The International Journal of Advanced Manufacturing Technology (2021) 115:1005–1019
https://doi.org/10.1007/s00170-021-07253-6

ORIGINAL ARTICLE

Harnessing fuzzy neural network for gear fault diagnosis with limited data labels

Kai Zhou[1] · Jiong Tang[2]

Received: 17 December 2020 / Accepted: 6 May 2021 / Published online: 17 May 2021

Abstract

Diagnosis and prognosis of gear systems play an important role in modern manufacturing. While first-principle-based inverse analysis is subject to various limitations, data-driven approaches such as many machine learning techniques have shown great promise in recent years. Nevertheless, major challenges remain. Machine learning generally requires large amount of high-quality training data which may not be available for many industrial systems. In particular, while gear faults are continuous in nature and exhibit many different scenarios, in practical situations owing to the high cost in data acquisition especially for fault scenarios, only a small number of discrete classes of faults, i.e., fault types and severities, can be recorded and employed in training. As such, the neural networks trained will need to deal with unseen faults when they are actually implemented. To tackle this challenge, in this research, we develop a fuzzy classification approach capable of handling fault scenarios that are not included in the training dataset. Through the integration of a fuzzification procedure, this fuzzy neural network (FNN) can produce classification outcome with probability and confidence level. An unseen fault scenario will be classified into the nearest fault class with probability, effectively yielding the diagnosis result under limited data. While fault features in gear vibration signals are hidden and have complex nonlinear relations with respect to fault scenarios, it is found that the kernel principal component analysis (KPCA) can enable the FNN to facilitate the correlation of fault features. Systematic case studies using experimental data acquired from a lab-scale gear system are carried out to validate the new approach.

Keywords Gear fault diagnosis · Unseen fault scenarios · Fuzzy neural network (FNN) · Kernel principal component analysis (KPCA) · Fuzzy classification

1 Introduction

Gear systems are widely used in modern manufacturing industry. Condition monitoring and fault diagnosis and prognosis of gears play an important role in ensuring the system integrity and performance. Different types of signals, such as vibration [32, 34], acoustic emission [46], and eddy current measurement [16], are employed to facilitate fault diagnosis. Among them, vibration signals are most commonly used because they can be readily measured through a variety of low-cost sensors and data acquisition systems. The physical features in gear vibration include the sideband frequencies and the keynote meshing frequency and its harmonics, which may offer insights to gear faults [27]. The vibration-based gear fault diagnosis methods include the model-based and model-free methods. The model-based methods are built upon the first principle where the effect of gear fault is reflected with the parametric variation of the healthy system that is numerically or analytically modeled. After updating the associated parameters in the numerical/analytical model by using experimental measurement, one can estimate the actual fault [8, 28]. This type of method resorts to an inverse analysis–based optimization procedure. A major challenge in these methods is the difficulty in establishing high-fidelity baseline model, as gear vibrations are intrinsically multi-scale problem with significant modeling uncertainties. In comparison, the model-free methods analyze experimental data directly to infer fault occurrence, thereby avoiding the challenge in establishing the baseline model. Many model-free gear fault diagnosis methods are built upon signal processing and feature

✉ Jiong Tang
jiong.tang@uconn.edu

[1] Department of Mechanical Engineering-Engineering Mechanics, Michigan Technological University, Houghton, MI, USA

[2] Department of Mechanical Engineering, University of Connecticut, Storrs, CT 06269, USA

extraction techniques, including but not limited to spectrum analysis [14], Wigner-Ville distribution [39], cyclostationary signals [5], envelope analysis [29], and various wavelet transforms [10, 24, 43]. While these methods have shown different levels of success, selection of specific features is often based on empirical judgement and experience.

With the rapid advancement of machine learning techniques, data-driven approaches have shown many advantageous aspects in gear fault diagnosis and prognosis [21, 35–37, 40, 42, 50]. Various neural networks and surrogate models can be trained to develop the mapping relation between signals and fault conditions, known as classifiers. To enhance the classification performance, they oftentimes are executed in conjunction with the aforementioned signal processing techniques. Li et al. [22, 23] proposed a combination of Kalman filter-based signal processing and the least square support vector machine to identify the fault types of planetary gearboxes. Dibaj et al. [13] developed a signal decomposition technique to extract the dominant vibration modes, upon which the support vector machine classifier was then applied to detect the defective status of gearbox system. Chen et al. [9] presented a gear fault diagnosis approach by integrating the empirical mode decomposition, singular value decomposition, and random forest. Unsurprisingly, like in many other disciplines, the deep learning neural network models have been increasingly investigated and employed for gear fault diagnosis due to their powerful inference capability. Indeed, when the amount of training data is large, deep learning can facilitate the analysis of raw vibration data even without going through pre-processing [2, 36, 37, 44, 49]. That is, when the parameters of a deep learning model are properly trained via large dataset, representative features can be automatically extracted in a hierarchy of conceptual abstractions. Jing et al. [19] developed a convolutional neural network to learn features from raw data, frequency spectrum, and combined time-frequency data, and indicated that it outperformed other inference methods. Wang et al. (2018) proposed an intelligent diagnosis scheme based upon the generative adversarial deep learning neural networks, in which the generator and discriminator networks were concurrently optimized to enhance the ability of fault classification. Li et al. [22, 23] performed planetary gear fault diagnosis using deep learning neural network with motor current signal. Wu et al. [41] constructed a one-dimensional convolutional neural network model for gear fault diagnosis, and verified its feasibility through gearbox challenge data and a planetary gearbox test rig.

Although machine learning for gear diagnosis and prognosis is extremely promising, the performance hinges upon both the quantity and quality of the training data. Fault occurrence in modern systems is relatively infrequent. In practical situations owing to the high cost in data acquisition especially for different fault scenarios, the inadequacy of labeled training samples, i.e., experimental data of known fault classes, is a common issue. As such, Cao et al. [6] proposed to employ deep convolutional neural network–based transfer learning (e.g., Alexnet) to solve the data quantity issue. In their approach, the issue of limited data is overcome by formulating a new neural network architecture that consists of two parts. Massive image data from ImageNet (http://www.image-net.org/challenges/LSVRC/2010/) were used first to train an original deep neural network model, the parameters of which were transferred to the new architecture as the first part. The second part of the architecture, an untrained neural network, then accommodated the gear fault diagnosis task and was further trained using experimentally generated gear fault data. It was found that this technique could effectively address the issue of insufficient labeled data for the system analyzed. Similar transfer-learning strategies were recently adopted by He et al. [18] and Chen et al. [11]. It is worth noting that, despite these advancements, there are still major challenges in terms of the quality of training data. In particular, while gear faults are continuous in nature and exhibit many different scenarios, only a small number of discrete classes of faults, e.g., fault with different severity levels, can be recorded and employed in training. As such, the training data do not have sufficient quality to distinguish different fault severity levels, and the neural networks trained will need to deal with unseen faults when they are actually implemented. Intuitively, a regression-type neural network may offer the extrapolation capability to handle the unseen fault scenarios. There are, nevertheless, significant obstacles. First and foremost, gear faults feature both categorical and continuous scenarios, which makes it nearly infeasible to develop simple regression to cover all fault scenarios. Secondly, even for those continuous scenarios (i.e., same fault with different severity levels), their relations with respect to the underlying features in vibration signals are generally nonlinear. These, coupled with the fact that the number of labels (i.e., fault scenarios) is usually small, make it difficult for the neural network to carry out extrapolation in a deterministic sense. Massive scenarios would be needed to produce reasonable regression analysis, which will yield extremely high data acquisition cost and computational cost.

An alternative way of handling unseen fault scenarios, as proposed in this research, is to adopt a non-deterministic, probability (i.e., confidence level)-based decision-making, which to a large extent mimics human cognizance. Instead of extrapolation, we want to fully utilize the existing labels in measurement data available. That is, when an unseen fault scenario is analyzed, the neural network will produce a probability assessment of its correlation with respect to all existing labels. We aim at developing a neural network that is capable of predicting correctly that an unseen scenario has the highest

probability of being correlated to the closest fault severity level in the training data. For example, if 30% chipping gear tip is an unseen scenario but the neural network can classify it to be highly similar with labeled scenario of 20% or 40% chipping tip, the neural network can indeed fulfill the diagnosis mission. To facilitate this, we leverage upon the concept of fuzzy neural network (FNN) which is a type of neuro-fuzzy inference systems [25]. FNN starts with the development of "fuzzy neurons" to describe the fuzzy rules, building the synaptic connections which incorporate fuzziness into the neural network. The perception in response to the linguistic statements is used as an input vector to a neural network with fully connected layers [7, 26]. There have been recent efforts in utilizing FNN concept toward machinery diagnosis. For example, Soualhi et al. [33] developed an FNN model to conduct diagnosis of critical components of the gear reducer. Ahuja et al. [1] integrated an FNN model with the continuous wavelet transform to identify the faults of two different types of gearboxes under nonstationary conditions. It should be noted that in previous investigations, the FNN concept and its linguistic description were employed to mainly handle the uncertain and imprecise data and the associated features, whereas the fault classifications were still deterministic with respect to existing labels. Nevertheless, the issue of limited fault labels, which commonly exists in manufacturing industry, has not yet been tackled. In this research, taking advantage of the versatile architecture of neuro-fuzzy inference, we plan to formulate a FNN where a fuzzification procedure is incorporated into the backpropagation network to specifically analyze unseen fault scenarios based on features extracted. Moreover, this new FNN is built upon a data pre-processing procedure through the kernel principal component analysis (KPCA) [12, 47, 48] that leads to a nonlinear feature transformation and dimension reduction.

The remainder of this paper is organized as follows. In Section 2, the architecture of the new FNN specifically tailored for gear fault diagnosis with limited labels is outlined, followed by the KPCA method employed to enable the probabilistic assessment. Section 3 provides FNN implementation details and systematic case studies with experimental investigation. Concluding remarks are summarized in Section 4.

2 Fuzzy fault diagnosis framework

In this section, we first present the fuzzy neural network (FNN) designed and tailored to allow gear fault diagnosis with limited labels. The FNN model construction and the subsequent classification analysis are further integrated with the kernel principal component analysis (KPCA) which aims at reducing the feature dimensionality and yielding the feature correlation for unseen scenarios.

2.1 Architecture of fuzzy neural network for fuzzy classification

FNN refers to the combination of fuzzy logic and neural network. It incorporates fuzzy algorithm to process information that is intrinsically fuzzy and then to learn the nature of data through a high-speed parallel structure [45]. The concept of FNN combines the advantages of both artificial neural networks and fuzzy qualitative modeling approaches, leading to adaptation capability, rapid learning capacity, and simplicity and ease in handling vagueness, uncertain, and imprecise data [15]. The fuzzy rules and the neural network in the FNN model function as fuzzification and defuzzification, respectively. The backpropagation training tunes optimally both the rule weights (i.e., the membership function parameters) and the weights in the fully connected layers. While previous investigations were mainly focused on deterministic classification, it was recognized that FNN would possess better generalization capability to overcome overfitting. Here, we investigate fuzzy classification with limited labels. We start from presenting the specific FNN architecture tailored for this purpose.

The FNN to be established consists of six layers, as shown in Fig. 1. The first layer is the usual input layer, which feeds the input information into the neural network. Each node of this layer carries one feature of input sample. As will be shown later, in the proposed classification system for fuzzy gear diagnosis, each input sample is a vector $[x_1, x_2, \ldots, x_n]$ representing the features of the principal component (PC) information of sampled raw vibration signal. n denotes the number of input variables. The node input-output function is given as,

$$g_i^{(1)} = x_i \quad i = 1, 2, \ldots, n \tag{1}$$

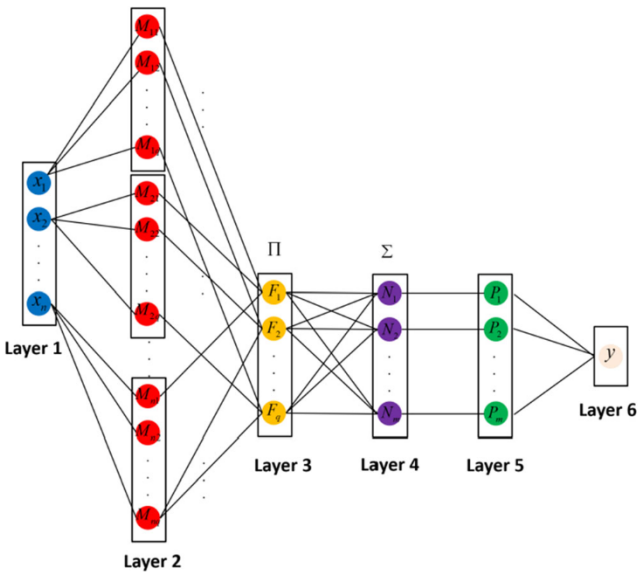

Fig. 1 FNN architecture

The second layer is the fuzzification layer, which is one of the core elements to incorporate fuzzy reasoning. The fuzzy rules are integrated into this layer, and they are characterized by membership functions mapping the point from input space into a membership value (or degree of membership). Mathematically, the membership value can be interpolated in terms of the output of the input layer as shown in Fig. 2. The node input-output functions can be described as

$$g_{ij}^{(2)} = g_{ij}^{(2)}\left(g_i^{(1)}, \boldsymbol{\beta}_j\right) \quad i = 1, 2, \ldots, n \quad j = 1, 2, \ldots, q \tag{2}$$

where $\boldsymbol{\beta}_j$ is the vector of parameters associated with the jth membership function, and q is the number of membership functions for each input. Membership functions can generally be modeled as basic functions such as piecewise linear functions, Gaussian distribution function, sigmoid curve, and quadratic and cubic polynomial curves [25]. The membership function usually can be selected either by looking into the distribution of physical data or by resorting to trial-and-error. Gaussian membership functions that are defined by two parameters have the advantage of simplicity, enabling the efficient computation in model training. Additionally, they are able to specify the fuzzy sets with smoothness and concise notation [30]. For this reason, we adopt Gaussian membership function in this research (More details will be presented in Section 3).

The third layer is the fuzzy reasoning/rule layer. It activates its affiliated rule nodes to take actions accordingly in terms of antecedents from the fuzzification layer. The output of this layer is the firing strength, which mathematically is represented as a multiplication form of related membership values. For example, given the fuzzification layer antecedent, i.e., if (x_1 is M_{1j}) and if (x_2 is M_{2j})… and if (x_n is M_{nj}), then

$$g_j^{(3)} = \prod_{i=1}^{n} g_{ij}^{(2)} \quad j = 1, 2, \ldots, q \tag{3}$$

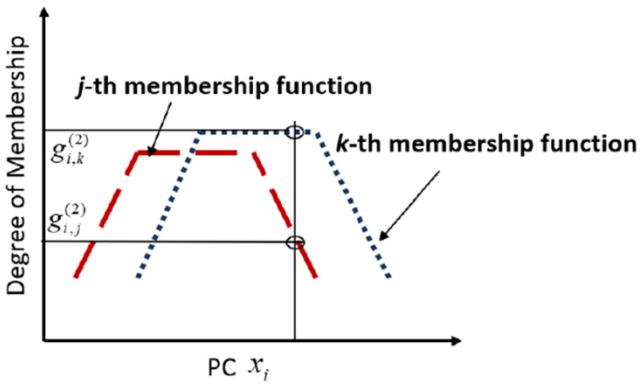

Fig. 2 Illustration of membership value interpolation

Here, only q firing strength items are involved in the above equation for illustration purposes. Theoretically, the maximum number of firing strength items can reach up to q^n. The inclusion of firing strength allows it to differentiate the samples in terms of their input information, i.e., membership values. Intuitively, the discrepancy of samples will become more notable when more input information is fed into the layer. In the fuzzy classification of gear faults, the fuzzy reasoning provides a form of many-valued logic, yielding the probabilities of different observed fault types that range from 0 to 1. The fourth layer is the defuzzification layer. Corresponding to the fuzzification layer, i.e., the second layer, the fourth layer will produce a quantifiable result in crisp logic, given fuzzy sets and corresponding membership values.

$$g_k^{(4)} = \phi\left(\sum_{j=1}^{q} \omega_{j,k} g_j^{(3)} + b_k\right) \quad k = 1, 2, \ldots, m \tag{4}$$

In training, the model using available dataset and known labels, the defuzzification layer is employed to yield deterministic results. On the other hand, in prediction and actual diagnosis, since the network will deal with unseen fault scenarios where our goal is not to produce crisp outcome of classification, the weight matrix of this layer will be replaced with an identity matrix. In other words, the defuzzification step is employed in training the model, but not involved for fuzzy classification.

Subsequently, the normalization function is performed through the Softmax layer, i.e., the fifth layer, in order to facilitate the training with numerical stability. The output of Softmax layer is the probability value of the target sample with respect to the specific known fault classes/labels, which is denoted as

$$g_k^{(5)} = \frac{g_k^{(4)}}{\sum_{i=1}^{m} g_i^{(4)}} \quad k = 1, 2, \ldots, m \tag{5}$$

where m is the number of outputs. The last layer is the final output layer that makes the decision based on the normalized probability values from the Softmax layer. The criterion for decision-making is simply expressed as

$$y = \operatorname{argmax}\left(\left[g_1^{(5)}, \ldots, g_k^{(5)}, \ldots, g_m^{(5)}\right]\right) \tag{6}$$

One may notice that the scale of the FNN model is highly dependent on the hyperparameters n, q, and m mentioned above. Once all those hyperparameters are finalized, the total number of unknown parameters to be optimized can be estimated through the equation below,

$$Q = n \times q \times l(\boldsymbol{\beta}) + (q^n + 1) \times m \tag{7}$$

where $l(\boldsymbol{\beta})$ denotes the number of parameters in the membership function. For example, Gaussian membership function usually is modeled by two parameters, i.e., the center and width. The first part at the right hand side of Eq. (7) indicates the number of unknowns in the fuzzification layer, while the second part represents the number of weights and biases in the defuzzification layer. The backpropagation optimization scheme is usually adopted to identify the best unknown parameters in associated layers [17]. In this research, the number of outputs, i.e., m, is the number of fault labels involved in training. n is the number of representative feature components extracted via KPCA analysis to be presented in the succeeding subsection, which is a key hyperparameter to be investigated. All the hyperparameters will be optimally determined to ensure both the computational accuracy and efficiency of FNN.

One key idea in this proposed FNN is that the training using existing labels and the actual classification will be conducted differently. For training, since all data samples are known to belong to respective classes/labels, all six layers of the FNN will be involved, aiming at accurately classifying all the data samples with respect to known classes/labels. For actual classification, since we expect the neural network to be able to deal with unseen scenarios and compare them with existing classes/labels to provide probabilistic assessment of similarity, only the first five layers will be utilized as the fifth layer is capable of providing the probabilistic assessment as outputs for diagnosis. Also as mentioned, the treatment in defuzzification layer (i.e., the fourth layer) is different in training and actual classification/diagnosis. Specifically, an additional fuzzification step will be employed in actual classification. This strategy can take full advantage of the versatility and fuzzy logic functionality of the FNN concept.

2.2 Kernel principal component analysis for data pre-processing and feature extraction

Although for the sake of simplicity it is tempting to directly utilize raw vibration signals for classification, owing to the high sampling frequency, these signals are generally of high dimension, requiring an FNN model with very high dimension as well which may pose a computational challenge. Therefore, feature extraction with dimension reduction appears to be a natural pre-processing step. Moreover, in order to take advantage of its fuzzy logic nature, the inputs to the FNN will need to be the key features of the signals rather than the raw time-domain responses, because it is much more pertinent to apply the fuzzification procedure, i.e., the second layer in FNN as indicated in the preceding section, to a small number of features extracted. As mentioned in Section 1, various feature extraction approaches exist for gear diagnosis, including frequency analysis, joint time-frequency analysis, and principal component analysis (PCA) techniques. It is worth noting that selecting feature extraction approaches for gear diagnosis remains to be an open and on-going research topic. In this research, we propose to incorporate kernel principal component analysis (KPCA) for pre-processing, owing to its capability of facilitating nonlinear dimensionality reduction and feature correlation.

Principal component analysis (PCA) is a statistical approach that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables, referred to as the principal components (PCs). The principal components are ranked in terms of feature variances among all data samples [20]. Through discarding the insignificant PCs, PCA can drastically reduce the data dimensionality in an interpretable way, such that only the most important information in the data is preserved. While the standard or linear PCA is formulated for representing the features in linear space, the KPCA allows us to extract and correlate features in a nonlinear manner [12, 47, 48], which is important for fuzzy classification of unseen scenarios. Let $\mathbf{D}_i$ be the ith ($i = 1, \cdots, N$) gear vibration response sample with dimension u, we assume that the data with zero mean is operated, which yields

$$\frac{1}{N}\sum_{i=1}^{N}\varphi(\mathbf{D}_i) = 0 \tag{8}$$

$\varphi(.)$ is the nonlinear transformation operator. The covariance matrix of data $\mathbf{D}$ can be calculated as

$$\mathbf{C} = \frac{1}{N}\sum_{i=1}^{N}\varphi(\mathbf{D}_i)\varphi(\mathbf{D}_i)^T \tag{9}$$

Correspondingly, we have the following eigenvalue problem,

$$\mathbf{C}\mathbf{v}_k = \lambda_k\mathbf{v}_k \tag{10}$$

Based on Eqs. (8) to (10), we can obtain

$$\frac{1}{N}\sum_{i=1}^{N}\varphi(\mathbf{D}_i)\left\{\varphi(\mathbf{D}_i)^T\mathbf{v}_k\right\} = \lambda_k\mathbf{v}_k \tag{11}$$

Equation (11) can be re-written as

$$\mathbf{v}_k = \sum_{i=1}^{N} r_{ki}\varphi(\mathbf{D}_i) \tag{12}$$

where $r_{ki} = \frac{\varphi(\mathbf{D}_i)^T\mathbf{v}_k}{N\lambda_K}$. Substituting Eq. (12) into (11) yields

$$\frac{1}{N}\sum_{i=1}^{N}\varphi(\mathbf{D}_i)\varphi(\mathbf{D}_i)^T\sum_{j=1}^{N} r_{kj}\varphi(\mathbf{D}_j) = \lambda_k\sum_{i=1}^{N} r_{ki}\varphi(\mathbf{D}_i) \tag{13}$$

Without loss of generality, the kernel function can be expressed as

$$\kappa(\mathbf{D}_i, \mathbf{D}_j) = \varphi(\mathbf{D}_i)^T \varphi(\mathbf{D}_j) \tag{14}$$

Multiplying both sides of Eq. (14) by $\varphi(\mathbf{D}_z)^T$ yields

$$\frac{1}{N}\sum_{i=1}^{N}\kappa(\mathbf{D}_z, \mathbf{D}_i)\sum_{j=1}^{N} r_{kj}\kappa(\mathbf{D}_i, \mathbf{D}_j) = \lambda_k \sum_{i=1}^{N} r_{ki}\kappa(\mathbf{D}_z, \mathbf{D}_i) \tag{15}$$

We can further re-write the above equation using matrix notation,

$$\mathbf{K}\mathbf{r}_k = \lambda_k N \mathbf{r}_k \tag{16}$$

where $\mathbf{r}_k = [r_{k1}, r_{k2} \ldots, r_{kN}]^T$ and $K_{i,\,j} = \kappa(\mathbf{D}_i, \mathbf{D}_j)$. Apparently, $\mathbf{r}_k$ and λ_k can be solved through an eigenvalue problem. The kernel principal components eventually can be obtained as

$$\mathbf{y}_k(\mathbf{D}) = \sum_{i=1}^{N} r_{ki}\kappa(\mathbf{D}, \mathbf{D}_i) \tag{17}$$

The PCs with larger eigenvalues λ_k are more dominant. By retaining the most dominant l PCs ($l \ll u$), feature extraction and dimension reduction can be achieved. In this research, these features will be used as inputs to the FNN.

Care must be taken regarding the fact that, regardless $\mathbf{D}$ has zero mean in its original space, it is not guaranteed to be centered. We can use the Gram matrix $\mathbf{K}'$ to substitute the kernel matrix $\mathbf{K}$ [3, 4]. The Gram matrix is given as

$$\mathbf{K}' = \mathbf{K} - \mathbf{1}_N\mathbf{K} - \mathbf{K}\mathbf{1}_N + \mathbf{1}_N\mathbf{K}\mathbf{1}_N \tag{18}$$

where $\mathbf{1}_N$ is the $N \times N$ matrix with all elements equal to $1/N$ [3]. An important step in KPCA is to define the structure of kernel $\kappa(\mathbf{D}_i, \mathbf{D}_j)$ (Eq. (14)). There exist several common choices, including Gaussian kernel (Eq. (19a)), polynomial kernel (Eq. (19b)), and sigmoid kernel (Eq. (19c)).

$$\kappa(\mathbf{z}_1, \mathbf{z}_2) = \exp\left(-\|\mathbf{z}_1 - \mathbf{z}_2\|^2 / 2\sigma^2\right) \tag{19a}$$

$$\kappa(\mathbf{z}_1, \mathbf{z}_2) = \left(\mathbf{z}_1^T\mathbf{z}_2 + a\right)^b \tag{19b}$$

$$\kappa(\mathbf{z}_1, \mathbf{z}_2) = \tanh\left(\alpha\mathbf{z}_1^T\mathbf{z}_2\right) \tag{19c}$$

$\kappa(\mathbf{z}_1, \mathbf{z}_2) = \left(\mathbf{z}_1^T\mathbf{z}_2 + a\right)^b$ As will be shown in the subsequent section, using KPCA and selecting Gaussian kernel can effectively lead to feature extraction and correlation for the proposed FNN. Fundamentally, KPCA is a generic, nonlinear feature extraction approach that is capable of exploiting the complicated spatial structure of high-dimensional features [38]. It is particularly suitable to fuzzy analysis of gear signals.

Integrating KPCA into FNN yields the fuzzy fault diagnosis framework as illustrated in Fig. 3.

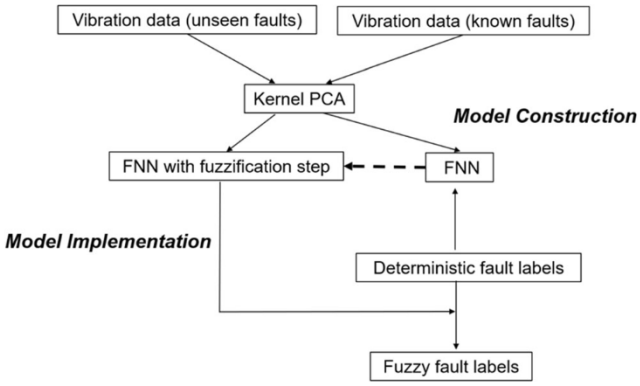

Fig. 3 Proposed fuzzy fault diagnosis framework

3 Fuzzy fault diagnosis practice with experimental investigation

In this section, the proposed fuzzy classification scheme is implemented to the fault diagnosis of a lab-scale gearbox system using experimental vibration measurement. Our focus is on how the FNN can handle limited labels in classification. The details of experimental data acquisition and pre-processing, specific FNN setup and tuning, and result discussion are presented as follows.

3.1 Experimental data acquisition and fault labeling

While many types of gear faults can occur, vibration responses measured are commonly used as information carriers that reflect the health status of a gearbox system. In this study, we employ a lab-scale two-stage gearbox system with replaceable gears shown in Fig. 4, upon which the experimental data is directly acquired for the subsequent fault diagnosis. The gear speed is controlled by a motor. The torque is supplied through a magnetic brake which can be adjusted by changing its input voltage. A 32-tooth pinion and an 80-tooth gear are installed on the first-stage input shaft. The second stage consists of a 48-tooth pinion and a 64-tooth gear. The input shaft speed and gear vibration signals are measured by a tachometer and an accelerometer, respectively. The signals are recorded through a dSPACE system (DS1006 processor board, dSPACE Inc.) with sampling frequency 20 KHz. The vibratory responses of a system involving gear mechanism are angle-periodic. In reality, while the gearbox system is recorded in a fixed sampling rate, the time-domain responses are generally not time-periodic due to speed variations under uncertainty. In this research, we apply the time synchronous averaging (TSA) approach to solve the nonstationary issue and minimize the effect of uncertainty caused by speed varying, where the time-even signals are resampled based on the shaft speed measured by the tachometer and averaged in angular domain

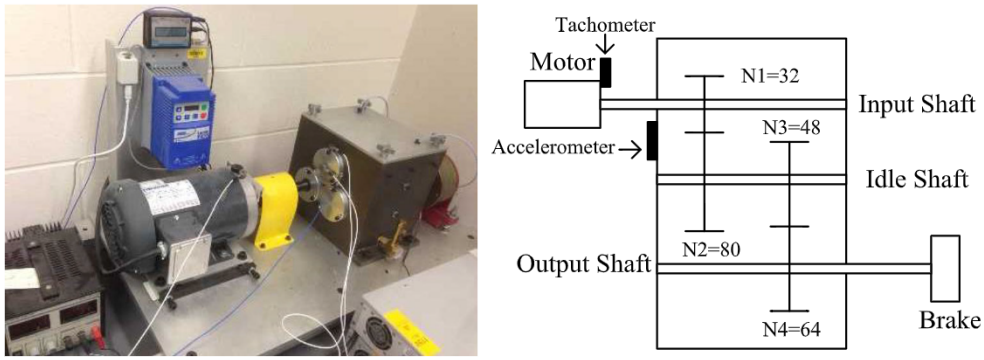

Fig. 4 Gearbox setup for data acquisition

[48]. As TSA converts the signals from the time-even to the angle-even representation, it can significantly reduce the non-coherent components in the system response.

In this research, 9 different gear conditions are introduced into the pinion on the input shaft, including the healthy condition, missing tooth, root crack, spalling, and chipping tip with 5 different severities/levels (to be detailed in Section 3.3), as shown in Fig. 5. Altogether, we have 9 classes or labels in the overall data collected. For each gear condition, 104 signals are collected. For each signal, 3600 angle-even data points are recorded in the course of 4 gear revolutions. Hence, we have a total of 936 (104 × 9) samples corresponding to 9 gear conditions. Each sample can be considered one data point containing 3600 raw features. The balance of data samples under different gear conditions is the prerequisite for ensuring the rationality of general classification analysis. All the data used in this study is made public at https://figshare.com/articles/Gear_Fault_Data/6127874/1. Since our focus in this research is the development of FNN to handle unseen fault scenarios, in the subsequent case demonstrations, we will use 8 gear conditions/classes in FNN training and examine the fuzzy classification results of the FNN using the remaining hold-out data as one unseen fault condition.

3.2 FNN configuration with hyperparameter setup

The core of this research is a specifically designed FNN to facilitate the training using existing labels and then the prediction that is capable of handling unseen scenarios. In what follows, we present the FNN configuration details for gear diagnosis.

Input layer According to the experimental setup, each input sample includes 3600 raw features, i.e., time series responses. Incorporating all these features into the neural network thus would require 3600 input nodes (i.e., $n = 3600$ in Eq. (1)), which makes it computationally costly to develop an FNN model. Furthermore, it is infeasible to formulate the subsequent fuzzification procedure with so many raw features. For the sake of feature extraction and dimension reduction, we carry out KPCA and employ the Gaussian kernel (Eq. (19a)). Hyperparameter n thus becomes the number of principal components (PCs). The best value of n will be determined through the FNN training convergence analysis to be shown in Section 3.3.

Fuzzification layer In this study, we choose the Gaussian membership function to describe the fuzzy sets. For illustration, the single cluster is used, which allows each input feature to be

Fig. 5 Nine fault types on pinions. For chipping tip, 5 different severities are included

Table 1 Fuzzy logic rules employed in FNN

Rule ID	Fuzzification layer #2/antecedent	Fuzzy reasoning layer #3/consequent
1	If (x_1 is M_{11}) AND (x_2 is M_{21}) AND … AND (x_n is M_{n1})	$\gamma_1 = \prod_{i=1}^{n} \nu_{i,1}$
2	If (x_1 is M_{12}) AND (x_2 is M_{22}) AND … AND (x_n is M_{n2})	$\gamma_2 = \prod_{i=1}^{n} \nu_{i,2}$
3	If (x_1 is M_{13}) AND (x_2 is M_{23}) AND … AND (x_n is M_{n3})	$\gamma_3 = \prod_{i=1}^{n} \nu_{i,3}$
……	……	……
m	If (x_1 is M_{1m}) AND (x_2 is M_{2m}) AND … AND (x_n is M_{nm})	$\gamma_m = \prod_{i=1}^{n} \nu_{i,m}$

processed by q numbers of membership functions (Eq. (2)). As such, the total number of membership functions in this layer is $n \times q$. To facilitate the subsequent fuzzy rule definition, here, we let $q = m$. As will be shown later, m is the number of observed/known fault classes employed in training (i.e., 8).

Fuzzy reasoning/rule layer Theoretically, the total number of fuzzy rules to cover all scenarios is m^n or q^n (when $q = m$). For methodology demonstration, here, we assume that the rules are subject to self-correlation, which reduces the rule number from m^n to m. The details of the rules employed are tabulated in Table 1. Such rule definition has the following merits: (1) each rule solely affects the probability output of the related fault type, which enables simple interpretation; and (2) the model stays as small-scale by only taking into account a small number of nodes, i.e., m in this layer.

Defuzzification layer The total number of nodes in this layer also is m. This layer performs like a fully connected layer. Given the m nodes in the preceding layer, there are $m \times m$ weights and m biases to be optimized. ReLU activation function is applied onto the nodes to realize the space mapping. As mentioned in Section 2.1, in the classification/prediction process, since our purpose is to obtain probabilistic assessment rather than crisp decision for unseen scenarios, we can just replace the optimized weight matrix ($m \times m$) and optimized bias vector ($m \times 1$) of defuzzification layer with a unit diagonal matrix and a zero vector.

Softmax layer m nodes produced in this layer indicate the probability values of the actual data sample with respect to the observed/labeled fault types. This yields the outcome of the FNN when used for diagnosis purposes.

Output layer Single node indicates the result of decision-making based on Eq. (6). This yields the outcome of FNN when used in training, where all data samples belong to certain known labels/scenarios.

M denotes the membership function; ν is the degree of membership through interpolation (same concept with $g^{(2)}$ shown in Eq. (2)); γ is the firing strength (same concept with $g^{(3)}$ shown in Eq. (3)).

3.3 FNN implementation and case demonstrations

In this section, we present FNN implementation details and demonstrate the fuzzy classification performance. As mentioned, 9 classes of gear conditions have been recorded. Among the gear faults, chipping tip is essentially a fault pattern with continuous severity, i.e., different chipping levels. It is infeasible to collect many severity levels in actual practice. In our data acquisition, we have obtained 5 different severity levels. In order to examine the fuzzy classification performance, in case demonstration, we purposely hold out one of the chipping tip class in network training. That is, we assume that only 4 chipping tip severities are available as

Table 2 Data split for training and testing of fuzzy fault diagnosis (case 1)

ID	Fault types	FNN output	Data size	FNN roles
1	Healthy	Labeled	104	Training
2	Missing tooth	Labeled	104	Training
3	Crack	Labeled	104	Training
4	Spalling	Labeled	104	Training
5	Chipping_tip_5 (least severe)	Labeled	104	Training
6	Chipping_tip_4	Labeled	104	Training
7	Chipping_tip_3	Labeled	104	Training
8	Chipping_tip_2	Unknown	104	Testing/validation
9	Chipping_tip_1 (most severe)	Labeled	104	Training

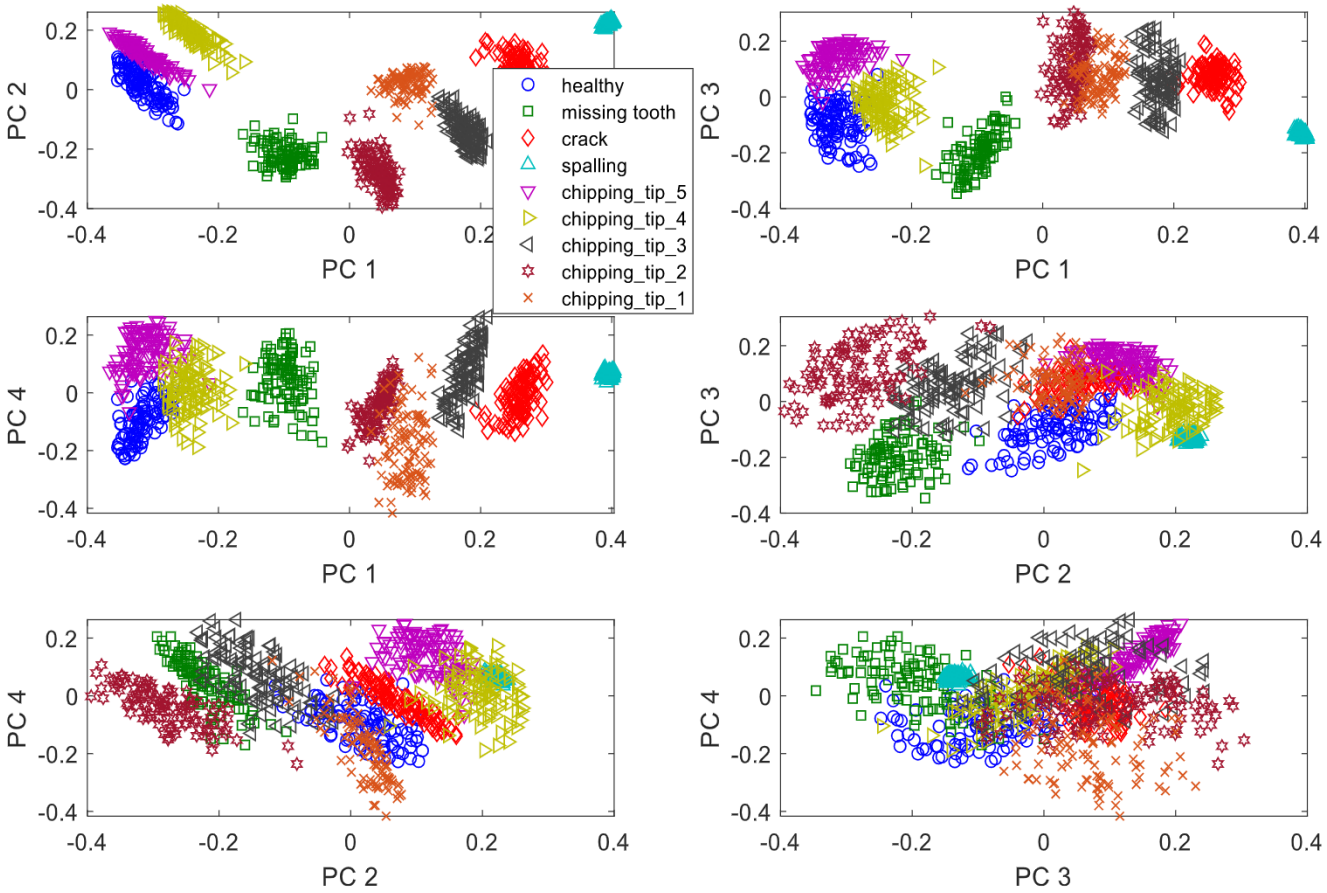

Fig. 6 Gaussian kernel PCs of gear vibration data

known labels to train the FNN, and we use the remaining chipping tip data to evaluate the FNN performance in terms of handling unseen fault scenarios. We analyze two different cases, i.e., holding out a severe chipping tip case, and holding out a weak chipping tip case.

3.3.1 Fuzzy classification of a severe chipping tip fault as the unseen fault scenario (case 1)

Altogether we have acquired 9 gear conditions, each with 104 vibration signal samples, yielding a total of 936 samples. The

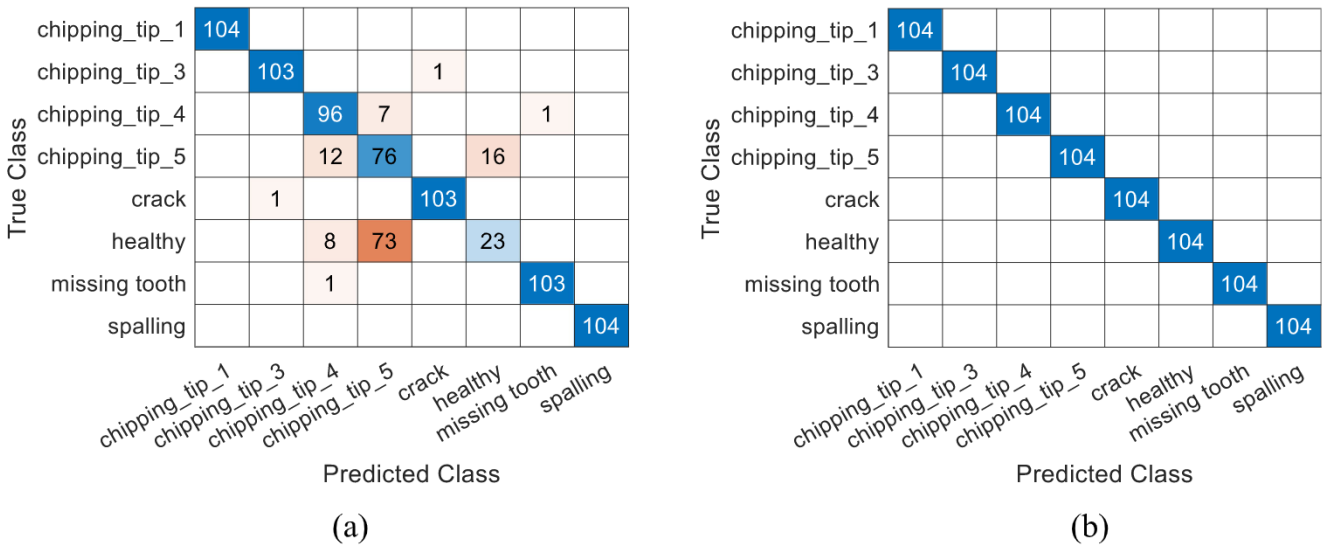

Fig. 7 Training confusion matrices under different numbers of PCs (case 1)

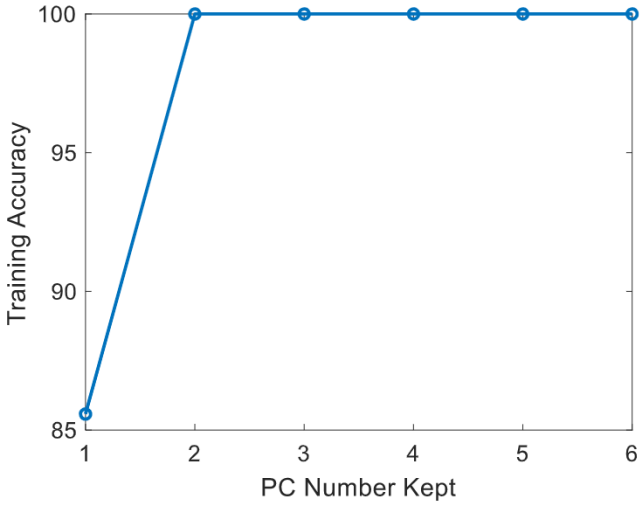

Fig. 8 FNN training accuracy curve versus PC number (case 1)

details of training and testing data split are given in Table 2. The module of the gear is 1.59 mm. In this table, chipping_tip_5 to chipping_tip_1 refer to 5 different levels of tip removal, i.e., 0.15 mm, 024 mm, 0.38 mm, 0.48 mm, and 0.69mm, respectively. Here, chipping_tip_5 represents the weakest level of chipping tip fault. In this first case demonstration, we hold out chipping_tip_2 with 0.48-mm tip removal, and use the other 8 labels/classes in FNN training. Therefore, 832 (104×8) samples corresponding to 8 fault types will be utilized to train the FNN model. Afterwards, we will examine the classification results. If the data of chipping_tip_2, which is unseen by the FNN in training, is classified as being close to chipping_tip_1 or chipping_tip_3, then the FNN can successfully fulfill the diagnosis mission.

Selection of FNN hyperparameter As pointed out in Section 3.2., the hyperparameters of FNN need to be optimally selected to ensure computational performance. To facilitate efficient training and classification with sufficient accuracy, the scale of FNN model will need to be properly decided. In this research, one hyperparameter to be identified is the number of PCs (i.e., n). The 4 lowest-order Gaussian kernel PCs calculated from the original gear vibration signals are shown in Fig. 6. The distribution of PCs over the entire collection of samples shows good differentiation ability of different fault types especially for the lower-order PCs. As the order of PC increases, the fault boundaries start to become unclear. In terms of computational efficiency, increasing the number of PCs will increase the dimension of FNN as the numbers of nodes in the subsequent layers will grow accordingly. In terms of computational accuracy, too small or too large numbers of PCs will cause the underfitting and overfitting issues, respectively. To facilitate the hyperparameter selection, a strategy based upon the FNN training convergence analysis is established as follows.

The FNN development and analysis are carried out using an in-house code, where the MATLAB Fuzzy Logic Toolbox is used to facilitate some basic operations. A total of 100 epoch size is adopted for training, which takes less than 1 min on a desktop with Intel CPU E5-2640 @2.40GHz (2 processors). The training accuracy of this classification analysis is reflected by the confusion matrix [31]. We employ a series of the numbers of principal components, i.e., n in an ascending order from 1 to 6 to examine the training accuracy tendency. For illustration, the confusion matrices obtained with $n = 1$ and $n = 2$, respectively, are given in Fig. 7a and 7b. Underfitting is clearly observed from Fig. 7a where n is selected as 1 because of large training error. Fundamentally, such underfitting is due to the lack of useful features introduced. On the contrary, when n increases to 2, as shown in Fig. 7b, 100% accuracy can be achieved. As the confusion matrix provides the comprehensive information, multiple metrics such as precision, recall, F1 score, and ACC (accuracy), can be further employed to quantify the training accuracy [31]. Here, we specifically use ACC. The training accuracy (i.e., ACC) tendency with respect to the number of PCs is obtained as shown in Fig. 8. Based on the convergence results, in this

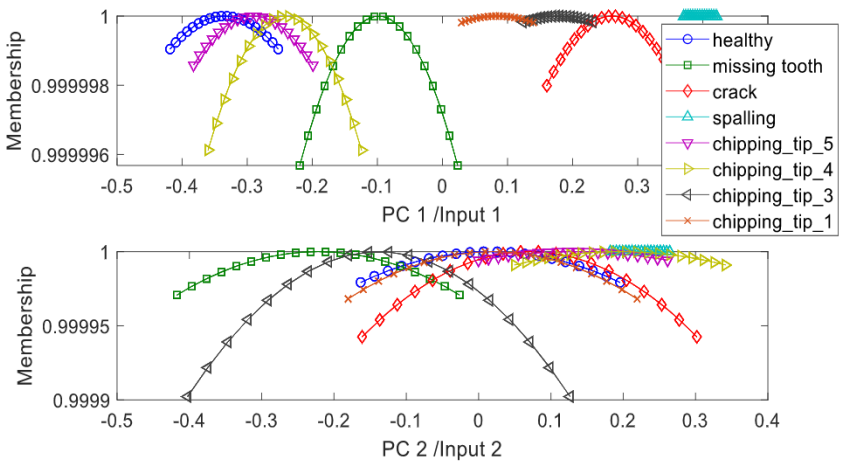

Fig. 9 Optimized Gaussian membership function (case 1)

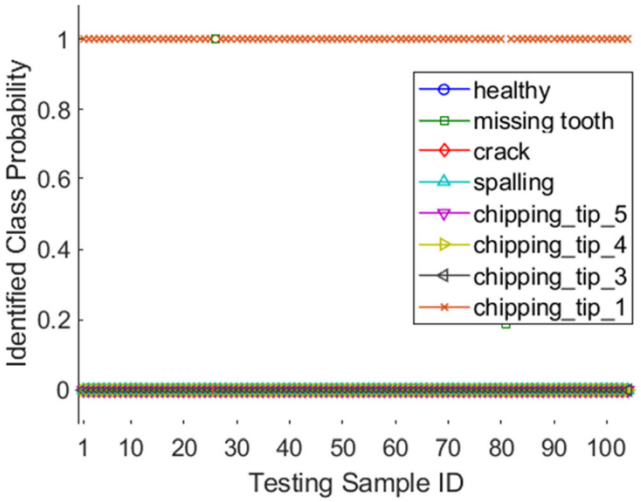

Fig. 10 Probability distribution of 104 testing samples through FNN classification (case 1)

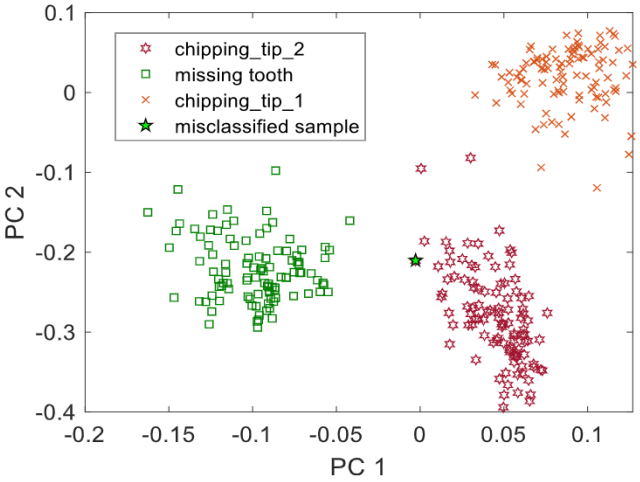

Fig. 12 Feature location of misclassified sample (case 1)

case, we select $n = 2$ for the subsequent fuzzy classification analysis. Interestingly, this shows the capability of KPCA in efficient feature extraction for this particular gearbox system.

FNN performance analysis and result demonstration For FNN training purposes, the original samples are transformed into the reduced-dimensional feature space. In total, 832 training samples of PCs are shuffled and fed into the FNN model for training that optimizes the unknown parameters in fuzzification and defuzzification layers. The unknown parameters in the fuzzification layer essentially refer to the parameters that characterize the membership functions. The membership functions are the core of this proposed methodology, allowing one to graphically represent the fuzzy sets. As a result, the membership functions can represent the physical behavior of the target system. In this study, the Gaussian membership functions are optimized and shown in Fig. 9. Eight membership functions are used to represent 8 observed/known fault types. The degree of membership can be interpolated via the associated PC value in the horizontal axis. It is found that the membership functions of the 1st PC have narrower width (i.e., variance) than that of the 2nd PC, because the lower-order PC is more discriminative as indicated in Fig. 6. In other words, the membership functions of the 1st PC indicate higher level of fuzziness than that of the 2nd PC.

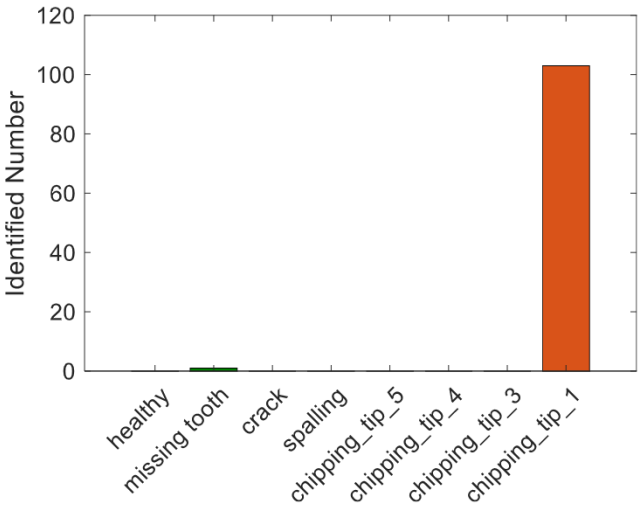

Fig. 11 Fuzzy classification accuracy (case 1)

After the FNN model is well-trained, a fuzzification step is incorporated, which takes advantage of the necessary optimized model parameters retained for testing/prediction. We now examine the performance of FNN in terms of unseen fault scenarios by testing the hold-out data of chipping_tip_2. As we have 8 known labels, the distribution of probability values of the testing dataset with respect to 8 labels over the entire testing space is shown in Fig. 10. Clearly, the majority of testing samples with actual fault, i.e., chipping_tip_2, are identified as being close to fault chipping_tip_1. This indicates that the FNN is capable of providing the closest fault classification with respect to known labels. The probability values appear to be well-separated, which shows the high confidence level of fuzzy classification result.

A fault scenario with the highest probability value will be assigned to the relevant testing sample as the most probable fault (Fig. 10). Counting the numbers of different fault types that are probabilistically identified over the entire testing space yields the classification accuracy result (Fig. 11). In this case study, if a testing sample belonging to chipping_tip_2 is classified either as chipping_tip_3 or as chipping_tip_1, i.e., the neighboring class, the classification is correct. Obviously, the result in Fig. 11 indicates very high fuzzy classification accuracy, i.e., 99% (103/104). Only 1 sample is misclassified as the missing tooth fault class.

Error investigation As can be observed from the above analysis, there exists one sample that is misclassified as

Table 3 Data split for training and testing of fuzzy fault diagnosis (case 2)

ID	Fault types	FNN output	Data size	FNN roles
1	Healthy	Labeled	104	Training
2	Missing tooth	Labeled	104	Training
3	Crack	Labeled	104	Training
4	Spalling	Labeled	104	Training
5	Chipping_tip_5 (least severe)	Labeled	104	Training
6	Chipping_tip_4	Unknown	104	Testing/validation
7	Chipping_tip_3	Labeled	104	Training
8	Chipping_tip_2	Labeled	104	Training
9	Chipping_tip_1 (most severe)	Labeled	104	Training

belonging to missing tooth fault class. Identifying the root cause of this misclassification can shed light on how features are extracted and how to minimize misclassification in future implementation. Taking advantage of KPCA analysis, we take a closer look at the PC distribution, i.e., only covering the fault classes of missing tooth, chipping_tip_1, and chipping_tip_2. As shown in Fig. 12, the misclassified sample is highlighted as a solid star. It can be observed that the feature of this sample is situated at the left boundary of the chipping_tip_2 feature cluster, which is closest to the missing tooth cluster. Considering that only 1 out of 104 samples exhibits such feature ambiguity, we can conclude that KPCA with 2 PCs is indeed effective in this particular implementation.

3.3.2 Fuzzy classification of a weak chipping tip fault as the unseen fault scenario (case 2)

To further validate the FNN, we formulate another case and revisit the FNN modeling and analysis for performance re-examination. In this second case, we hold out the data samples belonging to chipping_tip_4. The data of chipping_tip_4 corresponds to 0.24-mm tip removal, which is a weak chipping tip fault. The details of training and testing data split in this second case are given in Table 3. Similarly, Gaussian kernel KPCA is employed for nonlinear feature reduction and transformation. We first implement the training convergence analysis to optimally determine the number of PCs, which again is 2 in this case. The same configuration of FNN model and associated parameters defined in Section 3.2 is utilized here. Model training leads to the optimized membership functions shown in Fig. 13. Apparently, the observation is consistent with that of Fig. 9.

We then carry out testing of FNN training using the hold-out data. In this case, the classification is considered correct if a testing sample is classified as being close to chipping_tip_5 or chipping_tip_3. In testing, 104 samples are fed into the FNN model to produce the probability distribution and the final classification result, shown in Figs. 14 and 15 respectively. We can immediately observe that the majority of testing samples are classified as being close to chipping_tip_5. The classification accuracy is calculated as 98%; i.e., 102 out of 104 samples are correctly classified. Error analysis is carried out for the 2

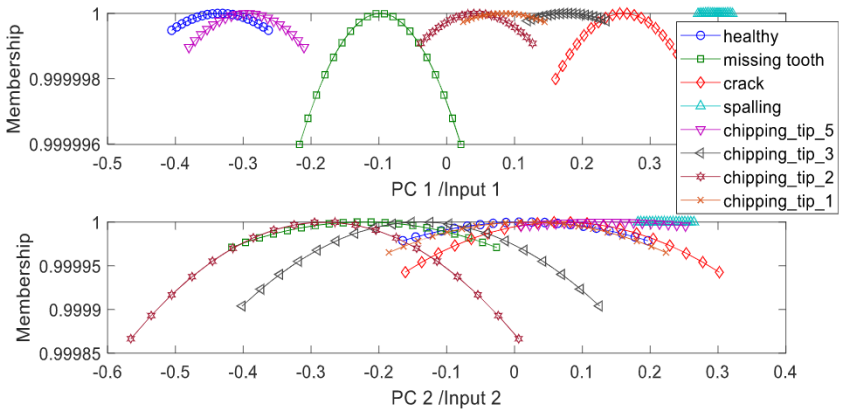

Fig. 13 Optimized Gaussian membership function (case 2)

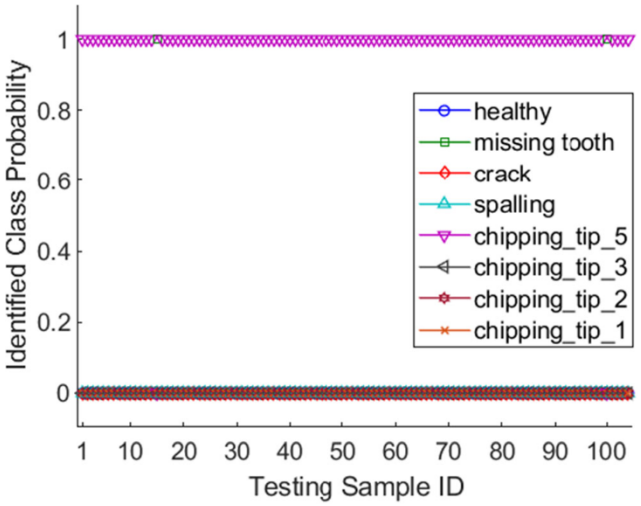

Fig. 14 Probability distribution of 104 testing samples through FNN classification (case 2)

misclassified samples. As shown in Fig. 16, the features of these 2 samples are situated at the bottom boundary of the feature cluster of chipping_tip_4, which is very close to the feature cluster of the missing tooth fault. These results again demonstrate the effectiveness of the proposed FNN to handle unseen fault scenarios.

4 Conclusion

The lack of sufficient data labels, i.e., fault types, in experimental data collected poses a major challenge for the practical implementation of gear fault diagnosis. The state-of-the-art machine learning techniques generally only allow the identification of gear faults within the known fault labels. To tackle this challenge, a new fuzzy classification method is

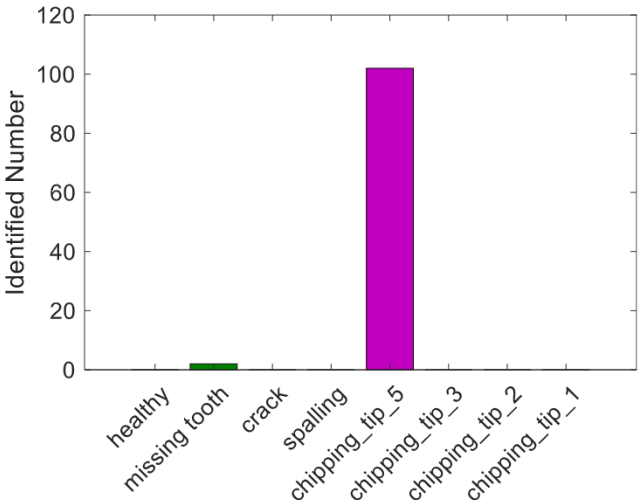

Fig. 15 Fuzzy classification accuracy (case 2)

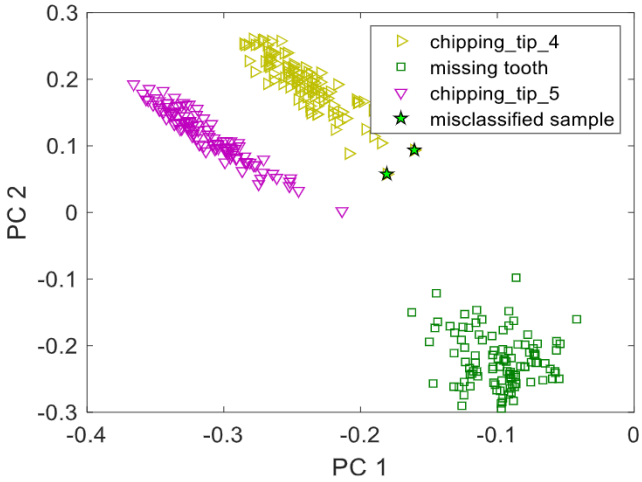

Fig. 16 Feature location of misclassified sample (case 2)

established in this research to deal with the gear fault diagnosis with limited data labels. A fuzzification procedure through tuning the architecture of the well-trained fuzzy neural network is adopted to achieve the classification of unseen fault scenarios that are not included in the training dataset. Particularly, those unseen fault scenarios will be classified based on their closeness with respect to known fault classes with probability. The vibration data acquired from a lab-scale gearbox system is preprocessed by means of the kernel principal component analysis (KPCA) in order to reduce the feature dimensionality and meanwhile capture the primary nonlinear features within data. This indeed facilitates the FNN model training and the subsequent classification analysis. The number of kernel principal components is optimally identified through FNN model training convergence analysis. Two cases with different levels of fault severities are carried out to examine the performance of this enhanced FNN in terms of dealing with unseen fault scenarios. In both cases, the FNN can successfully classify the unseen data as being close to the neighboring fault scenarios with 99% and 98% accuracy, respectively, thereby fulfilling the diagnosis purpose in actual practice.

Author contribution K Zhou and J Tang worked together to generate the conception of the work. K Zhou carried out algorithm development and data analysis and interpretation, and drafted the paper. J Tang provided advisement to K Zhou, and also provided critical revision of the paper.

Funding This research is supported by the National Science Foundation under grant IIS-1741174.

Data availability Not applicable.

Code availability Not applicable.

Declarations

Ethical approval Not applicable.

Consent to participate Not applicable.

Consent for publication Not applicable.

Competing interests The authors declare no competing interests.

References

1. Ahuja AS, Ramteke DS, Parey A (2020) Vibration-based fault diagnosis of a bevel and spur gearbox using continuous wavelet transform and adaptive neural-fuzzy inference system. Adv Intelligent Syst Comp 1096:473–496
2. Amarouayache IIE, Saadi MN, Guersi N, Boutasseta N (2020) Bearing fault diagnostics using EEMD processing and convolutional neural network methods. Int J Adv Manuf Technol 107(9-10):4077–4095
3. Bishop CM (2006) *Pattern recognition and machine learning*. *Springer*, Cambridge
4. Boyd S, Vandenberghe L (2018) 2018, Introduction to applied linear algebra: vectors, matrices, and least squares. Cambridge University Press, Cambridge
5. Buzzoni M, Antoni J, D'Elia G (2018) Blind deconvolution based on cyclostationarity maximization and its application to fault identification. J Sound Vib 432:569–601
6. Cao P, Zhang S, Tang J (2018) Pre-processing-free gear fault diagnosis using small datasets with deep convolutional neural network-based transfer learning. IEEE Access 6:26241–26253
7. Chen G, Zhang W (2018) Comprehensive evaluation method for performance of unmanned robot applied to automotive test using fuzzy logic and evidence theory. Comput Ind 98:48–55
8. Chen Z, Zhang J, Zhai W, Wang Y, Liu J (2017) Improved analytical methods for calculation of gear tooth fillet-foundation stiffness with tooth root crack. Eng Fail Anal 82:72–81
9. Chen J, Zhou D, Guo Z, Lin J, Lyu C, Lu C (2019) An active learning method based on uncertainty and complexity for gearbox fault diagnosis. IEEE Access 7:9022–9031
10. Cheng G, Cheng YL, Shen LH, Qiu JB, Zhang S (2013) Gear fault identification based on Hilbert-Huang transform and SOM neural network. Measurement: Journal of the International Measurement Confederation 46(3):1137–1146
11. Chen C, Shen F, Xu J, Yan R (2021) Domain adaptation-based transfer learning for gear fault diagnosis under varying working conditions. IEEE Trans Instrum Meas 70:9146579
12. Deng X, Tian X (2013) Nonlinear process fault pattern recognition using statistics kernel PCA similarity factor. Neurocomputing 121: 298–308
13. Dibaj A, Ettefagh MM, Hassannejad R, and Ehghaghi MB, (2019), "Fine-tuned variational mode decomposition for fault diagnosis of rotary machinery," Structural Health Monitoring, in press.
14. Fakhfakh T, Chaari F, Haddar M (2005) Numerical and experimental analysis of a gear system with teeth defects. Int J Adv Manuf Technol 25(5-6):542–550
15. Fuller R, 2013, *Introduction to neural-fuzzy systems*, Springer Sci Bus Media, 2013.
16. Gao B, He Y, Woo WL, Tian GY, Liu J, Hu Y (2016) Multidimensional tensor-based inductive thermography with multiple physical fields for offshore wind turbine gear inspection. IEEE Trans Ind Electron 63(10):6305–6315
17. Haykin S, (1998) *Neural networks: a comprehensive foundation (2nd edition)*, Prentice Hall.
18. He Z, Shao H, Wang P, Lin JJ, Cheng J, Yang Y (2020) Deep transfer multi-wavelet auto-encoder for intelligent fault diagnosis of gearbox with few target training samples. Knowl-Based Syst 191:105313
19. Jing L, Zhao M, Li P, Xu X (2017) A convolutional neural network based feature learning and fault diagnosis method for the condition monitoring of gearbox. Measurement: Journal of the International Measurement Confederation 111:1–10
20. Jolliffe IT, (2002) *Principal component analysis*, Springer Science & Business Media.
21. Li Y, Cheng G, Liu C, Chen X (2018) Study on planetary gear fault diagnosis based on variational mode decomposition and deep neural networks. Measurement: Journal of the International Measurement Confederation 130:94–104
22. Li Y, Feng K, Liang X, Zou MJ (2019a) A fault diagnosis method for planetary gearboxes under non-stationary working conditions using improved vold-Karlman filter and multi-scale sample entropy. J Sound Vib 439:271–286
23. Li F, Pang X, Yang Z (2019b) Motor current signal analysis using deep neural network for planetary gear fault diagnosis. Measurement: Journal of the International Measurement Confederation 145:45–54
24. Li Z, Yan X, Yuan C, Peng Z, Li L (2011) Virtual prototype and experimental research on gear multi-fault diagnosis using wavelet-autoregressive model and principal component analysis method. Mech Syst Signal Process 25(7):2589–2607
25. Liu PY, and Li HX, (2004) *Fuzzy neural network theory and application*, World Sci.
26. Moosavian A, Khazaee M, Ahmadi H, Khazaee M, Najafi G (2015) Fault diagnosis and classification of water pump using adaptive neuro-fuzzy inference system based on vibration signals. Struct Health Monit 14(5):402–410
27. Nguyen CD, Prosvirin A, Kim JM (2020) A reliable fault diagnosis method for a gearbox system with varying rotational speeds. Sensors 20(11):3105
28. Pandya Y, Parey A (2013) Failure path based modified gear mesh stiffness for spur gear pair with tooth root crack. Eng Fail Anal 27: 286–296
29. Randall RB, Antoni J, Chobsaard S (2001) The relationship between spectral correlation and envelope analysis in the diagnostics of bearing faults and other cyclostationary machine signals. Mech Syst Signal Process 15(5):945–962
30. Sadollah A, (2018) *Introductory chapter: which membership function is appropriate in fuzzy system?, Fuzzy logic based in optimization methods and control systems and its applications*, IntechOpen
31. Sammut C, and Webb GI, (2010) *Encyclopedia of Machine Learning and Data Mining*, Springer.
32. Saravanan N, Cholarirajan S, Ramachandram KI (2009) Vibration-based fault diagnosis of spur bevel gear box using fuzzy technique. Expert Syst Appl 36:3119–3135
33. Soualhi M, Nguyen KTP, Soualhi A, Medjaher K, Hemsas KE (2019) Health monitoring of bearing and gear faults by using a new health indicator extracted from current signals. Measurement: Journal of the International Measurement Confederation 141:37–51
34. Villa LF, Renones A, Peran JR, De Miguel LJ (2012) Statistical fault diagnosis based on vibration analysis for gear test-bench under non-stationary conditions of speed and load. Mech Syst Signal Process 29:436–446
35. Wang J, Ma Y, Zhang L, Gao RX, Wu D (2018a) Deep learning for smart manufacturing: methods and applications. J Manuf Syst 48: 144–156

36. Wang J, Yan J, Li C, Gao RX, Zhao R (2019b) Deep heterogeneous GRU model for predictive analytics in smart manufacturing: application to tool wear prediction. Comput Ind 111:1–14
37. Wang J, Gao RX, Yuan Z, Fan Z, Zhang L (2019a) A joint particle filter and expectation maximization approach to machine condition prognosis. J Intell Manuf 30(2):605–621
38. Wang Q, (2012), "Kernel principal component analysis and its applications in face recognition and active shape models," arXiv: 1207.3538 [cs.CV].
39. Wang S, Cai G, Zhu Z, Huang W, Zhang X (2015) Transient signal analysis based on Levenberg-Marquardt method for fault feature extraction of rotating machines. Mech Syst Signal Process 54:16–40
40. Wang Z, Wang J, Wang Y (2018b) An intelligent diagnosis scheme based on generative adversarial learning deep neural networks and its application to planetary gearbox fault pattern. Neurocomputing 310:213–222
41. Wu C, Jiang P, Ding C, Feng F, Chen T (2019) Intelligent fault diagnosis of rotating machinery based on one-dimensional convolutional neural network. Comput Ind 108:53–61
42. Yampikulsakul N, Byon E, Huang S, Sheng S, You M (2014) Condition monitoring of wind power system with nonparametric regression analysis. IEEE Trans Energy Convers 29(2):288–299
43. Yan R, Gao RX, Chen X (2014) Wavelets for fault diagnosis of rotary machines: a review with applications. Signal Process 96:1–15
44. Youcef Khodja A, Guersi N, Saadi MN, Boutasseta N (2020) Rolling element bearing fault diagnosis for rotating machinery using vibration spectrum imaging and convolutional neural networks. Int J Adv Manuf Technol 106(5-6):1737–1751
45. Zhang J, Qin W, Wu LH, Zhai WB (2014) Fuzzy neural network-based rescheduling decision mechanism for semiconductor manufacturing. Comput Ind 2014:1115–1125
46. Zhang Y, Lu W, Chu F (2017) Planet gear fault localization for wind turbine gearbox using acoustic emission signals. Renew Energy 109:449–460
47. Zhang C, Nie F, Xiang S (2010) A general kernelization framework for learning algorithms based on kernel PCA. Neurocomputing 73(4-6):959–967
48. Zhang S, Tang J (2018) Integrating angle-frequency domain synchronous averaging technique with feature extraction for gear fault diagnosis. Mech Syst Signal Process 90:711–729
49. Zhao R, Yan R, Chen Y, Mao K, Wang P, Gao RX (2019) Deep learning and its applications to machine health monitoring. Mech Syst Signal Process 115:213–237
50. Zhang Z, Wang Y, Wang K (2013) Intelligent fault diagnosis and prognosis approach for rotating machinery integrating wavelet transform, principal component analysis and artificial neural networks. Int J Adv Manuf Technol 68(1-4):763–773