks, they have the important advantage of providing an insight on the linguistic relationship between variables [29].

A promising approach is to consider the explanation capabilities (interpretability) of the diagnosis model as an explicit design goal within a systematic synthesis process, and to perform a comprehensive evaluation on the impact of the interpretability on other performance criteria. This approach aims at answering the following questions: i) how can we endow a model with explanatory capabilities using a systematic approach? and ii) what is the impact of this objective on commonly used diagnosis performance criteria?

1063-6706 (c) 2018 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2018.2878200, IEEE Transactions on Fuzzy Systems

TABLE III: Applied fuzzy formalisms in each one of the activities of the bearing fault diagnosis process.

Feat. extraction	Feature selection	Data fusion	Fault detection	Fault Classification	Fault prognosis
Fuzzy entropy	Fuzzy measures	Evidence theory	Rule-based	Rule-based	Fuzzy numbers
	Fuzzy-rough sets	Sugeno integral	Neuro-fuzzy	Neuro-fuzzy	Neuro-fuzzy
	Orthogonal fuzzy neighborhood discriminant analysis		Fuzzy clustering	Fuzzy clustering	Fuzzy clustering
				Fuzzy relation equations	Fuzzy-grey models
				Fuzzy SVM	
				Fuzzy SVDD	
				Ensembles	
				Fuzzy K-NN	
				Probabilistic Fuzzy Sys	
				Type-2 fuzzy sets	

The multi-objective and non-linear nature of the diagnosis problem makes it well suited to be tackled by a multi-objective optimization framework. This framework provides the possibility to choose from a working set of many potential alternative non-dominated solutions exhibiting different balances between interpretability and accuracy. In this sense, this approach can be viewed as a generalization of existing design methodologies (e.g., those described in Section IV-A), since it includes them as a special case when interpretability is neglected. The multi-objective optimization approach has been applied in other domains with quite interesting results, e.g. [194]. However, no work was found in our search that employed the multi-objection approach to bearing fault diagnosis. A single objective optimization approach that takes into account interpretability and accuracy in this domain is found in [30], [31] though.

VII. Conclusions

A bearing is a fundamental mechanical component and one of the primary cause of failure in rotary machines. Given the omnipresence of such machines (e.g., generators, turbines) early detection of bearing faults is not only a technically challenging problem but also an economically relevant one.

A first systematic review of works dealing with fuzzy formalisms for bearing fault diagnosis was presented. More than 150 papers were considered. Most of these were published in more than 50 different journals; Conferences papers were also considered.

By fault diagnosis we mean fault detection, fault classification, or fault prognosis. Sixteen formalisms have been identified and summarized. The two more notable formalisms are the neuro-fuzzy approach and clustering. Curiously enough, individually the fuzzy c-means algorithm is the most used algorithm among all algorithms employing fuzzy formalisms.

Different fuzzy formalisms have been used in the different steps of the bearing fault diagnosis process and found a wide range of real-world applications. Most of the existing work uses fuzzy formalisms for fault classification purposes. However, fuzzy entropy is often used for feature extraction. Fuzzy measures and (fuzzy) rough sets are being used for feature selection. Evidence theory is used as an information fusion framework.

The presented review provides a sound background for new studies on this subject. It identifies some literature gaps such as the lack of systematic performance comparisons among the available formalisms for a given bearing fault diagnosis activity and proposes a promising future research direction: multi-objective optimization of fault identification accuracy and interpretability.

Acknowledgements

The work was funded in part by the Prometeo Project of SENESCYT, Ecuador, the National Key Research & Development Program, China (2016YFE0132200), and by FCT, Portugal (SFRH/BSAB/128153/2016).

References

[1] "Report of large motor reliability survey of ind. and commercial installations, part I," *IEEE Trans. Ind. Appl.*, vol. IA-21, no. 4, pp. 853–864, July 1985.

[2] "Report of large motor reliability survey of industrial and commercial installations, part II," *IEEE Trans. Ind. Appl.*, vol. IA-21, no. 4, pp. 865–872, July 1985.

[3] L. Navarro, M. Delgado, J. Urresty, J. Cusido, and L. Romeral, "Condition monitoring system for characterization of electric motor ball bearings with distributed fault using fuzzy inference tools," in *Instru. and Measure. Tech. Conf. (I2MTC), 2010 IEEE*, May 2010, pp. 1159–1163.

[4] A. Oskouei and M. Esmaeili, "Neuron-fuzzy-based acoustic emission output parameters prediction for rotary device monitoring," in *ASME Int. Mech. Eng. Congr. and Exposition, Proc. (IMECE)*, vol. 12, 2012, pp. 437–442.

[5] R. B. Randall and J. Antoni, "Rolling element bearing diagnostics - a tutorial," *Mech. Syst. and Signal Process.*, vol. 25, no. 2, pp. 485 – 520, 2011.

[6] H. Kumar, P. Srinivasa, G. Vijay, and R. Rao, "Wavelet transform for bearing condition monitoring and fault diagnosis: A review," *Int. J. of Condition Monitoring and Diagnostic Eng. Management*, vol. 17, no. 1, pp. 9–23, 2014.

[7] Y. Lei, Z. He, Y. Zi, and H. Q., "Fault diagnosis of rotating machinery based on multiple ANFIS combination with gas," *Mech. Syst. and Signal Process.*, vol. 21, no. 5, pp. 2280–2294, 2007.

[8] Z. Xu, J. Xuan, T. Shi, B. Wu, and Y. Hu, "A novel fault diagnosis method using PCA and art-similarity classifier based on Yu's norm," *Key Eng. Materials*, vol. 413-414, pp. 569–574, 2009.

[9] G. Vijay, S. Pai, N. Sriram, and R. Rao, "Radial basis function neural network based comparison of dimensionality reduction techniques for effective bearing diagnostics," *Proc. Institution of Mech. Engineers, Part J: J. of Eng. Tribology*, vol. 227, no. 6, pp. 640–653, 2013.

[10] J. Ben Ali, L. Saidi, A. Mouelhi, B. Chebel-Morello, and F. Fnaiech, "Linear feature selection and classification using pnn and sfam neural networks for a nearly online diagnosis of bearing naturally progressing degradations," *Eng. Applicat. of Art. Intell.*, vol. 42, pp. 67–81, 2015.

[11] X. Liu, L. Ma, S. Zhang, and J. Mathew, "Feature group optimisation for machinery fault diagnosis based on fuzzy measures," *Australian J. of Mech. Eng.*, vol. 5, no. 2, pp. 191–197, 2008.

[12] X. Zhao, C. Zhao, H. Gao, and G. Wu, "Knowledge mining for fault diagnosis based on rough sets theory," in *IEEE Int. Conf. on Fuzzy Syst.*, June 2008, pp. 744–749.

[13] C. Li, J. V. de Oliveira, M. Cerrada, F. Pacheco, D. Cabrera, V. Sanchez, and G. Zurita, "Observer-biased bearing condition monitoring: From fault detection to multi-fault classification," *Eng. Applicat. of Art. Intell.*, vol. 50, pp. 287 – 301, 2016.

1063-6706 (c) 2018 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2018.2878200, IEEE Transactions on Fuzzy Systems

[14] W. Abed, S. Sharma, and R. Sutton, "Diagnosis of bearing fault of brushless dc motor based on dynamic neural network and orthogonal fuzzy neighborhood discriminant analysis," in *Control, 2014 UKACC Int. Conf. on*, 2014, pp. 378–383.

[15] W. Abed, S. Sharma, R. Sutton, and A. Motwani, "A robust bearing fault detection and diagnosis technique for brushless dc motors under non-stationary operating conditions," *J. of Control, Automation and Elect. Syst.*, vol. 26, no. 3, pp. 241–254, 2015.

[16] P. Goode and M.-Y. Chow, "A hybrid fuzzy/neural system used to extract heuristic knowledge from a fault detection problem," in *Fuzzy Syst., Proc. Third IEEE Conf. on*, Jun 1994, pp. 1731–1736 vol.3.

[17] B. Satish and N. Sarma, "A fuzzy BP approach for diagnosis and prognosis of bearing faults in induction motors," in *2005 IEEE Power Eng. Society General Meeting*, vol. 3, 2005, pp. 2291–2294.

[18] A. Mahamad and T. Hiyama, "Fault classification performance of induction motor bearing using ai methods," in *Ind. Electronics and Applicat. (ICIEA), 2010 the 5th IEEE Conf. on*, June 2010, pp. 56–61.

[19] Z.-X. Wang, C. Chang, X. German, and W. Tan, "Online fault detection of induction motors using independent component analysis and fuzzy neural network," in *Advances in Power System Control, Operation and Management (APSCOM 2009), 8th Int. Conf. on*, Nov 2009, pp. 1–6.

[20] M. Ballal, Z. Khan, M. Mishra, and R. Sonolikar, "Fuzzy system for the detection of winding insulation condition and bearing wear in induction motor," in *First India A Conf., 2004. Proc. IEEE INDICON 2004.*, Dec 2004, pp. 564–569.

[21] I. Bediaga, X. Mendizabal, I. Etxaniz, and J. Munoa, "An integrated system for machine tool spindle head ball bearing fault detection and diagnosis," *IEEE Instrum. Meas. Mag.*, vol. 16, no. 2, pp. 42–47, 2013.

[22] H. Abu-Rub, S. Ahmed, A. Iqbal, H. Toliyat, and M. Rahimian, "Incipient bearing fault detection for three-phase brushless dc motor drive using ANFIS," in *Diag. for Elect. Mach, Power Electronics Drives (SDEMPED), 2011 IEEE Int. Symp. on*, Sept 2011, pp. 620–625.

[23] J. Suwatthikul and S. Sornmuang, "Fault detection and diagnosis of a motor bearing shield," in *Proc. 6th IEEE Int. Conf. on Intell. Data Acquisition and Advanced Computing Syst.: Tech. and Applicat., IDAACS'2011*, vol. 1, 2011, pp. 332–335.

[24] K. Li, X. Ping, H. Wang, P. Chen, and Y. Cao, "Sequential fuzzy diagnosis method for motor roller bearing in variable operating conditions based on vibration analysis," *Sensors*, vol. 13, no. 6, pp. 8013–8041, 2013.

[25] X.-M. Han, "Crank bearing wear fault diagnosis of emulsion pump based on fuzzy support vector machine," in *2009 2nd Int. Conf. on Inform. and Computing Sci., ICIC 2009*, vol. 4, 2009, pp. 23–26.

[26] Z. Ye, C. Guoqiang, J. Yang, J. Zhang, Q. Shi, and H. Zeng, "Fault diagnosis of railway rolling bearing based on wavelet analysis and FCM," *Int. J. of Digital Content Tech. and its Applicat.*, vol. 5, no. 3, pp. 47–58, 2011.

[27] S. Zanoli and G. Astolfi, "Faults diagnosis for a centrifugal machine using the mahalanobis distance," in *IFAC Proc. Volumes*, vol. 8, no. PART 1, 2012, pp. 444–449.

[28] R. Zeng, L. Zhang, Y. Xiao, J. Mei, B. Zhou, H. Zhao, and J. Jia, "A method combining order tracking and fuzzy c-means for diesel engine fault detection and isolation," *Shock and Vibration*, vol. 2015, 2015.

[29] J. V. de Oliveira, "Semantic constraints for membership function optimization," *IEEE Trans. Syst., Man, Cybern. A*, vol. 29, no. 1, pp. 128–138, Jan 1999.

[30] E. Zio and G. Gola, "A neuro-fuzzy technique for diagnosing faults in rotating machinery," in *Proc. European Safety and Reliability Conf. 2006, ESREL 2006*, vol. 1, 2006, pp. 247–254.

[31] ——, "A neuro-fuzzy technique for fault diagnosis and its application to rotating machinery," *Reliability Eng. & System Safety*, vol. 94, no. 1, pp. 78 – 88, 2009.

[32] R. L. Winder and W. E. E. Littmann, *Bearing damage analysis*. National Bureau of Standard Publication, 1976.

[33] R. Tiwari, V. Gupta, and P. Kankar, "Bearing fault diagnosis based on multi-scale permutation entropy and adaptive neuro fuzzy classifier," *JVC/J. of Vib. and Control*, vol. 21, no. 3, pp. 461–467, 2015.

[34] I. Attoui, N. Boutasseta, N. Fergani, B. Oudjani, and A. Deliou, "Vibration-based bearing fault diagnosis by an integrated DWT-FFT approach and an adaptive neuro-fuzzy inference system," in *2015 3rd Int. Conf. on Control, Eng. Inform. Tech. (CEIT)*, May 2015, pp. 1–6.

[35] F. Pandansari, O. Wahyunggoro, and F. D. Wijaya, "Bearing faults analysis using modified ANFIS architecture in AC motors implementation," in *2016 8th Int. Conf. on Inform. Tech. and Elect. Eng. (ICITEE)*, Oct 2016, pp. 1–6.

[36] I. Attoui, N. Fergani, N. Boutasseta, B. Oudjani, and A. Deliou, "A new timefrequency method for identification and classification of ball bearing faults," *J. of Sound and Vib.*, vol. 397, pp. 241 – 265, 2017.

[37] Z. Xu, J. Xuan, T. Shi, B. Wu, and Y. Hu, "A novel fault diagnosis method of bearing based on improved fuzzy ARTMAP and modified distance discriminant technique," *Expert Syst. Applicat.*, vol. 36, no. 9, pp. 11 801 – 11 807, 2009.

[38] B. Li, P. yuan Liu, R. xi Hu, S. shan Mi, and J. ping Fu, "Fuzzy lattice classifier and its application to bearing fault diagnosis," *Ap Soft Computing*, vol. 12, no. 6, pp. 1708 – 1719, 2012.

[39] C.-F. Lin and S.-D. Wang, "Fuzzy support vector machines," *Trans. Neur. Netw.*, vol. 13, no. 2, pp. 464–471, Mar. 2002.

[40] H.-C. Kuo, L.-J. Wu, and J.-H. Chen, "Neural-fuzzy fault diagnosis in a marine propulsion shaft system," *J. of Materials Process. Tech.*, vol. 122, no. 1, pp. 12 – 22, 2002.

[41] H. Wang and P. Chen, "Fault diagnosis for a rolling bearing used in a reciprocating machine by adaptive filtering technique and fuzzy neural network," *WSEAS Trans. on Syst.*, vol. 7, no. 1, pp. 1–6, 2008.

[42] ——, "Sequential diagnosis for rolling bearing using fuzzy neural network," in *IEEE/ASME Int. Conf. on Advanced Intell. Mechatronics*, 2008, pp. 56–61.

[43] F. Zhao, J. Chen, L. Guo, and X. Li, "Neuro-fuzzy based condition prediction of bearing health," *J. of Vibration and Control*, vol. 15, no. 7, pp. 1079–1091, 2009.

[44] P. Fu, L. Jiang, W. Li, and A. Hope, "Bearing faults diagnosis based on b-spline neurofuzzy networks," in *Elect. and Control Eng. (ICECE), 2010 Int. Conf. on*, June 2010, pp. 540–543.

[45] G. Marichal, M. Artés, and J. Garcia-Prada, "An intelligent detection and classification system of faulty bearings," in *ASME Int. Mech. Eng. Congr. and Exposition, Proceedings*, vol. 15, 2010, pp. 229–237.

[46] G. Marichal, M. Arts, J. G. Prada, and O. Casanova, "Extraction of rules for faulty bearing classification by a neuro-fuzzy approach," *Mech. Syst. and Signal Process.*, vol. 25, no. 6, pp. 2073 – 2082, 2011.

[47] C. Chen, B. Zhang, and G. Vachtsevanos, "Prediction of machine health condition using neuro-fuzzy and bayesian algorithms," *IEEE Trans. Instrum. Meas.*, vol. 61, no. 2, pp. 297–306, Feb 2012.

[48] Z. Huang, Z. Yu, Z. Li, and Y. Geng, "A fault diagnosis method of rolling bearing through wear particle and vibration analyses," *Ap Mechanics and Materials*, vol. 26-28, pp. 676–681, 2010.

[49] T. Boukra, "Identifying new prognostic features for remaining useful life prediction using particle filtering and neuro-fuzzy system predictor," in *2015 IEEE 15th Int. Conf. on Environment and Elect. Eng. (EEEIC)*, June 2015, pp. 1533–1538.

[50] J. Liu, W. Wang, and F. Golnaraghi, "An enhanced diagnostic scheme for bearing condition monitoring," *IEEE Trans. Instrum. Meas.*, vol. 59, no. 2, pp. 309–321, 2010.

[51] H.-C. Kuo and H.-K. Chang, "A new symbiotic evolution-based fuzzy-neural approach to fault diagnosis of marine propulsion systems," *Eng. Applicat. of Art. Intell.*, vol. 17, no. 8, pp. 919–930, 2004.

[52] H. Abu-Rub, S. Ahmed, A. Iqbal, M. Rahimian, and H. Toliyat, "Incipient bearing fault diagnostics for inverter fed induction motor drive using ANFIS," in *19th Int. Conf. on Elect. Mach., ICEM 2010*, 2010, p. XP031779668.

[53] M. Seera and C. Lim, "Online motor fault detection and diagnosis using a hybrid fmm-cart model," *IEEE Trans. Neural Netw.*, vol. 25, no. 4, pp. 806–812, 2014.

[54] P. Fu, L. Jiang, A. Hope, and W. Li, "Fault diagnosis of bearings based on time-delayed correlation and demodulation as well as b-spline fuzzy neural networks," *Lecture Notes in Elect. Eng.*, vol. 67 LNEE, pp. 757–763, 2010.

[55] P. Fu, L. Jiang, W. Li, and A. Hope, "Bearing faults diagnosis based on b-spline neurofuzzy networks," in *Proc. - Int. Conf. on Elect. and Control Eng., ICECE 2010*, 2010, pp. 540–543.

[56] K. Li, P. Chen, and S. Wang, "An intelligent diagnosis method for rotating machinery using least squares mapping and a fuzzy neural network," *Sensors*, vol. 12, no. 5, pp. 5919–5939, 2012.

[57] J. Altmann and J. Mathew, "Multiple band-pass autoregressive demodulation for rolling-element bearing fault diagnosis," *Mech. Syst. and Signal Process.*, vol. 15, no. 5, pp. 963–977, 2001.

[58] M. Ballal, Z. Khan, H. Suryawanshi, and R. Sonolikar, "Adaptive neural fuzzy inference system for the detection of inter-turn insulation and bearing wear faults in induction motor," *IEEE Trans. Ind. Electron.*, vol. 54, no. 1, pp. 250–258, 2007.

[59] Y. Lei, Z. He, and Y. Zi, "Application of a novel hybrid intelligent method to compound fault diagnosis of locomotive roller bearings," *J. of Vibration and Acoustics, Trans. of the ASME*, vol. 130, no. 3, 2008.

1063-6706 (c) 2018 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2018.2878200, IEEE Transactions on Fuzzy Systems

[60] ——, "A new approach to intelligent fault diagnosis of rotating machinery," *Expert Syst. Applicat.*, vol. 35, no. 4, pp. 1593 – 1600, 2008.

[61] M. Yilmaz and E. Ayaz, "Adaptive neuro-fuzzy inference system for bearing fault detection in induction motors using temperature, current, vibration data," in *IEEE EUROCON*, 2009, pp. 1140–1145.

[62] W. Sui and D. Zhang, "Fault diagnosis of roller bearing conditions using ANFIS," *Ap Mechanics and Materials*, vol. 16-19, pp. 886–890, 2009.

[63] L. Zhang, G. Xiong, H. Liu, H. Zou, and W. Guo, "Bearing fault diagnosis using multi-scale entropy and adaptive neuro-fuzzy inference," *Expert Syst. Applicat.*, vol. 37, no. 8, pp. 6077 – 6085, 2010.

[64] J. Latuny and R. Entwistle, "A bearing fault classifier using artificial neuro-fuzzy inference system (ANFIS) based on statistical parameters and daubechies wavelet transform features," in *Advances in Ap Mechanics Res., Conf. Proc. - 7th Australasian Congr. on Ap Mechanics, ACAM 2012*, 2012, pp. 165–174.

[65] L. Ma, J. Kang, and C. Zhao, "Research on condition monitoring of bearing health using vibration data," *Ap Mechanics and Materials*, vol. 226-228, pp. 340–344, 2012.

[66] H. Ertunc, H. Ocak, and C. Aliustaoglu, "ANN- and ANFIS-based multi-staged decision algorithm for the detection and diagnosis of bearing faults," *Neural Computing and Applicat.*, vol. 22, no. SUPPL.1, pp. 435–446, 2013.

[67] J. Zhang, W. Ma, and L. Ma, "A fault diagnosis method based on ANFIS and bearing fault diagnosis," in *Proc. - 2014 Int. Conf. on Inform. Sci., Electronics and Elect. Eng., ISEEE 2014*, vol. 2, 2014, pp. 1274–1278.

[68] M. Saeidi, J. Zarei, H. Hassani, and A. Zamani, "Bearing fault detection via Park's vector approach based on ANFIS," in *Proc., 2014 Int. Conf. on Mechatronics and Control, ICMC 2014*, 2014, pp. 2436–2441.

[69] I. Attoui, A. Boudiaf, N. Fergani, B. Oudjani, N. Boutasseta, and A. Deliou, "Vibration-based gearbox fault diagnosis by DWPT and PCA approaches and an adaptive neuro-fuzzy inference system," 2016, pp. 234–239.

[70] F. Cheng, L. Qu, and W. Qiao, "A case-based data-driven prediction framework for machine fault prognostics," in *2015 IEEE Energy Conversion Congr. and Exposition (ECCE)*, Sept 2015, pp. 3957–3963.

[71] S. Rajabi, M. Samanazari, H. R. Momeni, and A. Ramezani, "Automated fault diagnosis of rolling element bearings based on morphological operators and M-ANFIS," in *2016 24th Iranian Conf. on Elect. Eng. (ICEE)*, May 2016, pp. 1757–1762.

[72] G. A. Carpenter, S. Grossberg, N. Markuzon, J. H. Reynolds, and D. B. Rosen, "Fuzzy ARTMAP: A neural network architecture for incremental supervised learning of analog multidimensional maps," *Neural Networks, IEEE Trans.*, vol. 3, no. 5, pp. 698–713, Sep. 1992.

[73] W. F. Godoy, R. H. C. Palcios, I. N. da Silva, A. Goedtel, and P. P. D. da Silva, "Intell. systems applied on the estimation of bearing faults in inverter-fed induction motors," in *IECON 2016 - 42nd A Conf. IEEE Ind. Electronics Society*, Oct 2016, pp. 1507–1512.

[74] I.-K. Jeong, M. Kang, J. Kim, J.-M. Kim, J.-M. Ha, and B.-K. Choi, "Enhanced det-based fault signature analysis for reliable diagnosis of single and multiple-combined bearing defects," *Shock and Vibration*, vol. 2015, 2015.

[75] Z. Xu, J. Xuan, T. Shi, B. Wu, and Y. Hu, "Application of a modified fuzzy ARTMAP with feature-weight learning for the fault diagnosis of bearing," *Expert Syst. Applicat.*, vol. 36, no. 6, pp. 9961 – 9968, 2009.

[76] Z. Tang and X. an, "Voting algorithm of fuzzy ARTMAP and its application to fault diagnosis," in *Fourth Int. Conf. on Fuzzy Syst. and Knowledge Discovery (FSKD 2007)*, vol. 4, Aug 2007, pp. 535–538.

[77] Z. Xu, H. Wu, and T. Shi, "Fault diagnosis of bearing based on selective ensemble of multiple fuzzy ARTMAP neural networks," *Ap Mechanics and Materials*, vol. 423-426, pp. 2480–2485, 2013.

[78] M. Jin, R. Li, Z. Xu, and X. Zhao, "Reliable fault diagnosis method using ensemble fuzzy ARTMAP based on improved bayesian belief method," *Neurocomputing*, vol. 133, pp. 309 – 316, 2014.

[79] Z. Xu, Y. Li, Z. Wang, and J. Xuan, "A selective fuzzy ARTMAP ensemble and its application to the fault diagnosis of rolling element bearing," *Neurocomputing*, vol. 182, pp. 25 – 35, 2016.

[80] V. Kaburlasos and V. Petridis, "Fuzzy lattice neurocomputing (FLN) models," *Neural Networks*, vol. 13, no. 10, pp. 1145 – 1170, 2000.

[81] S. Wadhwani, A. Wadhwani, S. Gupta, and V. Kumar, "Detection of bearing failure in rotating machine using adaptive neuro-fuzzy inference system," in *2006 Int. Conf. on Power Electronics, Drives and Energy Syst., PEDES '06*, 2006.

[82] Y. Pan, J. Chen, and G. Dong, "A hybrid model for bearing performance degradation assessment based on support vector data description and fuzzy c-means," *Proc. Institution of Mech. Engineers, Part C: J. of Mech. Eng. Sci.*, vol. 223, no. 11, pp. 2687–2695, 2009.

[83] P. Fu, W. Li, and W. Cao, "Fault diagnosis of bearings based on time-delayed correlation demodulation and fuzzy clustering," *Advanced Materials Res.*, vol. 211-212, pp. 510–514, 2011.

[84] S.-Q. Cao, X.-M. Zuo, A.-X. Tao, J.-M. Wang, and X.-Z. Chen, "A bearing intelligent fault diagnosis method based on cluster analysis," *Ap Mechanics and Materials*, vol. 152-154, pp. 1628–1633, 2012.

[85] W. Liu and J. Han, "A fuzzy clustering-based binary threshold bispectrum estimation approach," *Neural Computing and Applicat.*, vol. 21, no. 1, pp. 385–392, 2012.

[86] ——, "Rolling element bearing fault recognition approach based on fuzzy clustering bispectrum estimation," *Shock and Vibration*, vol. 20, no. 2, pp. 213–225, 2013.

[87] Y. Liang, L. Jia, G. Cai, and J. Liu, "A new approach to diagnose rolling bearing faults based on AFD," *Lecture Notes in Elect. Eng.*, vol. LNEE 288, no. Vol. 2, pp. 573–582, 2014.

[88] C. Wang, H. Li, W. Xiang, and D. Zhao, "A new signal classification method based on EEMD and FCM and its application in bearing fault diagnosis," *Ap Mechanics and Materials*, vol. 602-605, pp. 1803–1806, 2014.

[89] C. Liu, X. Huang, and X. Luo, "Roller bearing fault diagnosis based on ELMD and fuzzy C-means clustering algorithm," *Ap Mechanics and Materials*, vol. 602-605, pp. 1698–1700, 2014.

[90] Y. S. N. Pan and Z. Yi, "Rolling bearing faults detection based on improved sparse component analysis," in *2016 IEEE Int. Conf. on Inform. and Automation (ICIA)*, Aug 2016, pp. 772–775.

[91] S. Lotfan, N. Salehpour, H. Adiban, and A. Mashroutechi, "Bearing fault detection using fuzzy C-means and hybrid C-means-subtractive algorithms," in *2015 IEEE Int. Conf. on Fuzzy Syst.*, Aug 2015, pp. 1–7.

[92] W. Jiang, Z. Wang, Y. Zhu, Z. Zheng, and B. Zhang, "Fault recognition method for rolling bearing integrating VMD denoising and FCM clustering," *J. of Inform. and Computational Sci.*, vol. 12, no. 16, pp. 5967–5975, 2015.

[93] F. Wang, J. Sun, D. Yan, S. Zhang, L. Cui, and Y. Xu, "A feature extraction method for fault classification of rolling bearing based on PCA," *J. of Physics: Conf. Series*, vol. 628, no. 1, 2015.

[94] N. Pan, X. Wu, and Y. Guo, "Bearing compound fault acoustic diagnosis based on improved blind deconvolution algorithm," *Trans. of the Canadian Society for Mech. Eng.*, vol. 39, no. 3, pp. 657–667, 2015.

[95] S. Dong, X. Xu, and R. Chen, "Application of fuzzy c-means method and classification model of optimized k-nearest neighbor for fault diagnosis of bearing," *J. Brazilian Society of Mech. Sci. and Eng.*, vol. 38, no. 8, pp. 2255–2263, 2016.

[96] S. Fu, K. Liu, Y. Xu, and Y. Liu, "Rolling bearing diagnosing method based on time domain analysis and adaptive fuzzy c-means clustering," *Shock and Vibration*, vol. 2016, 2016.

[97] H. Yu, H. Li, and B. Xu, "Rolling bearing degradation state identification based on lcd relative spectral entropy," *J. of Failure Anal. and Prevention*, vol. 16, no. 4, pp. 655–666, 2016.

[98] L. Jiang, Y. Liu, X. Li, and A. Chen, "Fault diagnosis of roller bearing based on bispectrum estimation and fuzzy cluster analysis," *Ap Mechanics and Materials*, vol. 36, pp. 129–134, 2010.

[99] W. Sui, C. Lu, and D. Zhang, "Bearing fault diagnosis based on feature weighted FCM cluster analysis," in *Proc. - Int. Conf. on Comput. Sci. and Software Eng., CSSE 2008*, vol. 5, 2008, pp. 518–521.

[100] Y. Zhang, X.-D. Liu, F.-D. Xie, and K.-Q. Li, "Fault classifier of rotating machinery based on weighted support vector data description," *Expert Syst. Applicat.*, vol. 36, no. 4, pp. 7928–7932, 2009.

[101] M. Farajzadeh-Zanjani, R. Razavi-Far, M. Saif, J. Zarei, and V. Palade, "Diagnosis of bearing defects in induction motors by fuzzy-neighborhood density-based clustering," in *2015 IEEE 14th Int. Conf. on Mach. Learn. and Applicat. (ICMLA)*, Dec 2015, pp. 935–940.

[102] Z. Xu, Y. Li, Z. Wang, and J. Xuan, "A novel clustering method combining art with yu's norm for fault diagnosis of bearings," *Shock and Vibration*, vol. 2016, 2016.

[103] K. Yu, T. R. Lin, and J. W. Tan, "A bearing fault diagnosis technique based on singular values of EEMD spatial condition matrix and gath-geva clustering," *Ap Acoust.*, vol. 121, pp. 33 – 45, 2017.

[104] L. Jiang, Y. Liu, X. Li, and A. Chen, "Degradation assessment and fault diagnosis for roller bearing based on ar model and fuzzy cluster analysis," *Shock and Vib.*, vol. 18, no. 1-2, pp. 127–137, 2011.

[105] L. Han, C. Li, L. Zhan, and X. L. Li, "Rolling bearing fault diagnosis method based on EEMD permutation entropy and fuzzy clustering," in

1063-6706 (c) 2018 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2018.2878200, IEEE Transactions on Fuzzy Systems

2015 Fifth Int. Conf. on Instru. and Measure., Comput., Communication and Control (IMCCC), Sept 2015, pp. 470–474.

[106] J. Zhang, W. Ma, L. Ma, and Z. He, "Fault diagnosis model based on fuzzy support vector machine combined with weighted fuzzy clustering," *Trans. of Tianjin University*, vol. 19, no. 3, pp. 174–181, 2013.

[107] G. Vijay, S. Pai, N. Sriram, and R. Rao, "Bearing diagnostics - a radial basis function neural network approach," *Proc. international conference on Condition Monitoring and Diagnostic Eng. Management (COMADEM)*, pp. 843–854, 2011.

[108] C. Li, L. Ledo, M. Delgado, M. Cerrada, F. Pacheco, D. Cabrera, R.-V. Snchez, and J. V. de Oliveira, "A bayesian approach to consequent parameter estimation in probabilistic fuzzy systems and its application to bearing fault classification," *Knowledge-Based Syst.*, 2017.

[109] P. Chen and T. Toyota, "Sequential fuzzy diagnosis for plant machinery," *JSME Int. J., Series C: Mech. Syst., Mach. Elements and Manuf.*, vol. 46, no. 3, pp. 1121–1129, 2003.

[110] P. Chen, M. Taniguchi, and T. Toyota, "Intell. diagnosis method of multi-fault state for plant machinery using wavelet analysis, genetic programming and possibility theory," in *Proc. - IEEE Int. Conf. on Robotics and Automation*, vol. 1, 2003, pp. 610–615.

[111] H. Wang and P. Chen, "Sequential fuzzy diagnosis for condition monitoring of rolling bearing based on neural network," *Lecture Notes in Comput. Sci.*, vol. 5264 LNCS, no. PART 2, pp. 284–293, 2008.

[112] ——, "Fuzzy diagnosis method for rotating machinery in variable rotating speed," *Sensors J., IEEE*, vol. 11, no. 1, pp. 23–34, Jan 2011.

[113] ——, "Intell. diagnosis method for rolling element bearing faults using possibility theory and neural network," *Comput. & Ind. Eng.*, vol. 60, no. 4, pp. 511 – 518, 2011.

[114] H. Xue, Z. Li, Y. Li, H. Jiang, and P. Chen, "A fuzzy diagnosis of multi-fault state based on information fusion from multiple sensors," *J. of Vibroengineering*, vol. 18, no. 4, pp. 2135–2148, 2016.

[115] L. Zhang, G. Xiong, H. Liu, and H. Zou, "Bearing condition identification using improved multiscale entropy and svms," *Key Eng. Materials*, vol. 419-420, pp. 817–820, 2010.

[116] L. Zhang, G. Xiong, H. Liu, H. Zou, and W. Guo, "Applying improved multi-scale entropy and support vector machines for bearing health condition identification," *Proc. Institution of Mech. Engineers, Part C: J. of Mech. Eng. Sci.*, vol. 224, no. 6, pp. 1315–1325, 2010.

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

[118] M. L. D. Wong, S. H. Lee, and A. K. Nandi, "Anomaly detection of rolling elements using fuzzy entropy and similarity measures." in *10th Int. Conf. on Vib. in Rotating Machinery*. Woodhead Publishing, 2012, pp. 693–702.

[119] J. Zheng, J. Cheng, and Y. Yang, "A rolling bearing fault diagnosis approach based on LCD and fuzzy entropy," *Mechanism and Mach. Theory*, vol. 70, pp. 441–453, 2013.

[120] 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 Mach. Theory*, vol. 78, pp. 187 – 200, 2014.

[121] Z. Jinde, J. Zhanwei, P. Ziwei, and Z. Kang, "VMD based adaptive multiscale fuzzy entropy and its application to rolling bearing fault diagnosis," in *2016 10th Int. Conf. on Sensing Tech. (ICST)*, Nov 2016, pp. 1–4.

[122] Y. Li, M. Xu, Y. Wei, and W. Huang, "Bearing fault diagnosis based on adaptive mutiscale fuzzy entropy and support vector machine," *J. of Vibroengineering*, vol. 17, no. 3, pp. 1188–1202, 2015.

[123] Y. Li, M. Xu, H. Zhao, and W. Huang, "Hierarchical fuzzy entropy and improved support vector machine based binary tree approach for rolling bearing fault diagnosis," *Mechanism and Mach. Theory*, vol. 98, pp. 114 – 132, 2016.

[124] Y. Li, M. Xu, R. Wang, and W. Huang, "A fault diagnosis scheme for rolling bearing based on local mean decomposition and improved multiscale fuzzy entropy," *J. of Sound and Vibration*, vol. 360, pp. 277 – 299, 2016.

[125] Y. Yang, H. Pan, L. Ma, and J. Cheng, "A fault diagnosis approach for roller bearing based on improved intrinsic timescale decomposition de-noising and kriging-variable predictive model-based class discriminate," *JVC/J. of Vibration and Control*, vol. 22, no. 5, pp. 1431–1446, 2016.

[126] K. Zhu and H. Li, "A rolling element bearing fault diagnosis approach based on hierarchical fuzzy entropy and support vector machine," *Proc. Institution of Mech. Engineers, Part C: J. of Mech. Eng. Sci.*, vol. 230, no. 13, pp. 2314–2322, 2016.

[127] H. Zhao, J. Wang, H. Han, and Y. Gao, "A feature extraction method based on HLMD and MFE for bearing clearance fault of reciprocating compressor," *Measure.*, vol. 89, pp. 34 – 43, 2016.

[128] 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, 2017.

[129] J. Zheng, H. Pan, and J. Cheng, "Rolling bearing fault detection and diagnosis based on composite multiscale fuzzy entropy and ensemble support vector machines," *Mech. Syst. and Signal Process.*, vol. 85, pp. 746 – 759, 2017.

[130] T. Liu, J. Singonahalli, and N. Iyer, "Detection of roller bearing defects using expert system and fuzzy logic," *Mech. Syst. and Signal Process.*, vol. 10, no. 5, pp. 595 – 614, 1996.

[131] S. da Silva Vicente, R. Fujimoto, and L. Padovese, "Rolling bearing fault diagnostic system using fuzzy logic," in *Fuzzy Syst., 2001. The 10th IEEE Int. Conf. on*, vol. 2, Dec 2001, pp. 816–819 vol.3.

[132] Y. Zhao, G. Zhang, J. Du, G. Wang, and G. Vachtsevanos, "Development of a distributed bearing health monitoring and assessing system," in *2004 8th Int. Conf. on Control, Automation, Robotics and Vision (ICARCV)*, vol. 1, 2004, pp. 474–478.

[133] F. Harrouche and A. Felkaoui, "Automation of fault diagnosis of bearing by application of fuzzy inference system (FIS)," *Mechanics and Industry*, vol. 15, no. 6, pp. 477–485, 2014.

[134] G. Nirmal and C. Gomathy, "Fault detection and classification method to prevent derailment of rolling stock," in *2015 Int. Conf. on Communications and Signal Process. (ICCSP)*, April 2015, pp. 0959–0963.

[135] I. Aydn, M. Karaköse, and E. Akin, "Combined intelligent methods based on wireless sensor networks for condition monitoring and fault diagnosis," *J. of Intell. Manuf.*, vol. 26, no. 4, pp. 717–729, 2015.

[136] D. Prince Winston, G. Sakthivel, and M. Saravanan, "Single parameter fault identification technique for three phase induction motor through wavelet analysis and fuzzy logic," *J. of Elect. Eng.*, vol. 16, no. 2, pp. 1–9, 2016.

[137] A. Raj and N. Murali, "Early classification of bearing faults using morphological operators and fuzzy inference," *IEEE Trans. Ind. Electron.*, vol. 60, no. 2, pp. 567–574, Feb 2013.

[138] V. Sugumaran and K. Ramachandran, "Automatic rule learning using decision tree for fuzzy classifier in fault diagnosis of roller bearing," *Mech. Syst. and Signal Process.*, vol. 21, no. 5, pp. 2237 – 2247, 2007.

[139] P. Jayaswal, S. Verma, and A. Wadhwani, "Application of ann, fuzzy logic and wavelet transform in machine fault diagnosis using vibration signal analysis," *J. of Quality in Maintenance Eng.*, vol. 16, no. 2, pp. 190–213, 2010.

[140] ——, "Development of EBP-artificial neural network expert system for rolling element bearing fault diagnosis," *JVC/J. of Vib. and Control*, vol. 17, no. 8, pp. 1131–1148, 2011.

[141] V. Sugumaran and K. Ramachandran, "Fault diagnosis of roller bearing using fuzzy classifier and histogram features with focus on automatic rule learning," *Expert Syst. Applicat.*, vol. 38, no. 5, pp. 4901–4907, May 2011.

[142] T. Boutros and M. Liang, "Mech. fault detection using fuzzy index fusion," *Int. J. of Mach. Tools and Manufacture*, vol. 47, no. 11, pp. 1702 – 1714, 2007.

[143] C. Martnez-Garca, G. L. Osorio-Gordillo, C. M. Astorga-Zaragoza, and V. Puig, "Fault detection on bearings coupled to permanent magnet DC motors by using a generalized takagi-sugeno PI observer," in *2016 3rd Conf. on Control and Fault-Tolerant Systems (SysTol)*, Sept 2016, pp. 579–584.

[144] Y. Forghani, H. Sadoghi Yazdi, and S. Effati, "An extension to fuzzy support vector data description (FSVDD)," *Pattern Anal. and Applicat.*, vol. 15, no. 3, pp. 237–247, 2012.

[145] Z. Shen, Z. He, X. Chen, C. Sun, and Z. Liu, "A monotonic degradation assessment index of rolling bearings using fuzzy support vector data description and running time," *Sensors*, vol. 12, no. 8, p. 10109, 2012.

[146] J. V. de Oliveira, "Neuron inspired learning rules for fuzzy relational structures," *Fuzzy Sets and Syst.*, vol. 57, no. 1, pp. 41 – 53, 1993.

[147] D. Stefanoiu, "Fault diagnosis by fuzzy-statistical reasoning," *J. of Intell. and Fuzzy Syst.*, vol. 15, no. 3-4, pp. 181–194, 2004.

[148] D. Stefanoiu and F. Ionescu, "A fuzzy-statistical reasoning model for bearings fault diagnosis," in *Proc. IEEE Int. Conf. on Sys., Man and Cybern.*, vol. 7, 2002, pp. 497–502.

[149] J. Yan, L. Lu, D. Zhao, and G. Wang, "Diagnosis of bearing incipient faults using fuzzy logic based methodology," in *Fuzzy Syst. and Knowledge Discovery (FSKD), 2010 Int. Conf. on*, vol. 3, Aug 2010, pp. 1229–1233.

1063-6706 (c) 2018 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2018.2878200, IEEE Transactions on Fuzzy Systems

[150] Z. Yao, G. Lou, X. Song, and Y. Zhou, "On-line fault diagnosis study for roller bearing based on fuzzy fault tree," in *CAR 2010 - 2nd Int. Asia Conf. on Informatics in Control, Automation and Robotics*, vol. 1, 2010, pp. 182–185.

[151] X. Yu and J. Liu, "Research on fault diagnosis and application for rolling element bearing based on fuzzy analysis," *Advanced Materials Res.*, vol. 189-193, pp. 1358–1361, 2011.

[152] W. Li, W. Pan, and S. Zhang, "Fault diagnosis using rough sets and BP networks," in *Mechanic Automation and Control Eng. (MACE), 2010 Int. Conf. on*, June 2010, pp. 585–588.

[153] F. Pacheco, M. Cerrada-Lozada, R.-V. Sánchez, D. Cabrera, C. Li, and J. V. de Oliveira, "Clustering algorithm using rough set theory for unsupervised feature selection," in *2016 Int. Joint Conf. on Neural Networks (IJCNN)*, July 2016, pp. 3493–3499.

[154] F. Pacheco, M. Cerrada, R.-V. Sanchez, D. Cabrera, C. Li, and J. V. de Oliveira, "Attribute clustering using rough set theory for feature selection in fault severity classification of rotating machinery," *Expert Syst. Applicat.*, vol. 71, pp. 69 – 86, 2017.

[155] R. Jensen and Q. Shen, "Semantics-preserving dimensionality reduction: rough and fuzzy-rough-based approaches," *IEEE Trans. Knowl. Data Eng.*, vol. 16, no. 12, pp. 1457–1471, Dec 2004.

[156] E. Juuso, M. Ruusunen, and G. Perigot, "Linguistic equation models for failure mode identification from multisensor vibration analysis," in *7th Int. Conf. on Condition Monitoring and Machinery Failure Prevention Technologies 2010, CM 2010/MFPT 2010*, vol. 2, 2010, pp. 1408–1420.

[157] S. Lahdelma and E. Juuso, "Advanced signal processing and fault diagnosis in condition monitoring," *Insight: Non-Destructive Testing and Condition Monitoring*, vol. 49, no. 12, pp. 719–725, 2007.

[158] G. Shafer, *A Mathematical Theory of Evidence*. Princeton: Princeton University Press, 1976.

[159] C. Chen and G. Vachtsevanos, "Bearing condition prediction considering uncertainty: An interval type-2 fuzzy neural network approach," *Robot. Comput.-Integr. Manuf.*, vol. 28, no. 4, pp. 509–516, Aug. 2012.

[160] J. Zarei, M. M. Arefi, and H. Hassani, "Bearing fault detection based on interval type-2 fuzzy logic systems for support vector machines," in *2015 6th Int. Conf. on Modeling, Simulation, and Ap Optimization (ICMSAO)*, May 2015, pp. 1–6.

[161] J. Mendel, F. Liu, and D. Zhai, "α-plane representation for type-2 fuzzy sets: Theory and applicat." *IEEE Trans. Fuzzy Syst.*, vol. 17, no. 5, pp. 1189–1207, Oct 2009.

[162] C. Wagner and H. Hagras, "Toward general type-2 fuzzy logic systems based on zslices," *IEEE Trans. Fuzzy Syst.*, vol. 18, no. 4, pp. 637–660, Aug 2010.

[163] H. Hassani, J. Zarei, M. M. Arefi, and R. Razavi-Far, "zSlices-based general type-2 fuzzy fusion of support vector machines with application to bearing fault detection," *IEEE Trans. Ind. Electron.*, 2017.

[164] J. M. Keller, M. R. Gray, and J. A. Givens, "A fuzzy k-nearest neighbor algorithm," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-15, no. 4, pp. 580–585, July 1985.

[165] J. van den Berg, U. Kaymak, and R. Almeida, "Conditional density estimation using probabilistic fuzzy systems," *IEEE Trans. Fuzzy Syst.*, vol. 21, no. 5, pp. 869–882, Oct 2013.

[166] H.-E. Lee, K.-H. Park, and Z. Z. Bien, "Iterative fuzzy clustering algorithm with supervision to construct probabilistic fuzzy rule base from numerical data," *IEEE Trans. Fuzzy Syst.*, vol. 16, no. 1, pp. 263–277, Feb 2008.

[167] L. Waltman, U. Kaymak, and J. van den Berg, "Maximum likelihood parameter estimation in probabilistic fuzzy classifiers," in *Fuzzy Syst.,14th IEEE Int. Conf. on*, May 2005, pp. 1098–1103.

[168] K. Loparo, "Bearings vibration data set. The Case Western Reserve University bearing data center." http://www.eecs.cwru.edu/laboratory/bearing/download.htm, 2003, [Accessed Oct. 2011].

[169] H. Tian, X.-Y. Kang, J.-N. Zhang, and S.-S. Han, "Application of fuzzy rough sets in patterns recognition of bearing," in *Quality, Reliability, Risk, Maintenance, and Safety Eng. (ICQR2MSE), 2012 Int. Conf. on*, June 2012, pp. 731–734.

[170] S. Sharma, W. Abed, R. Sutton, and B. Subudhi, "Corrosion fault diagnosis of rolling element bearing under constant and variable load and speed conditions," *IFAC-PapersOnLine*, vol. 48, no. 30, pp. 49 – 54, 2015.

[171] X. Liu, L. Ma, and J. Mathew, "Machinery fault diagnosis based on fuzzy measure and fuzzy integral data fusion techniques," *Mech. Syst. and Signal Process.*, vol. 23, no. 3, pp. 690 – 700, 2009.

[172] Y. He, F. Chu, and B. Zhong, "A study on group decision-making based fault multi-symptom-domain consensus diagnosis," *Reliability Eng. and System Safety*, vol. 74, no. 1, pp. 43–52, 2001.

[173] W.-X. Xu, J.-W. Tan, and H. Zhan, "Research and application of the improved dst new method based on fuzzy consistent matrix and the weighted average," *Advanced Materials Res.*, vol. 1030-1032, pp. 1764–1768, 2014.

[174] X. Sun, J. Tan, Y. Wen, and C. Feng, "Rolling bearing fault diagnosis method based on data-driven random fuzzy evidence acquisition and dempster-shafer evidence theory," *Advances in Mech. Eng.*, vol. 8, no. 1, 2016.

[175] C. Li, J. V. de Oliveira, R. Sánchez, M. Cerrada, G. Zurita, and D. Cabrera, "Fuzzy determination of informative frequency band for bearing fault detection," *J. of Intell. and Fuzzy Syst.*, vol. 30, no. 6, pp. 3513–3525, 2016.

[176] Y. Huang, L. Gong, S. Wang, and L. Li, "A fuzzy based semi-supervised method for fault diagnosis and performance evaluation," in *Advanced Intell. Mechatronics (AIM), 2014 IEEE/ASME Int. Conf. on*, July 2014, pp. 1647–1651.

[177] A. Fernandez, L. Gonzalez, I. Bediaga, A. Gaston, and J. Hernandez, "Automatic fault detection and diagnosis implementation based on intelligent approaches," in *Emerging Technologies and Factory Automation, 2005. ETFA 2005. 10th IEEE Conf. on*, vol. 1, Sept 2005, pp. 8–586.

[178] H. Ertunc, H. Ocak, and C. Aliustaoglu, "ANN- and ANFIS-based multi-staged decision algorithm for the detection and diagnosis of bearing faults," *Neural Computing and Applicat.*, vol. 22, no. 1, pp. 435–446, 2013.

[179] X. Chen, H. Zeng, and Z. Li, "A multi-fault diagnosis method of rolling bearing based on wavelet-PCA and fuzzy k-nearest neighbor," *Ap Mechanics and Materials*, vol. 29-32, pp. 1602–1607, 2010.

[180] Z. Li, X. Yan, Z. Tian, C. Yuan, Z. Peng, and L. Li, "Blind vibration component separation and nonlinear feature extraction applied to the nonstationary vibration signals for the gearbox multi-fault diagnosis," *Measure.*, vol. 46, no. 1, pp. 259 – 271, 2013.

[181] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, Sep. 1995.

[182] Y. Zhan, Q. Tan, and Y. Zhang, "The application of morphology analysis and BTFSVM to intelligent fault diagnosis on the bearing of ships," in *2009 2nd Int. Symp. on Knowledge Acquisition and Modeling, KAM 2009*, vol. 3, 2009, pp. 233–236.

[183] S. Shen, Y. Zhan, and Q. Tan, "The application of morphology analysis and RFFSVM to intelligent fault diagnosis on the bearing of ships," in *2010 Int. Conf. on E-Health Networking, Digital Ecosystems and Technologies, EDT 2010*, vol. 1, 2010, pp. 386–389.

[184] W. Zhao and L. Wang, "Rolling bearing fault diagnosis based on wavelet packet feature entropy-MFSVM," *Advanced Materials Res.*, vol. 121-122, pp. 813–818, 2010.

[185] J. Zhang, W. Ma, J. Lin, L. Ma, and X. Jia, "Fault diagnosis approach for rotating machinery based on dynamic model and computational intelligence," *Measure.*, vol. 59, pp. 73 – 87, 2015.

[186] P. Fazendeiro and J. V. de Oliveira, "Observer-biased fuzzy clustering," *IEEE Trans. Fuzzy Syst.*, vol. 23, no. 1, pp. 85–97, 2015.

[187] C. Li and J. Zhou, "Semi-supervised weighted kernel clustering based on gravitational search for fault diagnosis," *ISA Trans.*, vol. 53, no. 5, pp. 1534–1543, 2014.

[188] M. Amar, I. Gondal, and C. Wilson, "Fuzzy logic inspired bearing fault-model membership estimation," in *2013 IEEE 8th Int. Conf. on Intell. Sensors, Netw. and Inform. Process*, vol. 1, 2013, pp. 420–425.

[189] Y. Yang, L. Zhang, L. Zhang, X. Cai, and S. Zhang, "The health degree evaluation of rolling element bearings using an improved BP neural network," *J. of Inform. and Computational Sci.*, vol. 9, no. 14, pp. 4217–4227, 2012.

[190] L. Zhang, Z. Wang, and S. Zhao, "Short-term fault prediction of mechanical rotating parts on the basis of fuzzy-grey optimising method," *Mech. Syst. and Signal Process.*, vol. 21, no. 2, pp. 856 – 865, 2007.

[191] J. L. Deng, "Introduction to grey system theory," *J. Grey Syst.*, vol. 1, no. 1, pp. 1–24, Nov. 1989.

[192] C. Li, M. Cerrada, D. Cabrera, R. V. Sanchez, F. Pacheco, G. Ulutagay, and J. V. D. Oliveira, "Some preliminary results on the comparison of fcm, gk, fcmfp and fn-dbscan for bearing fault diagnosis," in *2017 International Conference on Sensing, Diagnostics, Prognostics, and Control (SDPC)*, Aug 2017, pp. 41–46.

[193] ——, "A comparison of fuzzy clustering algorithms for bearing fault diagnosis," *J. of Intell. and Fuzzy Syst*, vol. 34, no. 6, pp. 3565–3580, 2018.

[194] P. Fazendeiro, J. V. de Oliveira, and W. Pedrycz, "A multiobjective design of a patient and anaesthetist-friendly neuromuscular blockade controller," *IEEE Trans. Biomed. Eng.*, vol. 54, no. 9, pp. 1667–1678, Sept 2007.

1063-6706 (c) 2018 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.