An Integrated Class-Imbalanced Learning Scheme for Diagnosing Bearing Defects in Induction Motors

Roozbeh Razavi-Far, *Member, IEEE*, Maryam Farajzadeh-Zanjani, *Student Member, IEEE*, and Mehrdad Saif, *Senior Member, IEEE*

***Abstract*—This paper focuses on the development of an integrated scheme for diagnosing bearing defects in induction motors, under the class-imbalanced condition. This scheme comprises of four main modules: segmentation, feature extraction, feature reduction, and fault classification. Various state-of-the-art techniques have been devised in the feature extraction and reduction modules to extract informative sets of features from a raw vibration signal, filter redundant features, and produce the most distinct features for the following module. The fault classification module adapts various state-of-the-art class-imbalanced learning techniques for diagnosing bearing defects. This module contains a novel imputation-based oversampling technique for class-imbalanced learning. This integrated diagnostic scheme is evaluated on three experimental scenarios with different imbalance ratios. The reasonable diagnostic performances confirm the ability of the proposed novel class-imbalanced learning technique in diagnosing bearing defects, independently from the imbalance ratios.**

***Index Terms*—Bearing defects, Case Western Reserve University (CWRU), dimensionality reduction (DR), fault diagnosis, feature extraction (FE), feature selection (FS), imbalanced condition, induction motors (IMs).**

I. Introduction

INDUCTION motors (IMs) have been extensively used as a primary source of power in various industries. As a result of their simplicity and rigidity, IMs gained a key role in many industries such as power plants, aerospace, and petrochemical industries. However, components' aging and different type of defects can result in IM failure and even system breakdown, and thus, prompt diagnosis of IM defects is strategically essential for industries, not only to improve reliability, efficiency, and productivity of the systems, but also to minimize the maintenance cost [1], [2].

Manuscript received April 4, 2017; revised July 9, 2017 and August 30, 2017; accepted September 13, 2017. Date of publication September 21, 2017; date of current version December 1, 2017. This work was supported by the Natural Sciences and Engineering Research Council of Canada. Paper no. TII-17-0676. *(Corresponding author: Roozbeh Razavi-Far.)*

The authors are with the Department of Electrical and Computer Engineering, University of Windsor, Windsor ON N9B 3P4, Canada (e-mail: roozbeh@uwindsor.ca; farajza@uwindsor.ca; msaif@uwindsor.ca).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TII.2017.2755064

Various faults can occur in IMs that are typically categorized into two groups: mechanical damages (e.g., bearing defects, air gap eccentricity, and broken rotor bars) and electrical damages (e.g., phase-to-ground, phase-to-phase, and turn-to-turn connections in stator winding) [3]. Recent studies on the source of faults in IMs show that bearing defect, with almost 69% of failures distribution, is the most frequent fault in IMs compared to stator winding, rotor bars, and shaft/coupling with 21%, 7%, and 3% of failure distribution, respectively, [3]. Hence, diagnosing bearing defects, which is the focus of this paper, and performing prompt corrective actions can help preventing system failure and reducing the cost of unscheduled downtime.

Contemporary schemes to diagnose bearing defects in IMs are commonly based on the mathematical model of the system. However, performance of these diagnostic schemes highly depends on the accuracy of the model, which is not easy to obtain and subjected to inevitable assumptions and conditions. In this respect, data-driven model-free techniques have been extensively used for diagnosing faults in IMs in recent years [4]–[6]. This paper also focuses on data-driven techniques for diagnosing bearing defects in IMs.

Bearing defects usually result in consecutive and periodic impulses in machine vibrations anytime a roller moves over the defective surface. Then, a prevalent technique for diagnosing bearing defects in IMs is based on the processing of the raw vibrational signals to extract informative features for the use of fault classifiers [7].

Various signal analysis techniques have been applied to the nonstationary and complex vibrational signals to extract discriminant features, including time-domain, frequency-domain [e.g., fast Fourier transform (FFT)] and time–frequency-domain [e.g., wavelet packet transform (WPT) and empirical mode decomposition (EMD)] features [3], [8]–[10]. However, actual vibrational spectra are usually represented by a considerable number of frequency or time-frequency components, i.e., high-dimensional feature sets, which cannot be easily handled. The extracted features are fed to the feature reduction (FR) module in order to filter redundant features, reduce the computational burden, and provide a more informative and discriminant sets of features for the fault classifiers to recognize faulty and healthy states [11].

The number of training samples for each class (i.e., normal and different faults) plays a vital role in the performance of the diagnostic systems; however, in designing of the most diagnostic classifiers, it has been assumed that the observed samples from different classes have almost a balanced class distribution. However, in real-world applications, systems (e.g., IMs) usually operate under the normal condition, and thus, samples of the normal state are expected to be more than samples of faulty classes. This can result in the collection of challenging imbalanced sets of samples, which further complicate the process of learning from the samples and diagnosing the faults. The main drawback with the imbalanced sets is that typical fault classifiers are usually biased with respect to (w.r.t.) the major class (i.e., normal state), and therefore, there is a higher misclassification rate for the samples of the minor classes (i.e., faulty classes). Although a considerable number of data-driven techniques have been applied for diagnosing bearing defects in IMs [11], [12], there is a need to design data-driven techniques for diagnosing bearing defects under the class-imbalance (CI) conditions. Hence, in this work, various state-of-the-art data-driven techniques have been designed for diagnosing bearing defects under the CI conditions.

The primary contribution of this paper is such a general integrated scheme for diagnosing bearing defects from sets with CI distributions of the samples. This diagnostic framework includes various state-of-the-art feature extraction (FE), feature selection (FS), and dimensionality reduction (DR) techniques along with various advanced techniques for the class-imbalanced learning (CIL). These CIL techniques include some state-of-the-art data-level, algorithm-level, and ensemble-based approaches. Besides, this work proposes a novel oversampling technique for the CIL. The primary novelty of this approach is in generating a set of incomplete samples representative of the minor classes and imputing them by resorting to the expectation maximization (EM) algorithm to produce new synthetic samples of the minor classes. The proposed diagnostic scheme is verified w.r.t. the standard and widely used Case Western Reserve University (CWRU) bearing datasets [13].

Over this diagnostic scheme, an empirical comparison of the performance of the data-driven techniques has been made with a triple objective. The former is to examine which technique outperforms the others in diagnosing bearing defects in IMs. The second one is to study the impacts of various FE and FR techniques in the diagnostic performance. The third one is the sensitivity analysis, where the performance of the diagnostic scheme is examined w.r.t. several datasets with different CI ratios. The diagnostic scheme is designed in a way that well-established conclusions can be extracted. The attained results show that the proposed novel technique outperforms other state-of-the-art CIL techniques in diagnosing bearing defects in terms of both performance measures and stability of the attained results.

The rest of this paper is organized as follows. Section II contains the problem statement for diagnosing bearing defects in IMs and describes the CI problem and the need for efficient data-driven fault diagnosis techniques in CI conditions. This section briefly presents various state-of-the-art techniques and also proposes a novel and efficient technique for the CIL. Section III describes main modules of the diagnostic scheme, including segmentation, FE, FR, and fault classification. Section IV presents and compares the experimental results. Section V summarizes and concludes this paper.

II. Problem Statement

The nature of the CI problem and the state-of-the-art CIL techniques are first presented as they are the prerequisites for the precise problem definition.

A. Class-Imbalanced Learning

The process of measuring and gathering raw data from different sources to analyze and detect anomalies, unusual trends, and faults is crucial to enhance the decision-making process [14]. Many of the industrial processes, such as IMs, usually operate in the normal state. Thus, it is very common for the diagnostic systems to collect a large number of samples of the normal state in the batch of data, while only a very few faulty samples could be collected in practice. Various diagnostic schemes have been applied to industrial processes; however, detecting faults under the CI condition is a challenging task with growing attention from both academia and industry.

An imbalanced dataset can be described as a set of samples, in which the proportion of the representative samples of one class is significantly larger than other classes. The amount of this proportion brings up the definition of the "imbalance ratio," which is an important factor in selecting a proper classification technique. The imbalance ratio indicates that the collected data are highly imbalance, moderate, or low. The major class in an imbalance dataset referred to a class with more number of samples, while the minor class is often the class of interest and should be detected with high accuracy. Standard classifiers usually require to accurately classify samples of the minor class as well as the major class. However, a CI distribution can result in misclassification of the minority class samples and jeopardizes the classification performance. Besides, these classifiers usually consider the classification accuracy as a performance evaluation measure. However, the classification accuracy can be biased toward the overrepresented classes and cannot effectively reveal the prediction of the representative minority class samples [15].

Therefore, in CI conditions, different techniques should be utilized to effectively classify faults or samples of the minor classes. These techniques are needed to construct classification models that are able to accurately diagnose the most important and infrequent samples of faults [16]. Several techniques have been proposed to tackle the CI problem [15]–[17]. These techniques, depending on how they deal with the CI problem, can be divided into three main categories; algorithm-level approaches, data-level approaches, and cost-sensitive methods [15]. The former category addresses the problem by designing or modifying a classification algorithm in a way that it considers the importance of rare samples as well as frequent ones [17]. The second category often exploits different data-processing techniques (i.e., sampling process) in order to rebalance the data distribution prior to FE, FS, and/or classification. These methods are generally more versatile because they can be integrated to any

classification model [17], [18]. The third category includes cost-sensitive methods, which can belong to two previous categories. They either use different misclassification cost associated with each class sample (i.e., data-level approach) or alter the training procedure to take costs (i.e., algorithm-level approach) in order to bias the classifier toward the rare class [15]. Apart from these categories, ensemble-based techniques are successfully applied for CIL. Ensemble techniques (e.g., bagging or boosting) are usually merged with one of the above categories in order to improve the accuracy [16], [17].

This paper primarily focuses on the development of the most efficient CIL techniques for diagnosing bearing defects under CI conditions. The proposed diagnostic scheme uses a state-of-the-art data-level approach, the synthetic minority oversampling technique (SMOTE) [18]. The SMOTE aims to generate some artificial samples to oversample the minor class [18]. Each newly generated sample is along the line between a selected sample of the minor class and its nearest neighbor [15], [18].

This integrated diagnostic scheme also uses state-of-the-art ensemble-based CIL approaches, including RUSBoost [16] and SMOTEbag [19], that are the most common and robust CIL techniques [17]. The RUSBoost algorithm is made of two important components: random undersampling (RUS) and AdaBoost.M2 [16]. The former is a RUS method, which reduces the number of samples of the major class and makes the RUSBoost learning process faster. The latter, a boosting technique, aims to construct a composite classifier, which is often more accurate than each individual weak classifier.

SMOTEBagging, the so-called SMOTEBag [19], is also based on two components: SMOTE and Bagging algorithms. The former creates new synthetic samples by interpolating samples of the minor class to construct various balance subsets $\mathcal{S}_t$ from the CI dataset $\mathcal{S}$ and, then, iteratively makes use of the more balanced and diverse subsets $\mathcal{S}_t$ to train T weak classifiers h_t and constructs a bagged ensemble. SMOTEBag enhances the diversity among the base classifiers by changing the ratio of bootstrap replicates and generating various synthetic samples by the SMOTE over the bagging iterations.

Besides, this diagnostic scheme also uses the weighted extreme learning machine (WELM) [20], which is a state-of-the-art algorithm-level approach. The WELM is a generalized single hidden layer feedforward network with flexible processing nodes, which also makes use of the cost information [20]. The WELM can also belong to cost-sensitive approaches, since it uses a misclassification cost in a weighting scheme.

B. Imputation-Based Class-Imbalanced Learning

Here, a novel oversampling technique has been developed, which is based on the imputation of missing values on the minority class samples. This oversampling technique is based on EM [21], which is referred to as EM imputation-based class-imbalanced learning (EMICIL). EMICIL aims to decrease the imbalance ratio by generating new synthetic samples similar to the minority class samples.

Algorithm 1: EMICIL.

INPUT: $\mathcal{S}$ is a CI dataset
DEFINITIONS:
$\mathcal{S}^{\min}$ is the subset of samples of the minor class
$\mathcal{S}^{\mathrm{maj}}$ is the subset of samples of the major class
$m^{\min}$ is the # of samples of the minor class
m^{maj} is the # of samples of the major class
η is a fixed rate $\eta = \lfloor 0.25 \times m^{\min} \rfloor$
C is the # of classes in $\mathcal{S}$
REQUIRE: $S_{os}^{\min} = \oslash$
1. Copy samples of major class into $\mathcal{S}^{\mathrm{maj}}$
for $i = 1, \ldots, C-1$ **do**
  2. Set a counter $t = 1$
  3. Copy samples of i^{th} minor class into $\mathcal{S}_t^{\min(i)}$
  while $m^{\mathrm{maj}} - m_t^{\min(i)} \geq \eta$ **do**
    4. Induce random missing on η samples of $\mathcal{S}_t^{\min(i)}$ to form $\mathcal{Z}_t^{\min(i)}$
    5. Call EMI subroutine to estimate missing values
    $$\left[\mathcal{S}_{\text{est}}^{\min(i)}\right] = \text{EMI}(\mathcal{Z}_t^{\min(i)})$$
    where EMI stands for the EM imputation [21]
    6. Use imputed samples of $\mathcal{S}_{\text{est}}^{\min(i)}$ to form $\mathcal{S}_{\text{imp}}^{\min(i)}$
    7. Merge the minor subsets
    $$\mathcal{S}_{t+1}^{\min(i)} = \mathcal{S}_{\text{imp}}^{\min(i)} \cup \mathcal{S}_t^{\min(i)}$$
    8. $t = t + 1$
  end
  9. Create the oversampled subset of minor classes
  $$\mathcal{S}_{os}^{\min} = \mathcal{S}_{t+1}^{\min(i)} \cup \mathcal{S}_{os}^{\min}$$
end
10. Return the balanced dataset $\mathcal{S}_b = \mathcal{S}_{os}^{\min} \cup \mathcal{S}^{\mathrm{maj}}$

In order to gain a balanced dataset, it initially splits the dataset into subsets of samples of the minor class $\mathcal{S}^{\min}$ and major class $\mathcal{S}^{\mathrm{maj}}$. EMICIL randomly draws a fixed number η of the minority class samples $\mathcal{S}_t^{\min}$, where t stands for the iteration number, to induce missing values at each iteration. It then randomly selects an attribute to induce the missing value on the η selected minority class samples and forms $Z_t^{\min}$. EMICIL then estimates missing values of $\mathcal{Z}_t^{\min}$ by resorting to the expectation maximization imputation (EMI) technique [21], [22] and returns $\mathcal{S}_{\text{est}}^{\min}$. It then selects only imputed samples of $\mathcal{S}_{\text{est}}^{\min}$ to form $\mathcal{S}_{\text{imp}}^{\min}$ and, then, merges these newly imputed samples that are representative of the minor class with the previous subset of the minority class samples $\mathcal{S}_t^{\min}$. This creates a new and larger collection $\mathcal{S}_{t+1}^{\min}$ of representative samples of the minor class for the subsequent iteration. This procedure iteratively continues to create and combine the minority class samples, until the convergence criterion has been met $m^{\mathrm{maj}} - m_t^{\min(i)} \prec \eta$. This allows us to produce the synthetic minority class samples until $m_t^{\min}$ approximately reaches to m^{maj}. The pseudocode of EMICIL is presented in Algorithm 1.

C. Expectation Maximization Imputation

EMI is an approach for maximum-likelihood estimation (MLE) from incomplete data [21]. EMI receives an incomplete set of samples of the minor class $\mathcal{Z}_t^{\min(i)}$ as inputs. Hereafter, the notation $\mathcal{Z}_t^{\min(i)}$ is replaced by $\mathcal{Z}$, for the sake of brevity. Let $\mathcal{Z} = (\mathcal{Z}_{\rm obs}; \mathcal{Z}_{\rm mis})$ represent the input data, where $\mathcal{Z}_{\rm obs}$ and $\mathcal{Z}_{\rm mis}$ stand for the observed and missing data, respectively.

MLE intends to estimate the most likely model parameters describing observed data. The data $\mathcal{Z}$ generated by a model can be represented through a probability or density function $p(\mathcal{Z}|\Theta)$. The joint density of $\mathcal{Z}$ can be formulated as

$$p(\mathcal{Z}, \mathcal{I}|\Theta, \xi) = p(\mathcal{Z}|\Theta)p(\mathcal{I}|\mathcal{Z}, \xi) \tag{1}$$

where Θ indicates the data generation mechanism, $\mathcal{I}$ stands for the missing data indicator matrix, and ξ stands for the missing data mechanism. The density of the observed data $p(\mathcal{Z}_{\rm obs}, \mathcal{I}|\Theta, \xi)$ can be attained through integration of (1) over $\mathcal{Z}_{\rm mis}$

$$p(\mathcal{Z}_{\rm obs}, \mathcal{I}|\Theta, \xi) = \int p(\mathcal{Z}_{\rm obs}, \mathcal{Z}_{\rm mis}|\Theta)p(\mathcal{I}|\mathcal{Z}_{\rm obs}, \mathcal{Z}_{\rm mis}, \xi)d\mathcal{Z}_{\rm mis}. \tag{2}$$

Assuming that the data are missing at random, $p(\mathcal{I}|\mathcal{Z}_{\rm obs}, \mathcal{Z}_{\rm mis}, \xi)$ becomes independent of $\mathcal{Z}_{\rm mis}$, and thus, the joint density of the observed data can be reformulated as follows:

$$p(\mathcal{Z}_{\rm obs}, \mathcal{I}|\Theta, \xi) = p(\mathcal{Z}_{\rm obs}|\Theta)p(\mathcal{I}|\mathcal{Z}_{\rm obs}, \xi). \tag{3}$$

This yields to estimate the model parameters Θ from a set of incomplete data, which maximizes the likelihood function $L(\Theta|\mathcal{Z}_{\rm obs})$. The likelihood of Θ given $\mathcal{Z}_{\rm obs}$, neglecting the missing data mechanism, can be described

$$L(\Theta|\mathcal{Z}_{\rm obs}) = \int p(\mathcal{Z}_{\rm obs}, \mathcal{Z}_{\rm mis}|\Theta)d\mathcal{Z}_{\rm mis}. \tag{4}$$

EMI contains two major steps. The former, E-step, calculates the conditional expectations of the missing data given the observed data, $\mathcal{Z}_{\rm obs}$, and the current estimate of the parameters, Θ^i. The latter, M-step, calculates the MLE of Θ from the expected complete data log-likelihood.

EMI initially calculates the expected observed data log-likelihood over the distribution of $\mathcal{Z}_{\rm mis}$, given Θ^i and $\mathcal{Z}_{\rm obs}$:

$$E(\Theta|\Theta^i) = \int l(\Theta|\mathcal{Z}_{\rm obs}, \mathcal{Z}_{\rm mis})p(\mathcal{Z}_{\rm mis}|\mathcal{Z}_{\rm obs}, \Theta^i)d\mathcal{Z}_{\rm mis}. \tag{5}$$

It then finds Θ^{i+1} through maximizing E w.r.t. Θ

$$\Theta^{i+1} = \underset{\Theta}{\rm argmax} E(\Theta|\Theta^i). \tag{6}$$

Maximizing $E(\Theta|\Theta^i)$ w.r.t. Θ is the same as increasing $l(\Theta|\mathcal{Z}_{\rm obs})$ [21]. This yields to reformulating the model $p(\mathcal{Z}|\Theta)$ as follows:

$$p(\mathcal{Z}_{\rm obs}, \mathcal{Z}_{\rm mis}|\Theta) = p(\mathcal{Z}_{\rm obs}|\Theta)p(\mathcal{Z}_{\rm mis}|\mathcal{Z}_{\rm obs}, \Theta) \tag{7}$$

which results in the following reformulation of the log-likelihood:

$$l(\Theta|\mathcal{Z}_{\rm obs}, \mathcal{Z}_{\rm mis}) = l(\Theta|\mathcal{Z}_{\rm obs}) + \ln p(\mathcal{Z}_{\rm mis}|\mathcal{Z}_{\rm obs}, \Theta). \tag{8}$$

By expressing the expectation of the right-hand side of (8) over the distribution of $\mathcal{Z}_{\rm mis}$, given Θ^i and $\mathcal{Z}_{\rm obs}$:

$$F(\Theta|\Theta^i) = \int \left[\ln p(\mathcal{Z}_{\rm mis}|\mathcal{Z}_{\rm obs}, \Theta)\right] p(\mathcal{Z}_{\rm mis}|\mathcal{Z}_{\rm obs}, \Theta^i)d\mathcal{Z}_{\rm mis}. \tag{9}$$

One can reformulate the log-likelihood $l(\Theta|\mathcal{Z}_{\rm obs})$ as follows:

$$l(\Theta|\mathcal{Z}_{\rm obs}) = E(\Theta|\Theta^i) - F(\Theta|\Theta^i). \tag{10}$$

EMI then aims to iteratively estimate Θ from $l(\Theta|\mathcal{Z})$ by estimating the maximum-likelihood based on the observed data log-likelihood $l(\Theta|\mathcal{Z}_{\rm obs})$. EM begins with initial estimates of Θ and iteratively performs the two steps until the estimated Θ and imputed values become stable [21]. This algorithm can find a global optimum solution. However, it has a slow convergence rate for the data with the large number of missing scores, and thus, EMICIL avoids inducing a large number of missing scores per iteration.

D. Problem Definition

The ultimate goal of this work is to design a diagnostic scheme, which is able to detect multiple bearing defects (i.e., inner race defect, outer race defect, and ball defect) in IMs, while the number of collected representative samples of each defect is less than the number of samples of the normal state. It is very common to collect vibration data with a skewed class distribution, since IMs often operate in the normal condition. A key requirement to design an efficient data-driven diagnostic system is then the ability of CIL and diagnosing multiple defects under the CI problem.

Here, drive and fan end bearing data from the CWRU Bearing Data Center are used as a case study. The database contains various files of the vibration signal, which is available in [13]. The CWRU bearing database has been widely used as a testbed to assess various data-driven and vibration-based diagnostic systems and has become a standard reference in recent years [12]. Three different diagnostic techniques are designed in [12] and applied to all datasets of the CWRU.

In order to better analyze the performance of the diagnostic scheme under different CI conditions, three datasets D_1, D_2, and D_3 with three different levels of imbalance, i.e., low, moderate, and high, are created from the CWRU bearing vibration records. These datasets contain samples of the normal state (i.e., major class) and samples of different defects (i.e., minor class) in the drive end (DE) and the fan end (FE). Two different defect widths, i.e., 0.007 and 0.021 in, a motor load of 2 hp, and a shaft speed of 1750 r/min with a sampling rate of 12 kHz are considered in this study. The number of selected samples for class of normal and defects, imbalance ratios, and defect width are specified in Table I. In this table, the imbalance ratio of (20:1), (100:1), and (150:1) means that the number of samples in the class of major is 20, 100, and 150 times greater than the number of samples of each minor class that can be categorized as low, moderate, and high CI (i.e., LCI, MCI, and HCI) datasets, respectively. In Table I, the entries inside parentheses represent the number of samples after segmentation of the signals (see Section III-A).

TABLE 1
NUMBER OF SAMPLES IN EACH DATASET

Class	Defect Width (in)	Length of Signal (Number of Samples)		
		D_1	D_2	D_3
Ball (DE)	0.007 and 0.021	46 080 (45)	9216 (9)	6144 (6)
Ball (FE)	0.007 and 0.021	46 080 (45)	9216 (9)	6144 (6)
Inner Race (DE)	0.007 and 0.021	46 080 (45)	9216 (9)	6144 (6)
Inner Race (FE)	0.007 and 0.021	46 080 (45)	9216 (9)	6144 (6)
Outer Race (DE)	0.007 and 0.021	46 080 (45)	9216 (9)	6144 (6)
Outer Race (FE)	0.007 and 0.021	46 080 (45)	9216 (9)	6144 (6)
Normal	–	921 600 (900)	921 600 (900)	921 600 (900)
Total	–	1 198 080 (1170)	976 896 (954)	958 464 (936)
Ratio	–	20:1	100:1	150:1
	–	LCI	MCI	HCI

III. DIAGNOSTIC SCHEME

Fig. 1 shows the general structure of the proposed diagnostic scheme. This scheme has four main modules: segmentation, FE, FR, and fault classification. The collected signals from the IM are initially segmented in the first module. The second module aims to extract useful information from the generated segments in various domains: time domain, frequency domain, and time–frequency domain. The third module contains two submodules: FS and DR. These two submodules work in parallel, where the former aims to use various techniques to select the most informative features, and the latter aims to reduce the dimension of the feature space by resorting to various transformations. The generated features are then fed to the fault classification module, which contains various CIL techniques. These CIL techniques learn the relation between the features and defects and diagnose bearing defects under the CI conditions.

The four modules of the proposed diagnostic scheme and the performance measures are first presented as they are the prerequisites for the precise experimental evaluations.

A. Segmentation

Each collected dataset is indeed a vibration signal contains l representative samples (see the row "Total" in Table I). These samples are collected at 12 kHz. However, the collected samples during this interval represent various periods of the rotational procedure, and it is necessary, therefore, to segment the whole signal into successive intervals, resembling various nonoverlapping samples collected at different time stamps. According to our experiments, fixing the length of segments to 1024 is a proper choice, beyond which the performance measures begin to decrease. This value is fixed for all scenarios resulting into m nonoverlapping segments (see the entries inside parentheses in Table I) of the length 1024, $m = l/1024$.

B. Feature Extraction

The diagnostic efficiency highly depends on the effectiveness of the applied FE techniques. The FE module aims to process high-frequency vibration signals and extract informative features in order to achieve more accurate performance measures. The FE module mainly focuses on three different domains: time, frequency, and time–frequency. This module makes use of the state-of-the-art techniques to extract informative features from the vibration signals.

TABLE II
EIGHT TIME-DOMAIN FEATURES OF THE kTH SEGMENT OF THE VIBRATION SIGNAL CONTAINING l_k ACCELERATION AMPLITUDES x_{ik}

Statistical Measure	Definition
Mean	$X_1 = \frac{\sum_{i=1}^{l_k} x_{ik}}{l_k}$
Root Mean Square	$X_2 = \left(\frac{\sum_{i=1}^{l_k} (x_{ik})^2}{l_k}\right)^{1/2}$
Skewness	$X_3 = \frac{\sum_{i=1}^{l_k} (x_{ik}-\mu_k)^3}{(l_k-1)\sigma_k^3}$
Kurtosis	$X_4 = \frac{\sum_{i=1}^{l_k} (x_{ik}-\mu_k)^4}{(l_k-1)\sigma_k^4}$
Crest Factor	$X_5 = \frac{\max(\lvert x_{ik}\rvert)}{\left(\frac{1}{l_k}\sum_{i=1}^{l_k} x_{ik}^2\right)^{1/2}}$
Impulse Factor	$X_6 = \frac{\max(\lvert x_{ik}\rvert)}{\frac{1}{l_k}\sum_{i=1}^{l_k} \lvert x_{ik}\rvert}$
Margin Factor	$X_7 = \frac{\max(\lvert x_{ik}\rvert)}{\left(\frac{1}{l_k}\sum_{i=1}^{l_k} \sqrt{\lvert x_{ik}\rvert}\right)^2}$
Entropy	$X_8 = \sum_{i=1}^{l_k} -P(x_{ik}) \log_2 P(x_{ik})$

1) Time-Domain Features: The simplest preprocessing technique, which is also robust to load variations, is to compute statistical time-domain features from the generated nonoverlapping segments of vibration signals. The FE module then receives m nonoverlapping segments of vibration signals from the previous module and, then, computes eight statistical measures including mean, root mean square, skewness, kurtosis, crest factor, impulse factor, margin factor, and entropy [7].

Considering the kth ($1 \leq k \leq m$) segment of the vibration signal $\mathcal{S}$, the eight features can be formulated in Table II, where l_k is the total number of samples in the kth segment and x_{ik} stands for the ith sample in the kth segment. μ_k and σ_k^2 in Table II stand for the mean and the variance of the samples in the kth segment, respectively. P in Table II stands for the probability mass function.

2) Frequency-Domain Features: To extract frequency-domain features, the fundamental frequencies of the vibrational signals with the bearing defects and their amplitudes usually must be determined *a priori*, which is not a realistic assumption in online monitoring applications, where the nonstationary conditions change the vibrational spectra. Here, the FE module initially performs the FFT on each segment of the vibration signal, analyzes its frequency spectrum, and extracts statistical features in the frequency domain. For the sake of a fair comparison, it calculates eight statistical features. These features can be

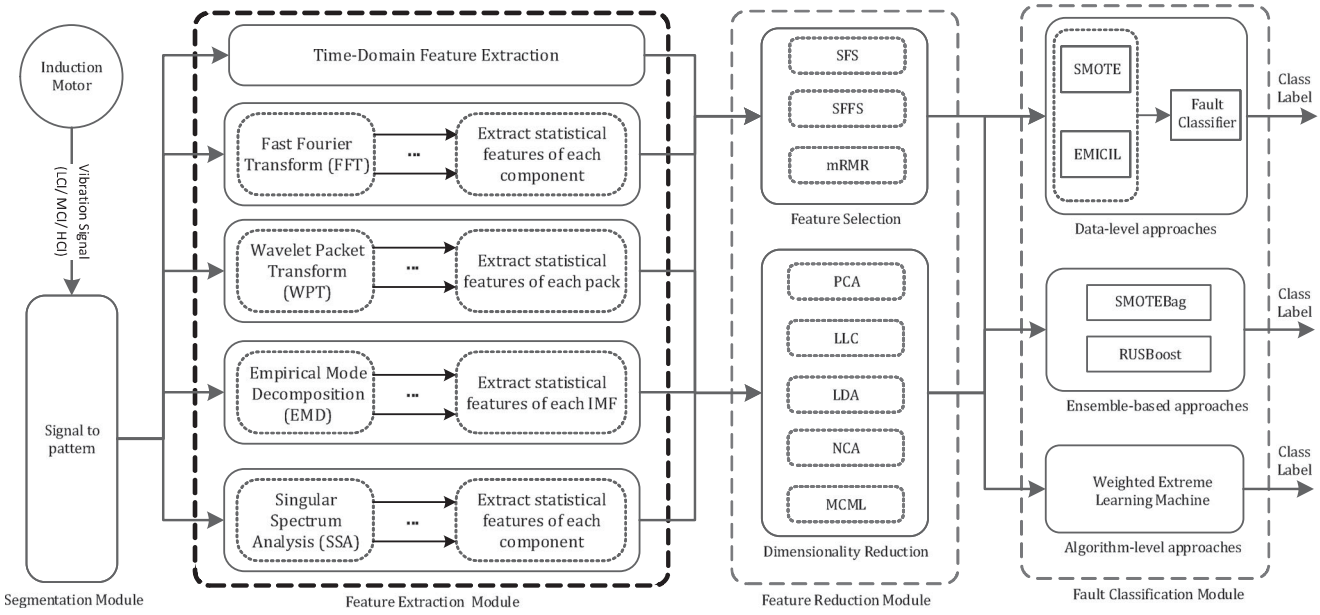

Fig. 1. Block diagram of the proposed diagnostic system for the CI condition.

formulated, similar to $\{X_1, \ldots, X_8\}$ in Table II, for each transformed segment (e.g., kth) of the vibration signal encompassing l_k frequency amplitudes f_{ik}.

3) Time–Frequency-Domain Features: Two state-of-the-art techniques are used to extract time–frequency-domain features from the vibration signal. These dual-domain analysis techniques are WPT and EMD that are widely used for nonstationary and nonlinear signal analysis. The WPT and EMD map raw vibration data to more explanatory and useful features aiming to enhance classification accuracy [23].

a) Wavelet packet transform: This is an extension of the wavelet transform, which aims to decompose the signal into different frequency subbands with higher resolution and provide local structure analysis of the spectrum [23].

The WPT makes use of low- and high-pass filters and iteratively decomposes both, approximation and detail coefficients, into two parts, unlike discrete wavelet transforms, which only consider the approximation coefficients for further decomposition, and thus, it provides a richest signal analysis [23].

Considering characteristics of the bearing vibration signals, here, the discrete Meyer wavelet along with the Shannon entropy is applied in the wavelet packet decomposition of the normal and faulty segments of the vibration signals.

Given an arbitrary segment of the time-domain vibration signal of l_k samples, the WPT with a tree depth of j results in 2^j final packs, i.e., $\{\mathbb{W}_{j,0}, \ldots, \mathbb{W}_{j,2^j-1}\}$. Each pack (i.e., kth), at depth of j, contains around Q wavelet coefficients, where $Q = l_k/2^j$. The FE module then extracts the eight different statistical features (listed in Table II) from the final 2^j packs of the tree by calculating the respective formula using the Q wavelet coefficients of each pack. This produces eight different statistical features for each wavelet packet and, thus, results in 256 features (i.e., 8×2^j).

b) Empirical mode decomposition: This is a state-of-the-art self-adaptive signal analysis technique [24] that can extract informative features from any complicated nonlinear and nonstationary signals. EMD breaks down the bearing vibration data into multiple orthogonal monocomponents, i.e., intrinsic mode functions (IMFs), so that the size of the generated IMFs equals the size of the given signal [24].

EMD initially identifies all the local maxima and minima of the signal and, consequently, computes the upper and lower envelopes by resorting to cubic fitting on the attained local maxima and minima. It then calculates the mean of the upper and lower envelopes of the signal to extract the detail. It repeats these steps on the detail signal until the mean of the detail signal becomes zero. It then extracts the current detail signal as the IMF component $\mathcal{M}_i(t)$ and, then, calculates the residual signal $\mathcal{U}_i(t)$ by subtracting the $\mathcal{M}_i(t)$ from the current signal. It finally makes use of the attained residual signal as a new signal for the subsequent iteration and iteratively continues to extract the subsequent IMFs. At each iteration, EMD extracts an $\mathcal{M}_i(t)$ with lower oscillation frequencies compared to previous iterations and consequently collects a residual $\mathcal{U}_i(t)$, so that the original signal $\mathcal{S}_k(t)$, $k = 1, \ldots, m$, can be reconstructed by means of all extracted IMFs and the monotonic residual function $\mathcal{U}_\gamma(t)$ as follows:

$$\mathcal{S}_k(t) = \sum_{i=1}^{\gamma} \mathcal{M}_i(t) + \mathcal{U}_\gamma(t) \tag{11}$$

where γ stands for the number of the extracted IMFs (here, $\gamma = 7$). The reader can refer to [24] for a more detailed explanation about the EMD algorithm. Each extracted IMF represents some frequency characteristics of the main vibration signal. To reveal these latent characteristics, eight statistical features (listed in Table II) are extracted from each IMF. This results in extraction of 56 statistical features from each segment of the vibration signal (i.e., $8 \times \gamma$).

4) Singular Spectrum Analysis (SSA): SSA is another state-of-the-art technique to analyze the vibration signals. It makes use of the Hankel matrix and singular value decomposition (SVD) for FE [25] and noise reduction from the vibration signals [25]. This can be considered as a nonparametric technique for time-series analysis that aims to express any 1-D vibration signal $\mathcal{S}_k(t)$ by a collection of multiple independent principal components, which may include white noise or different trends. This technique can be summarized as follows.

1) It initially embeds l_k data points of the signal $\mathcal{S}_k = (x_1, x_2, \ldots, x_{l_k})$ into the Hankel (trajectory) matrix A as follows:

$$A = [X_1 : \ldots : X_n] = \begin{pmatrix} x_1 & x_2 & \ldots & x_n \\ x_2 & x_3 & \ldots & x_{n+1} \\ \vdots & \vdots & \ddots & \vdots \\ x_m & x_{m+1} & \ldots & x_{l_k} \end{pmatrix}_{m \times n}. \tag{12}$$

Hence, signal $\mathcal{S}_k$ is projected into n lagged vector of size m, where $n = l_k - m + 1$ and $X_i = \{x_i, \ldots, x_{i+m-1}\}^T \in \mathcal{R}^m, (i = 1, \ldots, n)$;

2) It then applies SVD on the trajectory matrix A to cancel out the noise, which results in two orthogonal matrices $U_{m\times m}$ and $V_{n\times n}$. Matrix A can be reconstructed through $A = U\Sigma V^T$, where Σ is a diagonal matrix, which is composed of the square roots of the eigenvalues ζ_i of $A^T A$. The trajectory matrix A can be reformulated as follows:

$$A = \sum_{i=1}^{q} \sqrt{\zeta_i} U_i V_i^T \tag{13}$$

where $\{\zeta_1 \geq \zeta_2 \ldots \geq \zeta_q \geq 0\}$ are eigenvalues, U_i denotes the respective orthonormal eigenvectors, $V_i = A^T U_i / \sqrt{\zeta_i}$ stand for the principal components, and $q = \min(m, n)$.

3) It then splits the total set of m components into Δ disjoint sets and, consequently, groups the eigentriples (i.e., $\sqrt{\zeta_i}, U_i, V_i$) of similar characteristics $\delta \in [1, \Delta]$ that are used then to reconstruct the trajectory matrix.

4) It then makes an average over the skew-diagonal elements of the newly grouped matrices to form a Hankel matrix.

This is true since these values contribute into the same element in the newly derived vector $\psi_\delta = \{\psi_{\delta 1}, \psi_{\delta 2}, \ldots, \psi_{\delta l_k}\} \in \mathcal{R}^{l_k}$, i.e., principal component. ψ_δ is indeed the projected vector of the δth disjoint matrix, which can be obtained through diagonal averaging of the elements of the set as follows:

$$\psi_{\delta i} = \begin{cases} \frac{1}{i} \sum_{j=1}^{i} \phi_{j,i-j+1}, & 1 \leq i \leq m \\ \frac{1}{m} \sum_{j=1}^{m} \phi_{j,i-j+1}, & m \prec i \prec n \\ \dfrac{1}{l_k - i + 1} \sum_{j=i-n+1}^{m} \phi_{j,i-j+1}, & n \leq i \leq l_k \end{cases} \tag{14}$$

where $\phi_{j,i-j+1}$ stands for the elements of the δth disjoint matrix. It finally extracts reconstructed signals $\psi_\delta = \{\psi_{\delta 1}, \psi_{\delta 2}, \ldots, \psi_{\delta l_k}\}$, $\delta = 1, \ldots, \Delta$, as the most informative features. The FE module then calculates eight statistical features (listed in Table II) from each reconstructed signal. This results in the extraction of 80 statistical features from each segment of the vibration signal (i.e., $8 \times \Delta$).

C. Feature Reduction

A crucial requirement for the practical implementation of an online diagnostic system is the capability of providing a small set of informative features to guarantee efficient learning and immediate decision making. This can be performed by resorting to FR techniques. The FR module (see Fig. 1) contains state-of-the-art techniques for FS and DR.

1) Feature Selection: FS techniques aim to select and preserve a proper subset of n' informative features from all the extracted features n in order to attain more precise performance measures. FS techniques usually include a selection criterion and a search strategy [26].

Here, two variants of linear forward selection (LFS) [26] are used for FS. In general, LFS is a wrapper-based forward selection strategy, which can limit the number of selected features in each forward selection step and, thus, significantly decreases the number of evaluations. These two LFS variants are sequential forward selection (SFS) and sequential floating forward search (SFFS) [26]. Besides, in this work, another state-of-the-art technique, so-called minimal redundancy maximal relevance (mRMR) [27], is used for FS.

SFS [26] initially creates an empty subset, evaluates each candidate feature along with the current subset, i.e., previously selected features, and, then, appends the best feature into the current subset. The SFS algorithm, then, has a hill-climbing search mechanism and terminates when the preset number of features has been reached [26].

SFFS is a variant of floating techniques [28], which retreats features as long as the evaluation criterion is improving.

mRMR is a two-step FS algorithm, which merges the mRMR criterion and the wrapper strategy [27]. This is a computationally inexpensive technique, which can find a small subset of informative features. All of these FS techniques use the wrapper approach for the subset evaluation.

2) Dimensionality Reduction: DR techniques, on the other hand, aim to project the large-scale set of features onto a lower dimensional data space so that the diagnosis process can be performed faster and even more efficient. Moreover, the obtained low-dimensional features can decrease the storage space. These state-of-the-art DR techniques include two unsupervised approaches: principal component analysis (PCA) and locally linear coordination (LLC) [29]. The FR module also includes three supervised DR approaches: linear discriminant analysis (LDA) [30], neighborhood component analysis (NCA) [31], and maximally collapsing metric learning (MCML) [32].

PCA is a nonparametric unsupervised technique that extracts hidden information from a set of features, while decreasing the dimension of the original feature set [33]. PCA linearly transforms the samples to a set of new orthogonal coordinates, where the direction of the maximum variation of the samples lies on the first coordinate (principal component), and the direction of the next largest variation of the samples lies on the subsequent coordinate along with others [33].

LLC [29] is a state-of-the-art unsupervised DR technique, which uses the EM algorithm in order to compute a mixture of locally linear models and adjusts them to project the samples onto a low-dimensional feature representation [29].

Multiclass LDA [30] aims to projects samples onto a new feature space with the maximum separation between the means of projected classes and the minimum variance within each projected class.

NCA [31] aims to learn a linear projection matrix to be used in the k-nearest-neighbors algorithm. NCA selects a low-rank matrix to learn a low-dimensional feature space from the high-dimensional feature space. NCA then uses the gradient descent optimization approach to find a projection matrix that maximizes the leave-one-out classification performance w.r.t. the training subset.

MCML [32] reduces the dimensionality of the feature set by resorting to an ideal learned metric. MCML attempts to find and learn this ideal metric by solving a convex optimization problem, which aims to collapse all samples of the same class to a particular point and, at the same time, project samples of other classes into a very far distance point [32].

D. Fault Classification

The collected sets of features form CI sets of samples (i.e., LCI, MCI, and HCI scenarios) regardless of the techniques used for FE and FR, and thus, this module makes use of the state-of-the-art techniques for CIL. These CIL techniques, as stated in Section II, include SMOTE, EMICIL, RUSBoost, SMOTEBag, and WELM. The fault classification module receives these CI sets of samples as input and learns the relations between the features and defects to diagnose bearing defects under the CI conditions.

These CIL techniques, except for WELM, use the decision tree C4.5 algorithm as a base classifier, since it has been extensively used in CI domains. The procedure of FS is adjusted accordingly, i.e., the wrapper for FS also makes use of the C4.5, except for the WELM. The k-fold, stratified, nested cross-validation (CV) strategy is used to evaluate the model performance and to tune the C4.5 parameters. The nested CV contains two loops. The outer CV divides the whole dataset into k outer splits in a stratified manner. Each of the outer splits is iteratively used as a test subset (holdout outer fold), while the remaining $k-1$ splits are used as a training subset. The inner CV loop further subdivides each training subset of the outer loop into k' inner splits. The inner CV loop aims to tune model parameters. It iteratively holds out one fold for validation and trains a model using $k'-1$ remaining splits and, then, evaluates the trained model through the validation subset (holdout inner fold) in order to find the optimal parameters. These optimal parameters are then used to retrain a model based on the overall training set in the outer loop. The performance of this model is then evaluated on the unseen test subset (holdout outer fold) to get an unbiased estimate.

Accuracy has been extensively used for performance evaluation under balanced conditions, but it is not a proper performance measure under CI conditions as it has bias over the major class. Thus, in this work, various performance measures that are suitable for the multi-CI problem are considered into account. These performance measures are weighted average of receiver operating characteristic (ROC) area, weighted average of Matthews correlation coefficient (MCC), and weighted average of F-measure, which can be calculated based on the confusion matrix. For the sake of brevity, hereafter, weighted averages of ROC, MCC, and F-measure are written as "ROC," "MCC," and "F-measure," respectively.

IV. Experimental Results

The proposed diagnostic scheme, presented in Fig. 1, uses various state-of-the-art FE and FR techniques to process input signals and provide proper sets of informative features for the fault classification module. The fault classification module contains a novel and several state-of-the-art CIL techniques for diagnosing bearing defects under the CI condition.

Three scenarios with different imbalance ratios, i.e., low, moderate, and high, are used to evaluate the diagnostic system. The number of samples for each class, imbalance ratios, and defect widths are specified in Table I. This allows analyzing the sensitivity of the performance of the diagnostic system w.r.t. various CI ratios.

Diagnostic results are compared in terms of the weighted averages of ROC, MCC, and F-measure. These performance measures are used to evaluate the performance of the CIL techniques and to compare the impacts of each FE and FR techniques in the diagnostic performance. Besides, this section aims to study and compare the diagnostic performance of CIL techniques w.r.t. FE and FR modules at different CI ratios. Figs. 2–6 contain all three different performance metrics, since the attained results based on these metrics have similar trends. These performance metrics are the outputs of the diagnostic system. These outputs can be achieved by feeding the data (D_1, D_2, D_3) to the FE module (any FE technique), feeding the extracted features to the FR module (any FR technique), and, then, sending the reduced sets of features to the fault classification module (any CIL technique).

Fig. 2 compares the state-of-the-art techniques in each module for diagnosing bearing defects over all three scenarios. Each panel of Fig. 2 contains all performance measures (circles) attained by any integrated technique (FE techniques $\times$ FR techniques $\times$ CIL techniques $\times$ scenarios $\times$ performance measures). Fig. 2(a) depicts the distribution of the performance measures attained through each FE technique. The boxes illustrate the distribution range of the performance measures (circles) between the first and third quartiles, solid squares represent the mean of the performance measures for each FE technique, solid lines indicate the median value of the performance measures attained through each FE technique, the dots represent the outliers, and dash lines represent the outlier range. These FE techniques are ranked w.r.t. the mean values as FFT, WPT, EMD, SSA, and time-domain analysis. The FFT also has the smallest box, which shows the minimum variation of the performance measures and the most stable FE technique. Time-domain features produce the largest window and also have the maximum number of outliers, i.e., the most unstable diagnostic performances.

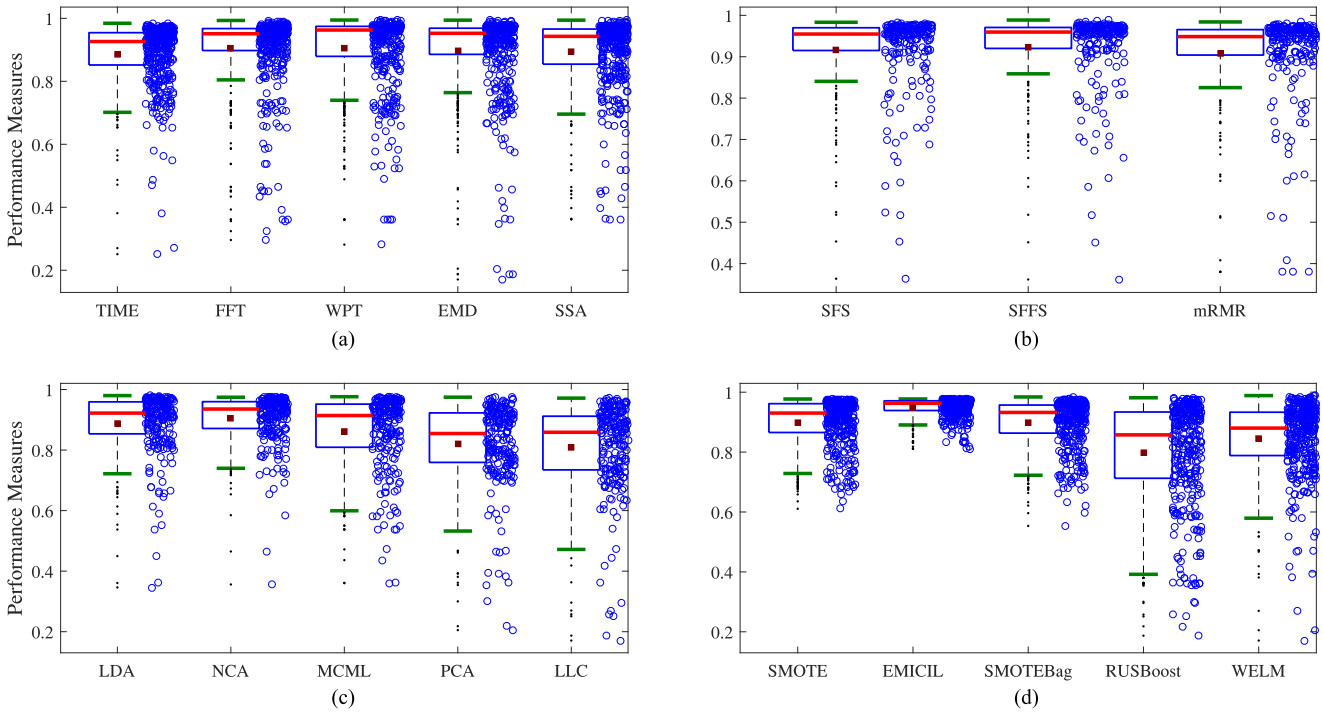

Fig. 2. Box plots represent distributions of the performance measures attained through each module.

Fig. 2(b) shows the distribution of the performance measures obtained through each FS technique. These FS methods are ranked w.r.t. the mean values as SFFS, SFS, and mRMR. SFFS also has the smallest box, which indicates the least variation and maximum stability among other competitors. However, the performance measures attained by these FS methods are very close to each other.

Fig. 2(c) represents the distribution of the performance measures achieved through each DR technique. These DR techniques are ranked w.r.t. the mean values as NCA, LDA, MCML, PCA, and LLC. NCA also has the smallest box, which indicates the least variation and the most stable DR technique. LLC yields the largest box and the maximum number of outliers, which represent the maximum instability among other DR techniques.

Fig. 2(d) illustrates the distribution of the performance measures attained by each CIL technique. These include all the performance measures attained through all FE and FR techniques. These fault classification techniques are ranked w.r.t. the mean values as EMICIL, SMOTE, SMOTEBag, WELM, and RUSBoost. EMICIL also has the smallest box, which represents the least variation and the most stable CIL technique. EMICIL significantly outperforms other CIL techniques. RUSBoost and WELM yield the largest boxes and the maximum number of outliers, respectively, which result in the most unstable diagnostic performances.

Fig. 3 shows the distribution of the performance measures (circles) obtained by each CIL technique through each FE technique for the scenario with the low imbalance ratio. The mean value of the performance measures obtained by each CIL technique is shown with a distinct marker (see the legend in Fig. 3). The mean values obtained through different FE techniques are connected by means of a distinct line for the sake of better comparison. This explanation is valid for the rest of figures in the paper except for the CI ratio and the module under study.

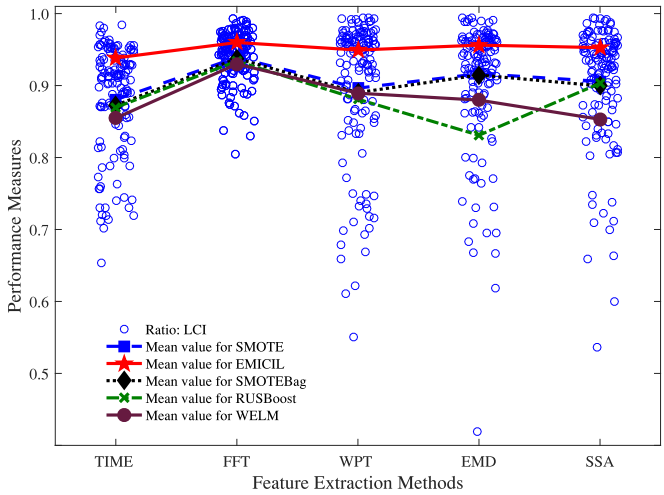

Fig. 3. Distributions of the performance measures on the scenario of the low imbalance ratio attained by each FE technique w.r.t. various CIL techniques (i.e., the mean values attained by each CIL technique are connected for a better visual evaluation).

Fig. 3 shows that EMICIL outperforms other CIL techniques regardless of the type of FE, i.e., EMICIL is insensitive to the type of extracted features. EMICIL is also the most stable fault classification technique. The rest of the CIL techniques are ranked as SMOTE, SMOTEBag, RUSBoost, and WELM. The last two are the most unstable CIL techniques.

Fig. 4 shows the distribution of the performance measures obtained by each CIL technique through each FE technique for the scenario with the medium imbalance ratio. Increasing the

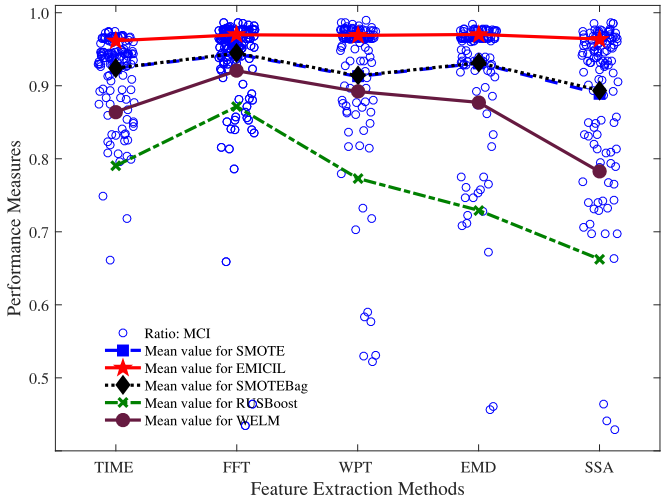

Fig. 4. Distributions of the performance measures on the scenario of the medium imbalance ratio attained by each FE technique w.r.t. various CIL techniques (i.e., the mean values attained by each CIL technique are connected for a better visual evaluation).

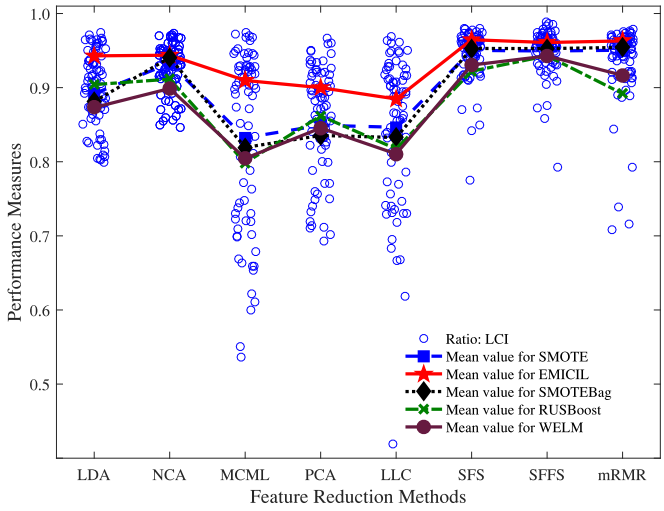

Fig. 5. Distributions of the performance measures on the scenario of the low imbalance ratio attained by each FR technique w.r.t. various CIL techniques (i.e., the mean values attained by each CIL technique are connected for a better visual evaluation).

imbalance ratio does not change the rank of CIL techniques except for WELM and RUSBoost. The attained results (not presented due to space limitation) for the scenario with the high imbalance ratio are very similar to Fig. 4 for the scenario with the medium imbalance ratio. This indicates that EMICIL outperforms other CIL techniques regardless of the type of extracted features and the imbalance ratios.

SMOTE and SOMTEBag are the subsequent ranks with a slight difference. WELM outperforms RUSBoost with the increasing of imbalance ratio, and thus, WELM and RUSBoost are ranked fourth and fifth for the scenarios with medium and high imbalance ratios.

Figs. 3 and 4 show that increasing the imbalance ratio results in significant reduction of the performance measures obtained by RUSBoost and WELM, i.e., these techniques are very sensitive to the imbalance ratio and their performances decrease by increasing the imbalance ratio. On the other hand, EMICIL outperforms other CIL techniques and is robust to the change of the imbalance ratio. The performance measures obtained by EMICIL are even improved by increasing the imbalance ratio, which is due to the use of EMI for producing a large number of distinct samples.

Fig. 5 shows the distribution of the performance measures (circles) obtained by each CIL technique through each DR and FS technique for the scenario with the low imbalance ratio. The mean value of the performance measures obtained by each CIL technique is shown with a distinct marker (see the legend in Fig. 5).

Fig. 5 shows that EMICIL outperforms other CIL techniques regardless of the type of DR and FS techniques used for the FR. EMICIL is also the most stable fault classification technique. The rest of the CIL techniques are ranked as SMOTE, SMOTEBag, RUSBoost, and WELM. The last two are the most unstable CIL techniques.

Fig. 6 shows the distribution of the performance measures obtained by each CIL technique through each DR and FS technique

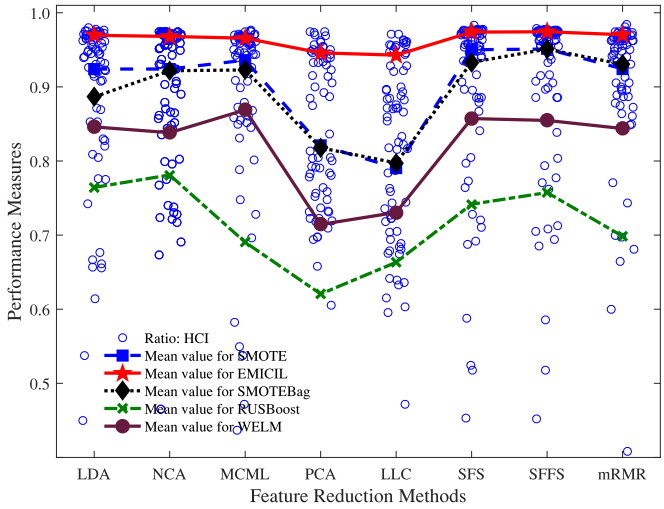

Fig. 6. Distributions of the performance measures on the scenario of the high imbalance ratio attained by each FR technique w.r.t. various CIL techniques (i.e., the mean values attained by each CIL technique are connected for a better visual evaluation).

for the scenario with the high imbalance ratio. This shows that increasing the imbalance ratio does not change the rank of CIL techniques except for WELM and RUSBoost. The attained results (not presented due to space limitation) for the scenario with the medium imbalance ratio is very similar to Fig. 6 for the scenario with the high imbalance ratio. This indicates that EMICIL outperforms other CIL techniques regardless of the type of FR and the imbalance ratio. SMOTE and SOMTEBag are the subsequent ranks with a slight difference. The WELM outperforms RUSBoost with the increasing of imbalance ratio, and thus, WELM and RUSBoost are ranked fourth and fifth for the scenarios with medium and high imbalance ratios.

Figs. 5 and 6 also confirm that increasing the imbalance ratio results in significant reduction of the performance measures obtained by RUSBoost and WELM, i.e., these techniques are

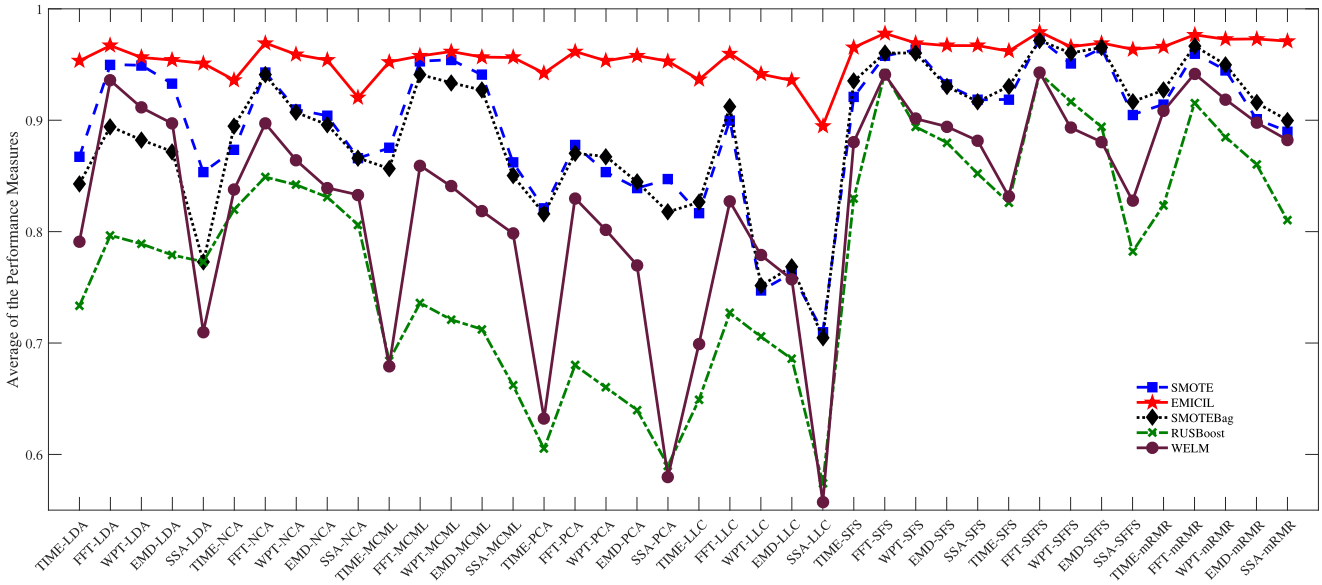

Fig. 7. Average of performance measures (three metrics) over three scenarios (LCI, MCI, and HCI) for each CIL technique w.r.t. each combination of FE and FR techniques.

very sensitive to the imbalance ratio. They also show that EMICIL outperforms other CIL techniques and is not very sensitive to the change of the imbalance ratio. They also indicate that the performance measures obtained by resorting to EMICIL are even improved by increasing the imbalance ratio. Figs. 5 and 6 also illustrate the rank of DR and FS techniques. In general, FS techniques outperform DR techniques as already represented in Fig. 2(b) and (c). FS techniques are also more stable compared to the DR techniques. NCA outperforms other DR techniques. The attained performance measures through NCA are comparable with the attained results through FS techniques. Fig. 6 also shows that, with increasing the imbalance ratio, the difference between the attained results through the FS techniques and the supervised DR techniques (NCA, LDA, and MCML) is reduced. On the other hand, the unsupervised DR techniques (PCA and LLC) underperform other competitors. This is particularly more significant for the scenario with the high imbalance ratio (see Fig. 6). The attained results in Figs. 5 and 6 show that the FS technique is followed by supervised and unsupervised DR techniques as second and third ranks, respectively.

Figs. 3–6 indicate that EMICIL, SMOTE, and SMOTEbag, regardless of the type of FE and FR, are more robust to increasing the imbalance ratio. Conversely, RUSboost and WELM are usually more sensitive to the imbalanced ratio, and their performance is degraded by increasing the ratio. This is due to the ability of EMICIL, SMOTE, and SMOTEbag in generation of the synthetic minority class samples, while the number of minority class samples is very limited in HCI.

A. Summary of Results

This section aims to focus on the integrated techniques to find out which combinations outperform the others in diagnosing bearing defects under CI conditions. Statistical tests are usually used to compare the attained results by different techniques however, in this work, a graphical presentation is used for comparison due to the integrated configuration of the techniques.

Fig. 7 illustrates the average of the performance measures (three metrics) over three CI scenarios (LCI, MCI, and HCI) for each CIL technique w.r.t. each combined FE–FR configuration. This can help to conclude that, given each FE–FR configuration, the CIL technique outperforms the others.

Fig. 7 shows that FFT–SFFS–EMICIL is the best configuration. It shows that the top-rank configurations use FFT for FE and EMICIL for fault classification. Top-rank configurations mostly make use of FS techniques (SFFS, SFS, and mRMR) and, then, supervised DR techniques.

Fig. 7 shows that SSA–LLC–WELM and SSA–LLC–RUSBoost are the lowest rank configurations. The five lowest ranked configurations include WELM and RUSBoost for classification, LLC and PCA for FR, and SSA for FE. Fig. 7 also shows that integrations of the unsupervised DR techniques with RUSBoost and WELM result in the lowest performance measures.

V. Conclusion

This paper proposes an efficient integrated scheme for diagnosing bearing defects in IMs, under the CI condition, which is quite relevant for actual diagnostic problems. This scheme has been evaluated on the CWRU bearing datasets. Besides, three different scenarios are used to evaluate the sensitivity of the diagnostic scheme to the CI ratio. The proposed scheme contains four modules including segmentation, FE, FR, and fault classification. In the FE module, several state-of-the-art techniques have been devised in the diagnostic scheme to extract

informative sets of features from the vibrational signal. The FR module includes various state-of-the-art FS and DR techniques, which aim to eliminate redundant features and generate discriminant and useful sets of features for the subsequent module. The fault classification module is made of some state-of-the-art data-level, algorithm-level, and ensemble-based approaches for CI learning, which are adopted for diagnosing bearing defects.

In the fault classification module, a novel oversampling technique, called EMICIL, for CIL has been developed, which is based on the EMI technique. The novelty of the EMICIL algorithm stands in producing a set of incomplete samples representative of each minor class and imputing them by resorting to the EM algorithm. This allows EMICIL to generate new synthetic samples of the minor class. EMICIL efficiently diagnoses bearing defects under the CI condition and improves performance measures. This improvement is more significant for the scenarios with the high CI ratios.

The achieved results show that EMICIL outperforms other state-of-the-art CIL techniques in diagnosing bearing defects in terms of both performance measures and stability of the attained results. The experiments suggest that EMICIL is a promising CIL approach, extensible to classification of CI datasets and other areas of industrial applications, which looks to be a worthwhile direction for future research. This integrated diagnostic scheme also enables an empirical comparison in each module to study the impacts of the state-of-the-art FE and FR techniques in the diagnostic performance.

References

[1] M. Riera-Guasp, J. A. Antonino-Daviu, and G. A. Capolino, "Advances in electrical machine, power electronic, and drive condition monitoring and fault detection: State of the art," *IEEE Trans. Ind. Electron.*, vol. 62, no. 3, pp. 1746–1759, Mar. 2015.

[2] R. K. Singleton, E. G. Strangas, and S. Aviyente, "The use of bearing currents and vibrations in lifetime estimation of bearings," *IEEE Trans. Ind. Informat.*, vol. 13, no. 3, pp. 1301–1309, Jun. 2017.

[3] P. Karvelis, G. Georgoulas, I. Tsoumas, J. Antonino-Daviu, V. Climente-Alarcn, and C. Stylios, "A symbolic representation approach for the diagnosis of broken rotor bars in induction motors," *IEEE Trans. Ind. Informat.*, vol. 11, no. 5, pp. 1028–1037, Oct. 2015.

[4] M. Cococcioni, B. Lazzerini, and S. Volpi, "Robust diagnosis of rolling element bearings based on classification techniques," *IEEE Trans. Ind. Informat.*, vol. 9, no. 4, pp. 2256–2263, Nov. 2013.

[5] X. Dai and Z. Gao, "From model, signal to knowledge: A data-driven perspective of fault detection and diagnosis," *IEEE Trans. Ind. Informat.*, vol. 9, no. 4, pp. 2226–2238, Nov. 2013.

[6] W. Li, S. Zhang, and S. Rakheja, "Feature denoising and nearest-farthest distance preserving projection for machine fault diagnosis," *IEEE Trans. Ind. Informat.*, vol. 12, no. 1, pp. 393–404, Feb. 2016.

[7] 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.*, vol. 64, no. 9, pp. 7210–7217, Sep. 2017.

[8] J. Antonino-Daviu, S. Aviyente, E. Strangas, and M. Riera-Guasp, "Scale invariant feature extraction algorithm for the automatic diagnosis of rotor asymmetries in induction motors," *IEEE Trans. Ind. Informat.*, vol. 9, no. 1, pp. 100–108, Feb. 2013.

[9] M. Van and H. Kang, "Bearing defect classification based on individual wavelet local fisher discriminant analysis with particle swarm optimization," *IEEE Trans. Ind. Informat.*, vol. 12, no. 1, pp. 124–135, Feb. 2016.

[10] B. Yang, R. Liu, and X. Chen, "Fault diagnosis for a wind turbine generator bearing via sparse representation and shift-invariant k-SVD," *IEEE Trans. Ind. Informat.*, vol. 13, no. 3, pp. 1321–1331, Jun. 2017.

[11] T. Rauber, F. de Assis Boldt, and F. Varejao, "Heterogeneous feature models and feature selection applied to bearing fault diagnosis," *IEEE Trans. Ind. Electron.*, vol. 62, no. 1, pp. 637–646, Jan. 2015.

[12] W. Smith and R. Randall, "Rolling element bearing diagnostics using the Case Western Reserve University data: A benchmark study," *Mech. Syst. Signal Process.*, vol. 64-65, pp. 100–131, 2015.

[13] Case Western Reserve University, Bearing Data Center, 2017, accessed: 2017-09-20. [Online]. Available: http://csegroups.case.edu/bearingdatacenter/home

[14] R. Razavi-Far, V. Palade, and E. Zio, "Optimal detection of new classes of faults by an invasive weed optimization method," in *Proc. Int. Joint Conf. Neural Netw.*, 2014, pp. 91–98.

[15] H. He and E. Garcia, "Learning from imbalanced data," *IEEE Trans. Knowl. Data Eng.*, vol. 21, no. 9, pp. 1263–1284, Sep. 2009.

[16] C. Seiffert, T. Khoshgoftaar, J. V. Hulse, and A. Napolitano, "RUSBoost: A hybrid approach to alleviating class imbalance," *IEEE Trans. Syst., Man, Cybern., Syst. A, Syst. Humans*, vol. 40, no. 1, pp. 185–197, Jan. 2010.

[17] M. Galar, A. Fernandez, E. Barrenechea, H. Bustince, and F. Herrera, "A review on ensembles for the class imbalance problem: Bagging-, boosting-, and hybrid-based approaches," *IEEE Trans. Syst., Man, Cybern., Syst. C, Appl. Rev.*, vol. 42, no. 4, pp. 463–484, Jul. 2012.

[18] N. Chawla, K. Bowyer, L. Hall, and W. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *J. Artif. Intell. Res.*, vol. 16, no. 1, pp. 321–357, 2002.

[19] S. Wang and X. Yao, "Diversity analysis on imbalanced data sets by using ensemble models," in *Proc. IEEE Symp. Comput. Intell. Data Mining*, 2009, pp. 324–331.

[20] W. Zong, G.-B. Huang, and Y. Chen, "Weighted extreme learning machine for imbalance learning," *Neurocomputing*, vol. 101, pp. 229–242, 2013.

[21] A. Dempster, N. Laird, and D. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Roy. Statist. Soc., Ser. B*, vol. 39, pp. 1–38, 1977.

[22] M. Farajzadeh-Zanjani, R. Razavi-Far, and M. Saif, "Efficient sampling techniques for ensemble learning and diagnosing bearing defects under class imbalanced condition," in *Proc. IEEE Symp. Ser. Comput. Intell.*, 2016, pp. 1–7.

[23] H. Keskes and A. Braham, "Recursive undecimated wavelet packet transform and DAG SVM for induction motor diagnosis," *IEEE Trans. Ind. Informat.*, vol. 11, no. 5, pp. 1059–1066, Oct. 2015.

[24] Y. Li, M. Xu, X. Liang, and W. Huang, "Application of bandwidth EMD and adaptive multi-scale morphology analysis for incipient fault diagnosis of rolling bearings," *IEEE Trans. Ind. Electron.*, vol. 64, no. 8, pp. 6506–6517, Aug. 2017.

[25] H. Jiang, J. Chen, G. Dong, T. Liu, and G. Chen, "Study on Hankel matrix-based SVD and its application in rolling element bearing fault diagnosis," *Mech. Syst. Signal Process.*, vols. 52/53, pp. 338–359, 2015.

[26] M. Gutlein, E. Frank, M. Hall, and A. Karwath, "Large-scale attribute selection using wrappers," in *Proc. IEEE Symp. Comput. Intell. Data Mining*, 2009, pp. 332–339.

[27] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 8, pp. 1226–1238, Aug. 2005.

[28] P. Pudil, J. Novovičová, and J. Kittler, "Floating search methods in feature selection," *Pattern Recognit. Lett.*, vol. 15, no. 11, pp. 1119–1125, Nov. 1994.

[29] Y. W. Teh and S. Roweis, "Automatic alignment of local representations," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2003, pp. 841–848.

[30] A. Martinez and A. Kak, "PCA versus LDA," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 2, pp. 228–233, Feb. 2001.

[31] J. Goldberger, G. Hinton, S. Roweis, and R. Salakhutdinov, "Neighbourhood components analysis," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2004, pp. 513–520.

[32] A. Globerson and S. Roweis, "Metric learning by collapsing classes," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2005, pp. 451–458.

[33] I. Jolliffe, *Principal Component Analysis*. New York, NY, USA: Springer, 2002.

Authors' photographs and biographies not available at the time of publication.