Efficient Sampling Techniques for Ensemble Learning and Diagnosing Bearing Defects under Class Imbalanced Condition

Maryam Farajzadeh-Zanjani, Roozbeh Razavi-Far, Mehrdad Saif
Department of Electrical & Computer Engineering
University of Windsor
401 Sunset Avenue, Windsor, N9B 3P4, Canada
Email: {farajza, roozbeh, msaif}@uwindsor.ca

***Abstract*—This paper focuses on sampling techniques to re-balance class distribution in a way that major and minor classes reach to almost equal number of the samples. A novel iterative over-sampling technique has been proposed which initially induces the missing values on the set of samples of the minor class and, then, imputes the missing scores to generate new synthetic samples of the minor class, in order to re-balance the class distribution. Two variations of the proposed over-sampling framework have been developed which make use of the Expectation Maximization and k-Nearest Neighbors imputation strategies. Moreover, the proposed over-sampling technique, which generates new samples for the minor class, has been integrated with a random under-sampling technique, which aims to simultaneously reduce the number of samples for the major class to speed up the process. The proposed sampling procedures have been used along with the ensemble of classifiers forming a diagnostic system. The constructed diagnostic scheme can efficiently diagnose multiple bearing defects in induction motors under class imbalanced condition.**

I. Introduction

Induction motors (IMs) play a significant role in industries and impact on a wide portion of industrial applications. Hence, IMs' performance and safety should be monitored to prevent unexpected failures and decrease downtime and maintenance cost of the system [1]. In other words, an appropriate monitoring technique to assess the automated system condition is needed to guarantee its reliability, efficiency and controllability. The monitored abnormalities in IMs are mainly related to defects which are occurred in critical components such as the bearings, the stator and the rotor [1].

Previous studies on IMs' source of failure hold the view that faults originated from the bearing cover a large proportion of failure distribution (e.g., 41%) in comparison with other components [2, 3]. Hence IM's failure diagnosis is concerned with bearing condition analysis has attracted the attention of many researchers [4, 5]. Generally, these defects may occur in different parts of the bearing as depicted in Figure 1. However, spalling of the races can be considered as the most frequent defect [5].

Many works have been focused on the processing of the bearing vibration signal to identify the system state, i.e., normal condition or any defects. Vibration signal is mainly analyzed in three different domains; Time, Frequency and Time-Frequency domains. Various fault diagnosis system make use of data-driven techniques to extract informative features from the bearing vibration signal. For instance, Ravi and Mohanty [6] used Fast Fourier Transformation (FFT) to analyze frequency-domain features, and Liu and Han [7] extracted several time-frequency features by means of Local Mean Decomposition (LMD). Moreover, time-domain features have been extensively studied due to their insensitivity to the change in motor load and the need for low computational efforts [8]. In this paper, time features have been considered to detect bearing defects.

Apart from feature extraction that is an essential task in bearing vibration analysis, a qualified intelligent classification algorithm is needed to diagnose samples of bearing vibration accurately and efficiently. Data-driven diagnostic techniques are usually make use of a classification algorithm to diagnose faults [9, 10]. Various classification algorithms such as fuzzy systems [11] and neural networks [12, 13] have been considered for fault diagnosis.

These fault classifiers aim to distinguish faulty (e.g., outer race defect, inner race defect or ball defect) and normal samples and determine the bearing health condition. However, these fault classification algorithms are typically based on the assumption that number of faulty samples are almost equal to number of normal samples. In other words, most of these fault classifiers are not designed for skewed-class data distribution, while collected data in industrial processes are often imbalanced [14]. In fact, IMs operate in the normal condition, hence, samples of normal class expected to be greater than faulty ones. Since class imbalance problem endangers the classification performance, some techniques should be applied to deal with this problem. One approach to handle class imbalance in the level of data is the use of sampling techniques, which aim to provide a balance dataset whether by under-sampling of the major class, e.g., Random Under-Sampling (RUS) or over-sampling of the minor class, e.g., Random Over-Sampling (ROS).

There exists other methods to tackle class imbalance problem by performing some modification on classification algorithms, i.e., defining the weight or the cost of contribution of samples in the classification task [15].

This work focuses on the data-level approach to create a dataset with the equal class distribution that can be used by the most of the fault classifiers. Hence, the benefits and the drawbacks of RUS and ROS as simple sampling methods along with state-of-the-art sampling techniques such as synthetic minority over-sampling technique (SMOTE) [16], are studied.

This paper proposes novel sampling techniques based on missing data imputation by means of the Expectation Maximization (EM) and the k-Nearest Neighbors (kNN) and applies them to diagnose bearing defects under the class imbalance condition.

The proposed diagnostic system uses an ensemble of classifiers, since, in general, combinations of sampling and ensemble schemes can lead to more versatile systems and obtain better performance [17]. The re-balanced set of samples obtained by the proposed sampling methods is fed to the ensemble of the fault classifiers (i.e., Adaboost.M1 and Bagging) to assess the efficiency of the proposed scheme in diagnosing bearing defects under the class imbalanced condition.

The reminder of the paper is structured as follows: Section II states the problem of classifying multiple defects under the class imbalance condition based on the bearing vibration data, and also briefly describes different units of the proposed diagnostic system. In Section III, the pre-processing unit is explained in detail containing two main components: feature extraction and sampling. The proposed sampling techniques to handle the class imbalance problem are also described in this section. The experimental results are presented in Section IV. This section also presents and compares the results of ensemble learning by means of feature subsets generated by the proposed sampling techniques and the state-of-the-art technique. Section V makes a conclusion of the paper.

II. Problem Statement

The large portion of IM's faults is related to bearing defects, hence, early detection and diagnosis of faults, which occurred in this component, is a substantial task. Diagnosing multiple bearing defects under the class imbalance condition is a challenging task since most of the classifiers are mainly devised for the class-balance distribution of data. Moreover, some of the classifiers can handle the class-imbalanced data for the binary class situations.

This paper aims to diagnose the bearing defects under the multi-class imbalanced condition. As in real world applications, data samples are often collected under skewed-class distribution, there is a need for a sampling technique to rebalance the data for the ease of classification and to facilitate training the fault classifiers.

In many works, vibration data selected form the case Western Reserve University (CWRU) Bearing Datasets [18] has been used to investigate the performance of data-driven diagnostic techniques [19]. In this work, the CWRU vibration data with 2% level of imbalance (LOI) is selected. LOI is calculated as the percentage of the proportion of the samples in minor class to the total number of the samples in the dataset. The imbalanced CWRU bearing datasets consist of the normal samples (i.e., class of major) and the faulty samples (i.e., minor class) situated in ball, inner race and outer race. The number of representative samples of the normal class is larger than the faulty ones, since IM often operates in the healthy condition.

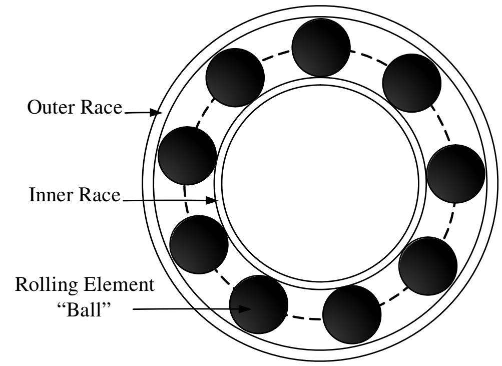

Fig. 1. Components of a ball bearing

A. Proposed Methodology

The proposed fault diagnostic scheme is illustrated in Figure 2. The diagnostic system is explained in two main units devoted for the pre-processing and the fault classification. In the former, statistical time-domain features are computed and sampling techniques are applied on the skewed-class vibration data that result in a balanced set of informative features. In the latter, the pre-processed and balanced set of samples are used to train an ensemble of fault classifiers (Adaboost.M1 and Bagging). Finally, the obtained results are compared with each other in order to find the best sampling method for the bearing fault diagnosis under the class imbalance condition.

The focus of this paper is to study the impact of different sampling techniques, which aim to tackle class imbalance condition, on the performance of the ensemble of fault classifier.

III. Pre-processing

A careful processing of the dataset before creating a classification model is a fundamental step in many applications. Pre-processing can prepare a qualified and discriminate set of features for an accurate classification and reliable analysis. Hence, a well-processed set of data can help to improve the performance of the fault classifiers.

A. Feature Extraction

The first step in the pre-processing, as it is shown in Figure 2, is to segment the vibration data with respect to the class labels. Once the representative samples of each class are divided into N none-overlapping folds (i.e., the intersection of all the folds is zero), five different statistical measures are calculated for each fold. These time-domain measures are defined in Table I to create the feature vectors [RMS, σ^2, Sk, Kr, $NSCM$], where m is number of samples in each fold, x_i is the i^{th} sample and μ is the mean value of the respective fold.

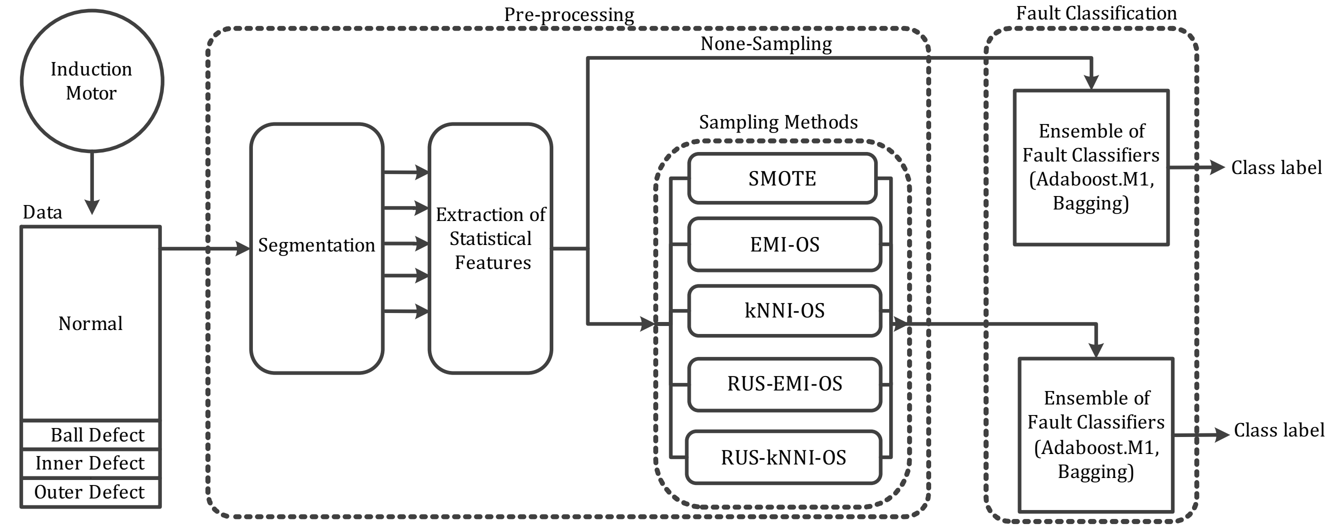

Fig. 2. General scheme of the proposed diagnostic system: the pre-processing unit, including novel sampling techniques and the fault classification unit.

TABLE I
TIME-DOMAIN FEATURES

Statistical Measures	Definition
Root mean square (RMS)	$\left(\frac{\sum_{i=1}^{m}(x_i-\mu)^2}{m}\right)^{1/2}$
Variance (σ^2)	$\frac{\sum_{i=1}^{m}(x_i-\mu)^2}{m-1}$
Skewness (Sk)	$\frac{\sum_{i=1}^{m}(x_i-\mu)^3}{(m-1)\sigma^3}$
Kurtosis (Kr)	$\frac{\sum_{i=1}^{m}(x_i-\mu)^4}{(m-1)\sigma^4}$
Normalized sixth central moment (NSCM)	$\frac{\sum_{i=1}^{m}(x_i-\mu)^6}{(m-1)\sigma^6}$

The first two steps of the pre-processing (i.e., segmentation and statistical feature extraction) do not change the level of imbalance. Therefore, the obtained data is not balanced (i.e., with 2% LOI) at this step and some sampling techniques are required to balance the size of the samples for each class.

B. Sampling Techniques

In order to prepare a dataset with an equal number of samples for each class, one may discard samples of the majority class (e.g., RUS) or increase samples of the minor class (e.g., ROS). Once the data is highly imbalanced, performing RUS on the majority of samples can result in information loss and performance reduction. On the other hand, ROS might lead to redundancy in datasets so that fault classifiers may not recognize the minority class significantly. To overcome these issues, SMOTE as an intelligent over-sampling technique has been used in this work. SMOTE creates some additional minor samples based on their k-Nearest Neighbors. Then, generated minority samples which are close and similar (i.e., not equal) to their nearest neighbors along with samples of the minor class form a balanced dataset [16].

Another approach for over-sampling of the minority class samples proposed in this paper is based on missing data imputation techniques. The missing data imputation-based over-sampling procedures have been performed by resorting to different imputation techniques such as Expectation Maximization (EMI) and k-Nearest Neighbors (kNNI). Imputation-based over-sampling techniques considered in this paper aim to generate new samples that are similar to the minority class samples to reduce the imbalance ratio.

In Figure 3, the proposed over-sampling schemes have been devised as parts of the pre-processing unit. Once the data at iteration t, $X^{(t)}$, is sorted based on the class labels, data related to the major class $X_{maj}^{(t)}$ and the minor class $X_{min}^{(t)}$ are obtained (see step 1). In multi-class datasets, there exist more than one class of minor, and similar procedure should be applied on each of the obtained minor class (i.e., there are more than one $X_{min}^{(t)}$ subset). In the step 2, missing values with rate of 20% are randomly induced on the minor class $X_{min}^{(t)}$ which results in an incomplete dataset, X_{min}^{mis}. Then, one of the imputation techniques, EMI or kNNI are employed to impute the missing values of the incomplete set X_{min}^{mis}. These missing data imputation procedures are explained in the following sections.

Once the estimated values for missing data are calculated and $\hat{X}_{min}^{est}$ is obtained, only these samples which contain imputed values are selected and gathered in $\hat{X}_{min}^{imp}$. After that these extracted samples are merged with the respective samples of the minority class, $X_{min}^{(t)}$, to form a new and larger set of samples of the minor class referred to as $X_{min}^{(t+1)}$, (see step 5). Then, the number of samples in $X_{min}^{(t+1)}$, $m_{min}^{(t+1)}$, is computed to see whether or not it is equal to the size of the subset of the major class and to see if the dataset is balanced.

EM imputation-based over-sampling (EMI-OS) approach makes use of the Expectation Maximization method in step 3 for missing data imputation. The EMI-OS procedure is stopped if $m_{min}^{(t+1)}$ is equal to the number of samples in the major class

$m_{maj}^{(t)}$. EMI-OS method is an over-sampling technique which increase the number of samples in the minor class iteratively until it reaches to $m_{maj}^{(t)}$.

kNN imputation-based over-sampling (kNNI-OS) approach uses k-Nearest Neighbor imputation method to create $\hat{X}_{min}^{est}$ (See step 3). In each iteration of kNNI-OS method, $m_{min}^{(t+1)}$ and $m_{maj}^{(t)}$ is compared to find whether or not faulty and normal classes have the same number of samples to stop the process.

Moreover, the over-sampling methods in highly imbalance multi-class datasets with large number of major class samples, $m_{maj}^{(t)}$, may increase a computational time to train the fault classifiers, and thus, under-sampling of the major class (e.g., RUS) has been performed along with the over-sampling techniques, i.e., EMI-OS and kNNI-OS.

An under-over-sampling technique originated from EMI-OS, is introduced which makes use of RUS combined with EMI-OS (i.e., referred to as RUS-EMI-OS for simplicity).

RUS-EMI-OS method applies random under-sampling without replacement on the major class $X_{maj}^{(t)}$ to reduce the number of samples to $m_{maj}^{(t+1)}$. In addition, EMI-OS is also performed simultaneously on the minor class $X_{min}^{(t)}$, to increase the number of samples to $m_{min}^{(t+1)}$. The stopping criterion in RUS-EMI-OS method is the equality of $m_{min}^{(t+1)}$ and $m_{maj}^{(t+1)}$, (See step 6). If the dataset is still class imbalance, then, the process continues until the stopping criterion has been met.

The Expectation Maximization (EM) algorithm, that can be considered as the core of EMI-OS method, is an iterative technique to compute maximum likelihood estimates of the parameters of models with partially observed data [20].

EM iterates between the expectation (E-step) which evaluates the posterior probabilities of the incomplete data, and maximization (M-step) which aims to update the model parameters θ (e.g., mean μ and covariance matrix Σ) using the posterior distribution of the missing data evaluated in the previous E-step. Figure 4 presents the pseudocode of the Expectation Maximization imputation (EMI) algorithm.

RUS-kNNI-OS method is also an under-over-sampling technique, which is similar to RUS-EMI-OS except the imputation procedure, which is based on kNNI. The main important part of the RUS-kNNI-OS and kNNI-OS method is the k-Nearest Neighbors imputation (kNNI) technique, which is based on the fact that the samples in the close proximity with other samples contain same characteristics [21]. Taking an incomplete sample from X_{min}^{mis} as a reference x_i, kNNI computes the distance from x_i to all available samples in X_{min}^{mis} and ranks them in descending order to determine its k nearest neighbors set as follows:

$$\mathcal{S}_i = \{\ x_p\ \}_{p=1}^{k} \tag{1}$$

It then replaces missing scores with the corresponding mean value of the k nearest samples:

$$\hat{x}_{ij} = \frac{\sum_{p=1}^{k} x_{pj}}{k} \tag{2}$$

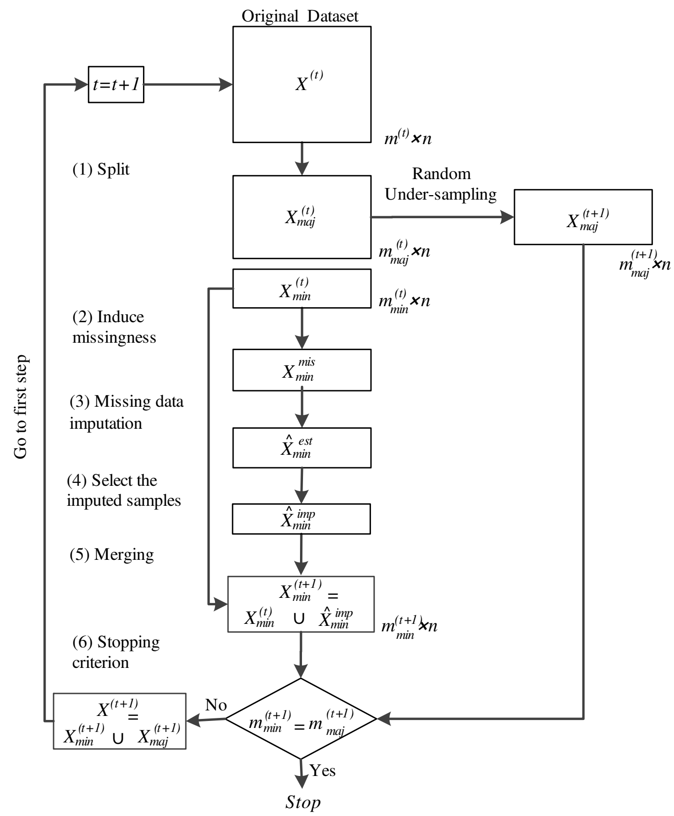

Fig. 3. The overall scheme of the proposed sampling technique based on the missing data imputation.

where x_{ij} stands for a missing score in the $j-th$ feature of the $i-th$ sample, k stands for the number of nearest neighbors and x_{pj} is the value of the $j-th$ feature of the $p-th$ nearest sample [22–24]. It finally returns the estimated matrix $\hat{X}_{min}^{est}$.

IV. Experimental Results

Vibration data from the CWRU bearing data center for loading condition of $2hp$ and shaft speed of $1750\ rpm$ has been used in this work. Bearing vibration consists of the normal condition of motor, also the ball defect (i.e., datasets 120 and 224), the inner race defect (i.e., datasets 107 and 211) and the outer race defect (i.e., datasets 132, 146, 159, 236, 248 and 260) with the defects width of 0.18 and 0.71 (mm).

Once feature extraction and segmentation procedures in the pre-processing unit has been performed, the number of normal samples, representative samples of the ball, the inner race and the outer race defects are decreased to 970, 20, 20 and 24, respectively. Hence, some over-sampling methods (i.e., SMOTE, EMI-OS, kNNI-OS) are exploited to add new samples to the minor class in order to increase the size of the representative samples of defect classes to be equal to the normal class. Moreover, two under-over-sampling techniques (i.e., RUS-EMI-OS, RUS-kNNI-OS) are also developed to decrease the size of the samples of the normal class by means of RUS while performing over-sampling on the samples of the

ASSUMING that $X_{min}^{mis} = (Y^{obs}, Y^{mis})$ in which Y^{obs} denotes the observed part and Y^{mis} denotes the missing part, and a joing distribution $p(Y^{obs}, Y^{mis}|\theta)$ over Y^{obs}, Y^{mis} and θ, EMI aims to maximize the likelihood function $p(Y^{obs}|\theta)$ w.r.t. θ.
SET the initial model parameters θ.
E-STEP: calculate $p(Y^{mis}|Y^{obs}, \theta^{old})$.
M-STEP: calculate θ^{new} as follows: $\theta_{new} = \underset{\theta}{argmax} \mathcal{L}(\theta, \theta^{old})$ where

$$\mathcal{L}(\theta, \theta^{old}) = \int p(Y^{mis}|Y^{obs}, \theta^{old}) \ln p(Y^{obs}, Y^{mis}|\theta) dY^{mis}$$

CHECK the stopping criterion to see if the estimated parameters or the imputed values stabilize. Otherwise $\theta^{old} \leftarrow \theta^{new}$ and go back to the E-step.

Fig. 4. The pseudocode for the EMI algorithm [20].

defective classes by means of EMI-OS or kNNI-OS, until the balanced data distribution has been reached.

In the fault diagnosis system studied in this work, Adaboost.M1 and Bagging as an ensemble of classifiers are considered to diagnose samples of normal and defects under the imbalanced condition. The base fault classifier to construct the ensemble approach is C4.5 decision tree algorithm. In order to obtain more reliable and accurate performance evaluation of the fault classifiers, a 10-fold cross validation has been performed and the confusion matrix has been studied. The two ensembles (i.e., Adaboost.M1 and Bagging) are combined with four different proposed sampling techniques along with the SMOTE which result in 10 different scenarios. Moreover, ensembles are applied on skewd-class data distribution without performing any sampling techniques. The attained confusion matrix for each scenario (i.e., there are 12 different scenarios in total) is exploited to analyze the performance of the respective fault classification model.

Considering accuracy for the performance evaluation of a classification model in a multi-fault imbalanced dataset cannot be a good representative of the classification task efficiency. Although, the accuracy is used as a performance measure in the balance data distribution, it tends to favor the class of major (i.e., samples from the normal class) in the imbalanced datasets that could compromise the performance evaluation [25].

Here, two proper evaluation measures, the weighted average of F-measure (i.e., hereafter referred to as F-measure) and the Macro average geometric (i.e., referred to as MAvG) are employed for the assessment of the fault classification models. The obtained confusion metrices are used to calculate these performance measures which can be formulated as follow [25]:

$$F-measure = \sum_{i=1}^{c} (\frac{m_i}{N} \cdot \frac{(2 \times precision_i \times recall_i)}{precision_i + recall_i}) \quad (3)$$

$$MAvG = (\prod_{i=1}^{c} \frac{\text{correctly classified samples for class i}}{m_i})^{\frac{1}{c}} \quad (4)$$

where the number of classes named as c, m_i shows the number of samples of the i^{th} class and N is the total size of the dataset.

In addition, $precision_i = TP_i/(TP_i + FP_i)$ and $recall_i = TP_i/(TP_i + FN_i)$, in which TP_i, FP_i and TN_i stands for the true positive, the false positive and the false negative regarding to class i, respectively.

Consider a confusion matrix Γ as depicted in Table II for a c-class problem where $\gamma_{p,q}$, ($p, q \in [1, c]$), referred to the value in row p and column q of the matrix Γ. Hence, TP_i, FP_i and TN_i can be calculated as follows:

TABLE II
CONFUSION MATRIX FOR THE MULTI-CLASS PROBLEM

		Predicted			
		class 1	class 2	$\cdots$	class c
Actual	class 1	$\gamma_{1,1}$	$\gamma_{1,2}$	$\gamma_{1,q}$	$\gamma_{1,c}$
	class 2	$\gamma_{2,1}$	$\gamma_{2,2}$	$\gamma_{2,q}$	$\gamma_{2,c}$
	$\cdots$	$\gamma_{p,1}$	$\gamma_{p,2}$	$\gamma_{p,q}$	$\gamma_{p,c}$
	class c	$\gamma_{c,1}$	$\gamma_{c,2}$	$\gamma_{c,q}$	$\gamma_{c,c}$

$$TP_i = \gamma_{i,i} \quad (5)$$

$$FP_i = \sum_{p=1}^{c} \gamma_{p,i} - \gamma_{i,i} \quad (6)$$

$$FN_i = \sum_{q=1}^{c} \gamma_{i,q} - \gamma_{i,i} \quad (7)$$

These performance measures, F-measure and MAvG, are computed for each of these 12 scenarios and results are organized in Table III. The last row of the table presents the mean value of the respective column (i.e., the average of F-measure and MAvg). From the table, it can be seen that sampling methods combined with Adaboost.M1 or Bagging achieve the performance of more than 0.98, while the obtained performance for fault classification without applying sampling methods (i.e., referred to as None) is 0.88 on average. The evaluation results of two ensembles are very close to each other, and their combination along with sampling techniques improve the performance. It is also noticeable that the use of sampling techniques along with the Adaboost.M1 presents slightly better performance compared to the Bagging combination.

In order to analyze the effect of sampling techniques in the classification task, Figures 5 and 6 are also provided. Figure 5 illustrates the value of the F-measure for different sampling techniques combined with the Adaboost.M1 and Bagging, separately.

From the Figures 5, it can be seen that EMI-OS and kNNI-OS over-sampling techniques have the highest performance among other techniques, while combination of EMI-OS and AdaBoost.M1 achieves a F-measure almost equal to 1 which means a perfect classification. RUS-EMI-OS and RUS-kNNI-OS are positioned as third and forth sampling techniques, respectively. Although RUS is a simple and fast under-sampling method, it may discard some important concepts from the data and results in a degradation of the classification performance.

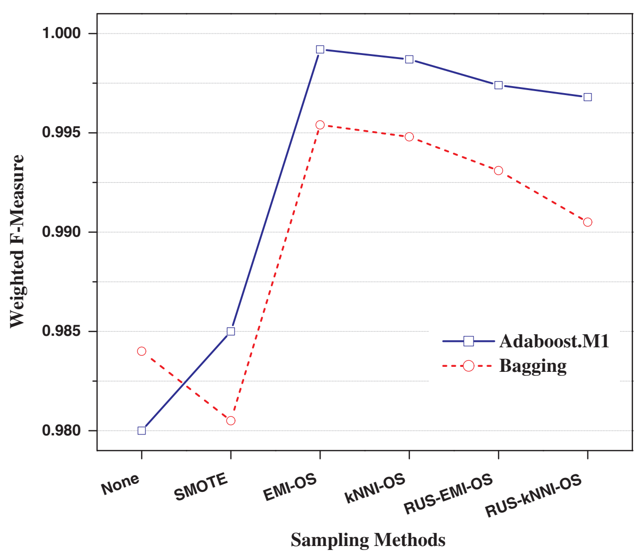

Fig. 5. The attained weighted F-measure by means of various sampling techniques and ensemble of the fault classifiers

The amount of MAvG which is calculated for each combination of sampling techniques and ensembles are compared together in Figure 6, to figure out the best classification model for the bearing fault diagnosis under the class imbalance condition with respect to MAvG. The performance trend depicted in Figure 6 shows that the integration of the sampling techniques and imputation methods could enhance the classification performance. Moreover, there is a considerable difference between the efficiency of the classification with sampling techniques (i.e., EMI-OS could obtain 0.999) or without any sampling methods (i.e., None is obtain 0.78). Figure 6 shows that SMOTE also follows the imputation-based sampling methods and can be ranked as fifth among all considered techniques.

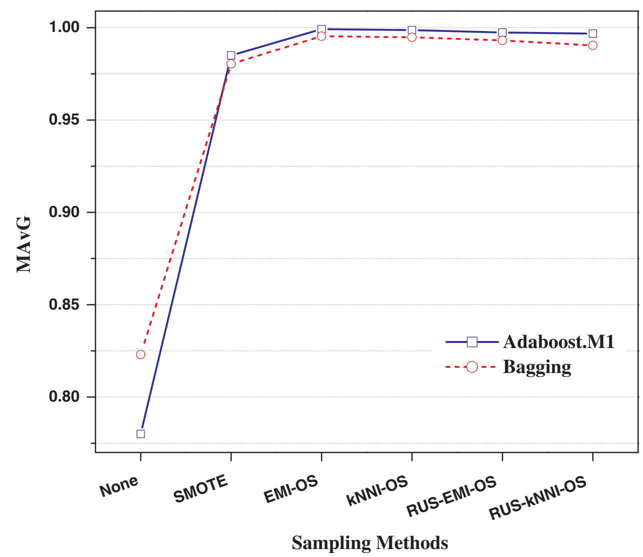

Fig. 6. The attained MAvG by means of various sampling techniques and ensemble of the fault classifiers

V. Conclusion

This paper proposed various sampling techniques to rebalance the class distribution of the samples and create proper inputs for the diagnostic system to diagnose the bearing defects (i.e., normal, ball defect, outer race defect and inner race defect). The proposed over-sampling technique initially induces missing values on the samples of the minor class and, then, imputes these missing values to generate new synthetic samples, while preserving the original samples of the minor class. It finally merges these samples to form a large set of samples of the minor class and, thus, balance the data. The Expectation Maximization (EM) and k-Nearest Neighbors (kNN) are used as missing data imputation techniques in variations of the proposed over-sampling technique. Besides, two under-over-sampling techniques are also developed, which perform random under-sampling on the samples of the major class along with the proposed imputation-based over-sampling procedure.

The proposed sampling techniques applied to five different time-domain features. These features are previously extracted from the bearing vibration signal. The feature extraction procedure and the proposed sampling techniques form the pre-processing unit, which aims to provide a set of class balanced features, to be fed to the ensemble of fault classifiers. Various sampling techniques have been devised in the pre-processing unit, creating different scenarios. Finally, the attained results are compared with respect to the weighted average of the F-mesaures and the MAvg. The experimental results have shown that the novel over-sampling techniques reach to the almost perfect classification performance for the bearing defects. The proposed under-over-sampling techniques follow the over-sampling techniques with a slight difference in terms of the performance measures.

TABLE III
PERFORMANCE MEASURES OBTAINED BY COMBINATION OF EACH SAMPLING METHOD AND ENSEMBLE OF THE FAULT CLASSIFIERS.

Performance Measures	None	SMOTE	EMI-OS	kNNI-OS	RUS-EMI-OS	RUS-kNNI-OS	Ensembles
F-Measure	0.98	0.985	0.999	0.998	0.997	0.996	Adaboost.M1
MAvG	0.78	0.985	0.999	0.998	0.997	0.996	
Average	0.88	0.985	**0.999**	0.998	0.997	0.996	
F-Measure	0.984	0.98	0.995	0.994	0.993	0.99	Bagging
MAvG	0.823	0.98	0.995	0.994	0.993	0.99	
Average	0.903	0.98	**0.995**	0.994	0.993	0.99	

This work could be extended for classification of imbalanced datasets. One of the concerns of the work is to study the relation between different sampling techniques and classification algorithms which can be studied in future research work.

ACKNOWLEDGMENT

This research was funded by The Natural Sciences and Engineering Research Council of Canada (NSERC).

REFERENCES

[1] S. Nandi and H. Toliyat, "Condition monitoring and fault diagnosis of electrical machines-a review," in *34th IEEE Industry Applications Conf.*, vol. 1, 1999, pp. 197–204.
[2] "Report of large motor reliability survey of industrial and commercial installations, part i," *IEEE Trans. on Industry App.*, vol. IA-21, no. 4, pp. 853–864, 1985.
[3] A. Bellini, F. Filippetti, C. Tassoni, and G. A. Capolino, "Advances in diagnostic techniques for induction machines," *IEEE Transactions on Industrial Electronics*, vol. 55, no. 12, pp. 4109–4126, 2008.
[4] M. Benbouzid, "Bibliography on induction motors faults detection and diagnosis," *IEEE Transaction on Energy Conversion*, vol. 14, pp. 1065–1074, 1999.
[5] J. Zarei, "Induction motors bearing fault detection using pattern recognition techniques," *Expert Systems with Applications*, vol. 39, pp. 68–73, 2012.
[6] V. Rai and A. Mohanty, "Bearing fault diagnosis using FFT of intrinsic mode functions in hilberthuang transform," *Mech. Syst. Sig. Proc.*, vol. 21, no. 6, pp. 2607 – 2615, 2007.
[7] H. Liu and M. Han, "A fault diagnosis method based on local mean decomposition and multi-scale entropy for roller bearings," *Mech. Mach. Theory*, vol. 75, pp. 67 – 78, 2014.
[8] M. Farajzadeh-Zanjani, R. Razavi-Far, M. Saif, J. Zarei, and V. Palade, "Diagnosis of bearing defects in induction motors by fuzzy-neighborhood density-based clustering," in *IEEE Int. Conf. on Machine Learning and Applications, ICMLA*, 2015.
[9] R. Razavi-Far, H. Davilu, V. Palade, and C. Lucas, "Neuro-fuzzy based fault diagnosis of a steam generator," in *7th IFAC Conf. on Fault Detection, Supervision and Safety of Technical Processes*, Barcelona, Spain, 2009, pp. 1180–1185.
[10] R. Razavi-Far and M. Kinnaert, "A multiple observers and dynamic weighting ensembles scheme for diagnosing new class faults in wind turbines," *Control Eng. Pract.*, vol. 21(9), pp. 1165–1177, 2013.
[11] Z. Xu, J. Xuan, T. Shi, B. Wu, and Y. Hu, "A novel fault diagnosis method of bearing based on improved fuzzy ARTMAP and modified distance discriminant technique," *Expert Syst. Appl.*, vol. 36, pp. 11 801–11 807, 2009.
[12] R. Razavi-Far and M. Kinnaert, "Incremental design of a decision system for residual evaluation: A wind turbine application," in *8th IFAC Conference Fault Detection, Supervision and Safety of Technical Processes*, vol. 8(1), 2012, pp. 343–348.
[13] M. Farajzadeh-Zanjani, R. Razavi-Far, and M. Saif, "Efficient feature extraction of vibration signals for diagnosing bearing defects in induction motors," in *Proceedings of IEEE Joint Congress on Neural Networks (IJCNN)*, 2016, pp. 4505–4511.
[14] H. He and E. A. Garcia, "Learning from imbalanced data," *IEEE Transactions on Knowledge and Data Engineering*, vol. 21, no. 9, pp. 1263–1284, 2009.
[15] X. Liu and Z. Zhou, "The influence of class imbalance on cost-sensitive learning: An empirical study," in *6th Int. Conf. on Data Mining (ICDM'06)*, 2006, pp. 970–974.
[16] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "Smote: Synthetic minority over-sampling technique," *J. Artif. Int. Res.*, vol. 16, no. 1, pp. 321–357, Jun. 2002.
[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. on Syst. Man, and Cybern., Part C*, vol. 42, no. 4, pp. 463–484, 2012.
[18] "Case western reserve university bearing data center website."
[19] T. Rauber, F. de Assis Boldt, and F. Varejao, "Heterogeneous feature models and feature selection applied to bearing fault diagnosis," *IEEE Trans. on Ind. Electronics*, vol. 62, no. 1, pp. 637–646, 2015.
[20] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society, Series B*, vol. 39, no. 1, pp. 1–38, 1977.
[21] P. J. Garca-Laencina, J.-L. Sancho-Gomez, A. R. Figueiras-Vidal, and M. Verleysen, "K nearest neighbours with mutual information for simultaneous classification and missing data imputation," *Neurocomputing*, vol. 72, pp. 1483–1493, 2009.
[22] P. Keerin, W. Kurutach, and T. Boongoen, "Cluster-based KNN missing value imputation for DNA microarray data," in *IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2012, pp. 445–450.
[23] R. Morin and D. Raeside, "A reappraisal of distance-weighted k-nearest neighbor classification for pattern recognition with missing data," *IEEE Trans. on Syst. Man, and Cybern.*, vol. 11, pp. 241–243, 1981.
[24] P. Batista and M. Monard, "An analysis of four missing data treatment methods for supervised learning," *Applied Artificial Intelligence*, vol. 17, pp. 519–533, 2003.
[25] J. Sanchez-Crisostomo, R. Alejo, E. Lopez-Gonzalez, R. Valdovinos, and J. Pacheco-Sanchez, "Empirical analysis of assessments metrics for multi-class imbalance learning on the back-propagation context," in *Advances in Swarm Intelligence*. Springer International Publishing, 2014, vol. 8795, pp. 17–23.