Expert Systems with Applications 40 (2013) 6788–6797

Contents lists available at SciVerse ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

A classifier fusion system for bearing fault diagnosis

Luana Batista, Bechir Badri, Robert Sabourin *, Marc Thomas

École de Technologie Supérieure, 1100, rue Notre-Dame Ouest, Montreál, QC H3C 1K3, Canada

ARTICLE INFO

Keywords:
Bearing fault diagnosis
Vibration analysis
Machine condition monitoring
Support vector machines
Iterative Boolean Combination
ROC curves
Classifier fusion

ABSTRACT

In this paper, a new strategy based on the fusion of different Support Vector Machines (SVM) is proposed in order to reduce noise effect in bearing fault diagnosis systems. Each SVM classifier is designed to deal with a specific noise configuration and, when combined together – by means of the Iterative Boolean Combination (IBC) technique – they provide high robustness to different noise-to-signal ratio. In order to produce a high amount of vibration signals, considering different defect dimensions and noise levels, the BEAring Toolbox (BEAT) is employed in this work. The experiments indicate that the proposed strategy can significantly reduce the error rates, even in the presence of very noisy signals.

© 2013 Elsevier Ltd. All rights reserved.

1. Introduction

Although the visual inspection of time- and frequency-domain features of measured signals is adequate for identifying machinery faults, there is a need for a reliable, fast and automated procedure of diagnosis (Samanta et al., 2004). Due to the increasing demands for greater product quality and variability, short product life-cycles, reduced cost, and global competition, automatic machine condition monitoring (MCM) has been gaining importance in the manufacturing industry (Liang et al., 2004). MCM systems allow for a significant reduction in the machinery maintenance costs, and, most importantly, the early detection of potential faults (Guo et al., 2005). Mass unbalance, rotor rub, shaft misalignment, gear failures and bearing defects are exemples of faults that may lead to the machine's breakdown (Samanta et al., 2004).

Besides the detection of the early occurence and seriousness of a fault, MCM systems may also be designed to identify the components that are deteriorating, and to estimate the time interval during which the monitored equipment can still operate before failure (Lazzerini and Volpi, 2011). These systems continuously measure and interpret signals (e.g., vibration, acoustic emission, infrared thermography, etc.), that provide useful information for identifying the presence of faulty symptoms.

The focus of this work is in rotating machines, which usually operate by means of bearings. Since they are the place where the basic dynamic loads and forces are applied, bearings represent a critical component. A defective bearing causes malfunction and may even lead to catastrophic failure of the machinery (Tandon and Choudhury, 1999). Vibration analysis has been the most employed methodology for detecting bearings defects (Thomas, 2011). Each time a rolling element passes over a defect, an impulse of vibration is generated. On the other hand, if the machine is operating properly, vibration amplitude is small and constant (Alguindigue et al., 1993). Another methodology successfully applied to this problem has been the acoustic emission (AE) (Elmaleeh and Saad, 2008; Tandon and Choudhury, 1999).

Automatic bearing fault diagnosis can be viewed as a pattern recognition problem, and several systems have been designed using well-known classification techniques, such as Artificial Neural Networks (ANNs) and Support Vector Machines (SVM). When these systems employ real vibration data obtained from bearings artificially damaged, they have to cope with a very limited amount of samples. Furthermore, with exception of a few works (Guo et al., 2005; Jack and Nandi, 2002) – which consider a validation set, besides the training and test sets –, the choice of the system's parameters, including the feature selection step, too often has been done by using the same datasets employed to train/test the classifiers. This may lead to biased classifiers that will hardly be able to generalize on new data. Another important aspect that has been little investigated in the literature is the presence of noise, which disturbs the vibration signals, and how this affects the identification of bearing defects (Lazzerini and Volpi, 2011).

In this paper, a classification system based on the fusion of different SVMs is proposed to detect early defects on bearings in the presence of high noise levels. Each SVM classifier is designed to deal with a specific noise configuration and, when combined together – by using the Iterative Boolean Combination (IBC) technique (Khreich et al., 2010) – they provide high robustness to different noise-to-signal ratio.

In order to produce a high amount of bearing vibration signals, considering different defect dimensions and noise levels, the BEAring Toolbox (BEAT) is employed in this work. BEAT is dedicated to the simulation of the dynamic behaviour of rotating ball bearings

* Corresponding author. Tel.: +1 (514) 396 8932; fax: +1 (514) 396 8595.
E-mail addresses: luana.bezerra@gmail.com (L. Batista), bechirbadri@yahoo.fr (B. Badri), robert.sabourin@etsmtl.ca (R. Sabourin), marc.thomas@etsmtl.ca (M. Thomas).

0957-4174/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.eswa.2013.06.033

in the presence of localized defects, and it was shown to provide realistic results, similar to those produced by a sensor during experimental measurements (Sassi et al., 2007).

This paper is organized as follows. Section 2 presents the state-of-the-art in automatic bearing fault diagnosis. Section 3 describes the experimental methodology, including datasets, measures used to evaluate the system performance, and the IBC technique. Finally, the experiments are presented and discussed in Section 4.

2. The state-of-the-art in automatic bearing fault diagnosis

Fig. 1 illustrates the general structure of a bearing. It is composed of six components: housing, outer race (OR), inner race (IR), rolling elements (RE) (i.e., rollers or balls), cage and shaft (Guo et al., 2005). As previously mentioned, the interaction of defects in rolling element bearings produces impulses of vibration. As these shocks excite the natural frequencies of the bearing elements, the analysis of the vibration signal in the frequency-domain, by means of the Fast Fourrier Transform (FFT), has been an effective method for predicting the health condition of bearings (Tandon and Choudhury, 1999).

Each defective bearing component produces frequencies, which allow for localizing different defects occurring simultaneously. BPFO (Ball Pass Frequency on an Outer race defect), BPFI (Ball Pass Frequency on an Inner race defect), FTF (Fundamental Train Frequency) and BSF (Ball Spin Frequency) – as well as their harmonics, modulating frequencies, and envelopes – are examples of frequency-domain indicators, calculated from kinematic considerations – that is, the geometry of the bearing and its rotational speed (Sassi et al., 2007).

It is worth noting that the shock amplitude is directly related to the defect dimension: the bigger the defect, the bigger the shock.

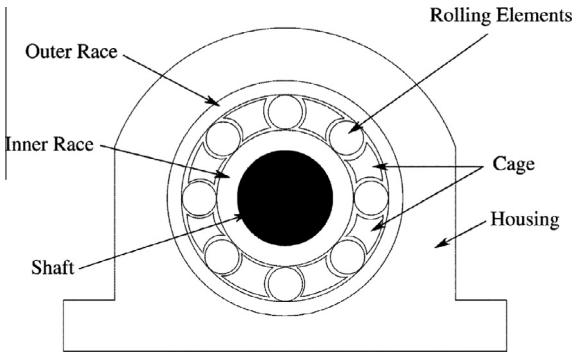

Fig. 1. Typical roller bearing, showing different component parts. Adapted from Jack and Nandi (2002).

Fig. 2 presents an example of a defect located in the outer race and its corresponding vibration signal.

Not only frequency- but also time-domain indicators have been widely employed as input features to train a bearing fault diagnosis classifier. Time-domain indicators are adimensional, and allow for representing the vibration signal through a single scalar value. For instance, *peak* is the maximum amplitude value of the vibration signal, RMS (Root Mean Square) represents the effective value (magnitude) of the vibration signal and Kurtosis describes the impulsive shape of the vibration signal. Table 1 presents the effectiveness (advantages and disadvantages) of some time-domain indicators in describing the presence (or absence) of faulty symptoms (Kankar et al., 2011; Sassi et al., 2008; Tandon and Choudhury, 1999).

A bearing fault diagnosis system may be designed to provide different levels of information about the defect (s). The first and simpler issue investigated in the literature is the detection of the presence or absence of a defect (Jack and Nandi, 2002; Samanta et al., 2004). The second issue is the determination of the defect location, which may occur in different components of a bearing (Alguindigue et al., 1993; Bhavaraju et al., 2010). Often, the type of defect is considered along with the defect location. For instance, some authors consider the following classes: sandblasting of IR/OR, indentation on the roll, unbalanced cage (Lazzerini and Volpi, 2011; Volpi et al., 2010), crack on IR/OR, spall on IR/OR, spalls on rollers (Widodo et al., 2009), generalized fault of two balls (Alguindigue et al., 1993), etc.

Finally, the severity of a bearing defect is the last and perhaps the most difficult information to be predicted. Through this information, it may be possible to estimate the duration during which the equipment can still operate safely. In the literature, this issue has been partially investigated, by associating a different class to each defect dimension (Cococcioni et al., 2009a, 2009b; Widodo et al., 2009). Cococcioni et al. (2009a), for example, have employed three classes for describing the seriousness of an "indentation on the roll", namely, light (450μm), medium (1.1 mm) and high (1.29 mm). The drawback of this strategy is that other defect dimensions are not considered by the classifier. A more suitable solution would be the estimation of defect dimensions as a regression problem.

Table 2 presents a summary of different systems reported in the literature, with their respective employed classification techniques, types of signal, descriptors (features), types of defects and datasets. It is important to mention that the bearing defects may be categorized as distributed or local. Distributed defects are due to unavoidable manufacting imperfections, such as surface roughness, waviness, misaligned races and off-size rolling elements (Sassi et al., 2007), whereas localized defects include cracks,

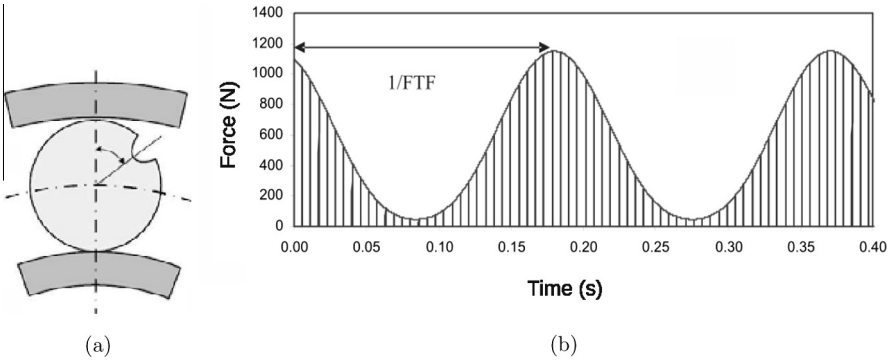

Fig. 2. Example of a hypothetical defect located in the rolling element (a) and its corresponding shock impulses (b), where FTF is the Fundamental Train Frequency (or cage frequency). Adapted from Sassi et al. (2007).

Table 1
Time-domain indicators.

Indicator	Advantage	Disadvantage
Peak	May indicate the presence of a defect even at the initial stage	The signal source is unknown; May create a false alarm
Root Mean Square (RMS)	Toward the end of the bearing life, the RMS level increases dramatically	Low sensitivity to indicate a defect at the initial stage; The signal source is unknown
Kurtosis	Low sensitivity to the variations of load and speed; Well suited for detecting a defect at the initial stage	When the defect is at an advanced stage, the Kurtosis value comes down to a value of an undamaged bearing; The signal source is unknown
Crest Factor (CF) Impulse Factor (IF)	Like Kurtosis, CF and IF are well suited for detecting a defect at the initial stage	Same problems as Kurtosis
Thikat (Sassi et al., 2008)	May indicate the presence of a defect at any rotational speed	Same problems as Kurtosis; No physical meaning; Needs the initial RMS value
Talaf (Sassi et al., 2008)	The *talaf* value constantly increases with the defect dimension; A slope change is an indication of impending failure; Indicates 4 levels of degradation	The signal source is unknown; No physical meaning; Needs the initial RMS value

Table 2
Survey of bearing fault diagnosis systems. (AE = Acoustic Emission, MLP = Multi-Layer Perceptron, SVM = Support Vector Machine, CHC = Convex Hull Classifier, PNN = Probabilistic Neural Network, RNN = Recirculation Neural Network, RBF = Radial Basis Function, GA = Genetic Algorithms, HMM = Hidden Markov Model, MFCC = Mel-Frequency Complex Cepstrum, SOM = Self-Organizing Maps, RVM = Relevance Vector Machine, QDC = Quadratic Discriminant Classifier, LDC = Linear Discriminant Classifier, PCA = Principal Component Analysis, ICA = Independent Component Analysis, EoC = Ensemble of Classifiers.)

Refs.	Classifiers	Signals	Features	Defect classes	Datasets
Kankar et al. (2011)	SVM MLP SOM	real vibration signals, artificial defects 5 different speeds	kurtosis, skewness, std (from wavelet coefficients) number of loaders, speed	faultless bearing, IR fault, OR fault, RE fault, fault in all components	150 samples 10-fold cross-validation
Bhavaraju et al. (2010)	MLP SOM	real vibration signals, artificial defects, 5 different speeds	kurtosis, skewness, std (from wavelet coefficients), number of loaders, speed	faultless bearing, IR fault, OR fault, RE fault, fault in all components	150 samples 50% training, 50% test
Lazzerini and Volpi (2011)	ensembles of MLPs	real vibration signals, artificial defects, 10 different noise levels	FFT parameters (performed forward feature selection)	faultless bearing, indentation on IR, indentation on the roll, sandblasting of IR, unbalanced cage	12740 samples 70% training, 30% test (100 trials)
Volpi et al. (2010)	one-class CHC	real vibration signals, artificial defects	FFT parameters (performed forward feature selection)	faultless bearing unbalanced cage, indentation on IR (450 μm), sandblasting of IR, indentation on the roll (450 μm, 1.1 mm and 1.29 mm)	12740 samples training with "fautless" class, test with all classes (30 trials)
Widodo et al. (2009)	RVM SVM	real AE and vibration signals, artificial defects, considered only low-speeds (e.g., 20 and 80 rpm)	statistical, time- and frequency-domain features selected with PCA/ICA	faultless bearing, crack on IR (0.1 mm), spall on IR (0.6 mm), crack on OR (0.1 mm), spall on OR (0.7 mm), spalls on rollers (1 mm and 1.6 mm)	105 samples cross-validation
Cococcioni et al. (2009b)	LDC, QDC, MLP, RBF NN	real vibration signals, artificial defects, 10 different noise levels	FFT parameters (performed forward feature selection)	faultless bearing, indentation on IR, indentation on the roll (450 μm, 1.1 mm and 1.29 mm), sandblasting of IR, unbalanced cage	12740 samples 70% training, 30% test (100 trials)
Cococcioni et al. (2009a)	LDC, QDC, MLP, EoC	real vibration signals, artificial defects, 5 frequency ranges	FFT parameters (performed forward feature selection)	faultless bearing, indentation on IR, indentation on the roll (450 μm, 1.1 mm and 1.29 mm), sandblasting of IR, unbalanced cage	12740 samples 70% training, 30% test (10 trials)

Table 3
Survey of bearing fault diagnosis systems (continuation).

Refs.	Classifiers	Signals	Features	Defect classes	Datasets
Sreejith et al. (2008)	MLP	real vibration signals, artificial defects	time-domain features	faultless bearing, RE fault, OR fault, IR fault	80 samples from CWRU bearing data center (Case Western Reserve University) 60% training, 40% test
Teotrakool et al. (2008)	SVM	motor current signals, artificial defects, 4 different speeds	RMS values from wavelet packet coefficients (feature selection with GA)	faultless bearing *vs.* OR fault; faultless bearing *vs.* cage fault	–
Lei et al. (2008)	improved fuzzy c-means	real vibration signals from locomotive roller bearings	time-domain features	fautless bearing slight rub faults on OR, serious flaking faults on OR	150 samples for clustering
Sugumaran et al. (2008)	one-class & multi-class SVMs	real vibration signals, artificial defects, 3 different speeds	Kurtosis and statistical features (selected with a decision tree)	faultless bearing, OR fault, IR fault, OR fault + IR fault	–
Sugumaran et al. (2007)	SVM, proximal SVM	real vibration signals, artificial defects, 3 different speeds	Kurtosis and statistical features (selected with a decision tree)	faultless bearing, OR fault, IR fault, OR fault + IR fault	600 samples 83% training, 17% test
Abbasion et al. (2007)	SVM	real vibration signals, artificial defects	Weibull negative log-likelihood function of time-domain signals	faultless bearing, IR-drive fault, IR-fan fault, RE-drive fault, RE-fan fault, OR-drive fault, OR-fan fault	63 samples for test
Rojas and Nandi (2006)	SVM	real vibration signals, speeds	FFT parameters and statistical features	faultless bearing, worn bearing, OR fault, IR fault, RE fault, cage fault	1920 samples 50% training, 50% test
Guo et al. (2005)	MLP, SVM	real vibration signals, defects artificially introduced, 16 different speeds	statistical, frequency- and time-domain features selected with GA	fautless bearing, worn bearing, cage fault, IR fault, OR fault, RE fault	2880 samples 1/3 training, 1/3 test 1/3 validation,

Table 4
Survey of Bearing Fault Diagnosis Systems (continuation).

Refs.	Classifiers	Signals	Features	Defect classes	Datasets
Purushotham et al. (2005)	HMMs (one per class)	real vibration signals, artificial defects, considered multiple faults	MFCC coefficients (wavelet analysis)	2 faults on IR + 1 fault on RE, 2 faults on OR + 1 fault on RE, one fault in each component	training, test (4 different splits)
Samanta et al. (2004)	MLP, RBF NN, PNN	real vibration signals, artificial defects	statistical and time-domain features selected with GA	fautless bearing *vs.* faulty bearing (OR fault)	288 samples 50% training, 50% test
Samanta et al. (2003)	MLP SVM	real vibration signals, artificial defects	statistical and time-domain features selected with GA	fautless bearing *vs.* faulty bearing (OR fault)	288 samples 60% training, 40% test
Samanta and Al-Balushi (2003)	MLP	real vibration signals, artificial defects	statistical and time-domain features	fautless bearing *vs.* faulty bearing (OR fault)	200 samples 60% training, 40% test
Lou and Loparo (2004)	neuro-fuzzy	real vibration signals, artificial defects, 4 different load values	std of wavelet coefficients	fautless bearing, IR fault, RE fault	24 samples 50% training, 50% test
Jack and Nandi (2002)	SVM, MLP	real vibration signals, artificial defects, 16 different speeds	statistical and frequency-domain features selected with GA	faultless (brand new bearing, worn bearing) *vs.* faulty (OR fault, IR fault, RE fault, cage fault)	2880 samples 1/3 training, 1/3 test, 1/3 validation
Jack and Nandi (2001)	SVM, MLP	real vibration signals, artificial defects, 16 different speeds	statistical and frequency-domain features	faultless bearing, worn bearing, OR fault, IR fault, RE fault, cage fault	960 samples 1/3 training, 1/3 test, 1/3 validation
Alguindigue et al. (1993)	RNN, MLP	real vibration signals, real and artificial defects	high- and low-frequency features	faultless bearing, fault on IR, generalized fault on IR, fault on OR, generalized fault on OR, artificial fault of a ball, generalized fault of two balls, generalized fault of all the components	the test set contained samples from the training set

pits and spalls on the rolling surfaces (Tandon and Choudhury, 1999). In Tables 2–4, only localized defects are considered.

Some authors have worked with signals obtained from multiple rotational speeds. With exception of Widodo et al. (2009) and Sugumaran et al. (2007) – which developed a different system for each rotational speed –, the classifiers have been trained/tested with data corresponding to several speeds simultaneously (Guo et al., 2005; Jack and Nandi, 2002; Rojas and Nandi, 2006; Teotrakool et al., 2008), and, sometimes, the rotational speed is employed as input-feature (Bhavaraju et al., 2010; Kankar et al., 2011). However, these systems consider either non-rotating loads or no-load conditions, which means that the shock amplitudes are not affected if the rotational speed changes. So far, no work investigated the case where a same system has to deal with different speeds under a rotating load.

Regarding non-rotating loads, few works have considered signals obtained from multiple load conditions. While (Bhavaraju et al., 2010; Kankar et al., 2011) employed the number of loaders (which goes from 0 to 2) as input-feature, (Lou and Loparo, 2004) acquired vibration data from four load values (0, 1, 2 and 3 Horse Power (HP)). In both cases, the signals regarding the different load conditions were employed to train/test a same classifier.

3. Methodology

The objective of this work is to detect the presence or absence of bearing defects by taking into account six levels of noise, i.e., signal-to-noise ratio ranging from 40 to 5 db. Noise robustness is achieved through the incorporation of noisy data during the training phase, along with the fusion of different SVMs, each one is designed to deal with a specific noise configuration.

The BEAT simulator (Sassi et al., 2007) is employed to generate vibration signals coming from the operation of a ball bearing type SKF 1210 ETK9. The rotational speed is 1800 RPM, subjected to a non-rotating load of 3000 N. From the simulated data, the following time-domain indicators are calculated: RMS, peak, Kurtosis, crest factor, impulse factor and shape factor. As frequency-domain indicators, BPFO, BPFI, 2BSF, as well as their first two hamonics are calculated. It is worth noting that the frequency-domain indicators employed in this work are normalized with respect to the rotational speed. Regarding the time-domain indicators, they are independent of the rotational speed when the load is non-rotating.

The rest of this section describes the datasets and the performance evaluation methods employed in the experiments, as well as the Iterative Boolean Combination technique.

3.1. Datasets

Six noise configurations (nc = 1,2,3,4,5,6) are considered in this paper, as indicated in Table 5. For each noise configuration, there is a specific database, that is, $DB_{(nc)}$. Each sample in the databases is composed of a set of frequency and temporal indicators, plus the defect diameter, d_{def}, related to each bearing component, i.e., $d_{def(OR)}$, $d_{def(IR)}$ and $d_{def(RE)}$. Eight classes of defects are defined in Table 6. The flag = 1 indicates that there is a defect in the corresponding component, while flag = 0 indicates the absence of defect. For instance, class 6 corresponds to two different defects occuring simultaneously: one in the outer race, and another in the ball. For the non-defective components, d_{def} goes from 0 mm to 0.016 mm. Regarding the defective components, d_{def} goes from 0.017 mm to 2.8 mm.

Since the objective of this work is to indicate the presence or absence of a bearing defect, regardless its location, only two classes are considered, i.e, *faultless* and *faulty*. The *faultless* class corresponds to the class 0 (see Table 6) and, in order to have two balanced classes, the *faulty* class contains subsets of samples from classes 1 to 7. Table 7 presents the way the samples are partitioned.

Table 5
Noise configurations (nc).

nc	Training/validation	Test
1	40 db	40, 30, 20, 15, 10, 5 db
2	40, 30 db	40, 30, 20, 15, 10, 5 db
3	40, 30, 20 db	40, 30, 20, 15, 10, 5 db
4	40, 30, 20, 15 db	40, 30, 20, 15, 10, 5 db
5	40, 30, 20, 15 10 db	40, 30, 20, 15, 10, 5 db
6	40, 30, 20, 15, 10, 5 db	40, 30, 20, 15, 10, 5 db

Table 6
Classes of defects.

	OR	IR	RE
class 0	0	0	0
class 1	1	0	0
class 2	0	1	0
class 3	0	0	1
class 4	1	1	0
class 5	1	0	1
class 6	0	1	1
class 7	1	1	1

Table 7
Data partitioning for each $DB_{(nc)}$ ($1 \leqslant nc \leqslant 6$).

	Positive class	Negative class
trn	3500	3500
vld	1750	1750
tst (per noise level)	1750	1750

Table 8
ROC AUC on validation data.

System	AUC
$S_{(nc=1)}$	1
$S_{(nc=2)}$	0.9999
$S_{(nc=3)}$	0.9999
$S_{(nc=4)}$	0.9996
$S_{(nc=5)}$	0.9992
$S_{(nc=6)}$	0.9989

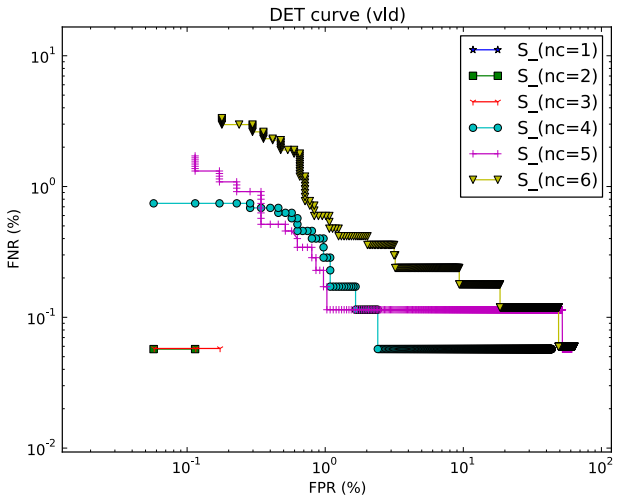

Fig. 3. DET curves of the selected systems $S_{(nc)}$, $1 \leqslant nc \leqslant 6$, using their respective validation sets (*vld*).

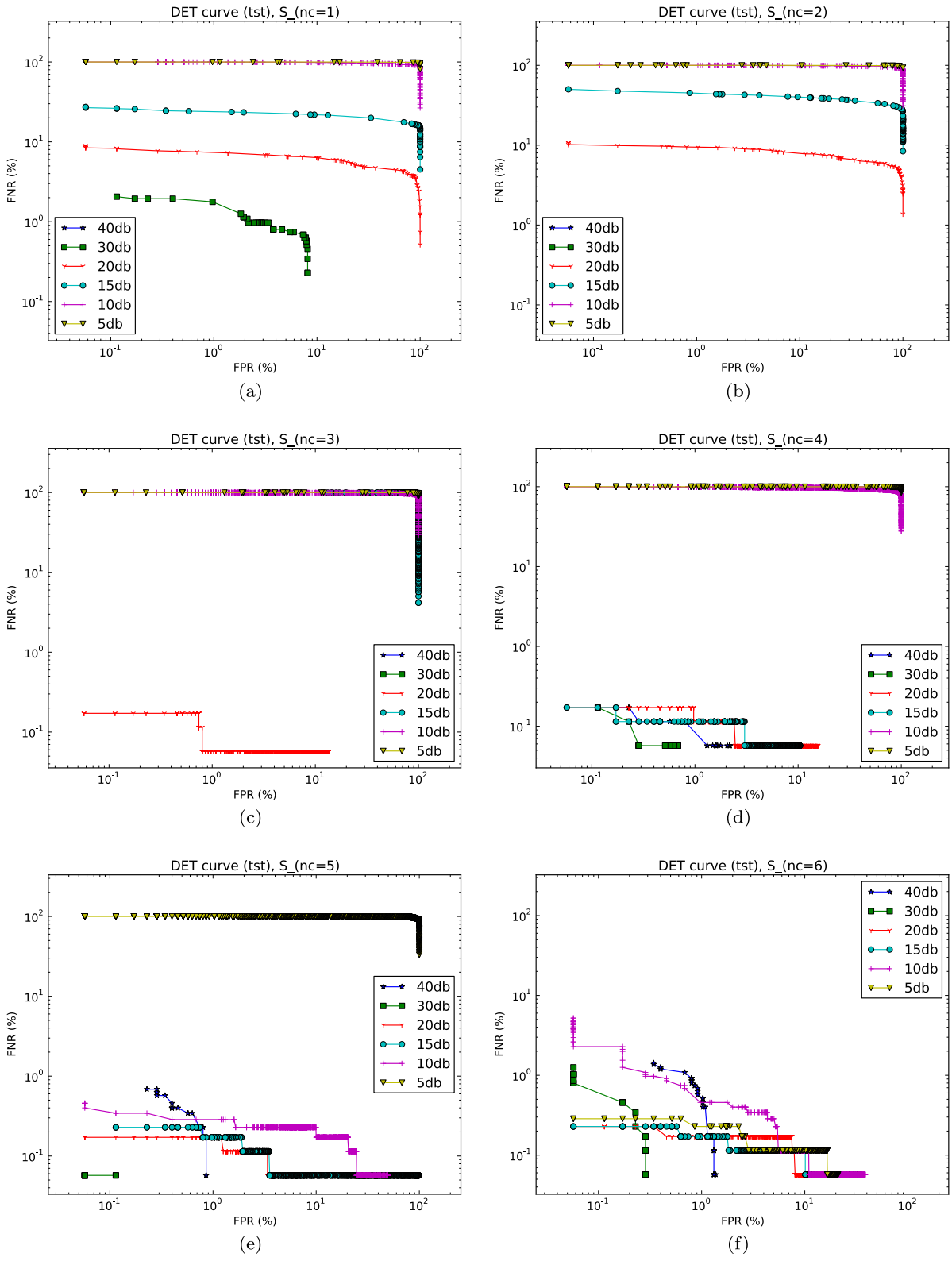

Fig. 4. DET curves of the selected systems $S_{(nc)}$, $1 \leqslant nc \leqslant 6$, using the test sets (*tst*).

3.2. Performance evaluation methods

The ROC (Receiving Operating Characteristics) curve – where the true positive rates (TPR) are plotted as function of the false positive rates (FPR) – is a powerful tool for evaluating, comparing and combining pattern recognition systems (Khreich et al., 2010). Several interesting properties can be observed from ROC curves. First, the AUC (Area Under Curve) is equivalent to the probability that the classifier will rank a randomly chosen positive sample higher than a randomly chosen negative sample. This measure is useful to characterize the system performance through a single scalar value. In addition, the optimal threshold for a given class

Table 9
Average EER $\pm\sigma$ (%) on test data over 10 trials.

tst	$S_{(nc=1)}$	$S_{(nc=2)}$	$S_{(nc=3)}$	$S_{(nc=4)}$	$S_{(nc=5)}$	$S_{(nc=6)}$
40 db	0.02 ± 0.02	0.05 ± 0.05	0.10 ± 0.07	0.17 ± 0.10	0.38 ± 0.17	0.57 ± 0.27
30 db	1.40 ± 1.09	0.00	0.01 ± 0.01	0.09 ± 0.07	0.14 ± 0.10	0.32 ± 0.16
20 db	5.38 ± 0.81	7.51 ± 1.01	0.07 ± 0.06	0.06 ± 0.06	0.08 ± 0.08	0.10 ± 0.06
15 db	20.98 ± 7.4	34.12 ± 6.7	ʁ	0.11 ± 0.03	0.16 ± 0.07	0.15 ± 0.06
10 db	ʁ	ʁ	ʁ	ʁ	0.27 ± 0.09	0.68 ± 0.26
5 db	ʁ	ʁ	ʁ	ʁ	ʁ	0.36 ± 0.08

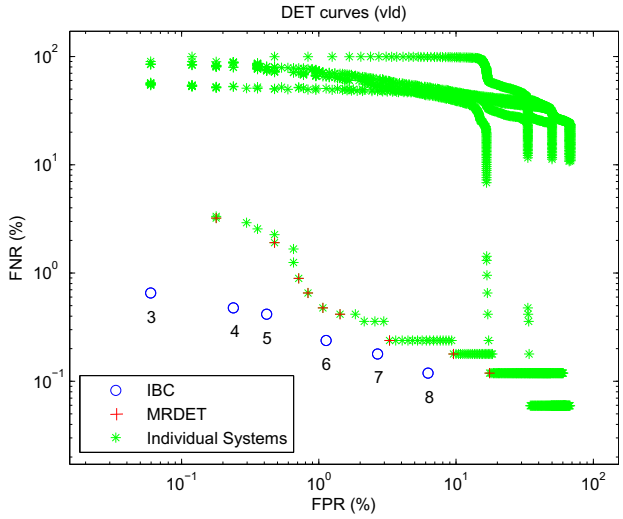

Fig. 5. DET curve obtained with IBC using a validation set containing all noise levels. The DET curves of the 6 individual systems and the Maximum Realizable DET curve (MRDET) are shown as well.

Table 10
Operating points of IBC DET curve.

Operating point	FNR (%)	FPR (%)	Average (%)
1	100.00	0.00	50.00
2	0.89	0.00	0.45
3	0.65	0.06	0.36
4	0.48	0.24	0.36
5	0.42	0.42	0.42
6	0.24	1.13	0.69
7	0.18	2.68	1.43
8	0.12	6.25	3.19
9	0.00	15.65	7.83
10	0.00	100.00	50.00

Table 11
Decision thresholds associated to the EER operating point.

Classifier	Threshold
c_1	0.9919
c_2	0.9816
c_3	0.9916
c_4	1.5587e−004
c_5	0.0095
c_6	0.0452

distribution lies on the ROC convex hull, which is defined as being the smallest convex set containing the points of the ROC curve. Finally, by taking into account several operating points, the ROC curve allows for analyzing these systems under different classification costs (Fawcett, 2006). A similar way to evaluate systems is through a DET (Detection Error Trade-off) curve, in which the false negative rates (FNR) are plotted as function of the false positive rates, generally, on a logarithmic scale.

In this work, ROC and DET curves are computed from the output probabilities provided by the classifiers. The validation set, *vld*, is used for this task. In order to test a given classifier, its corresponding ROC operating points (thresholds) are applied to the set, *tst*. Results on test are shown as well in terms of equal error rate (EER), which is obtained when the threshold is set to have the false negative rate approximately equal to the false positive rate.

3.3. Iterative Boolean Combination (IBC)

Ensembles of classifiers (EoCs) have been used to reduce error rates of many challenging pattern recognition problems. The motivation of using EoCs stems from the fact that different classifiers usually make different errors on different samples. When the response of a set of C classifiers is averaged, the variance contribution in the bias-variance decomposition decreases by $\frac{1}{C}$, resulting in a smaller classification error (Tumer and Ghosh, 1996).

It has been recently shown that the Iterative Boolean Combination (IBC) (Khreich et al., 2010) is an efficient technique for combining systems in the ROC space. IBC iteratively combines the ROC curves produced by different classifiers using all Boolean functions (i.e., $a \vee b$, $\neg a \vee b$, $a \vee \neg b$, $\neg(a \vee b)$, $a \wedge b$, $\neg\ a \wedge b$, $a \wedge \neg b$, $\neg(a \wedge b)$, $a \oplus b$, and $a \equiv b$), and does not require prior assumption that the classifiers are statistically independent. At each iteration, IBC selects the combinations that improve the Maximum Realizable ROC (MRROC) curve – i.e., the convex hull obtained from all individual ROC curves – and recombines them with the original ROC curves until the MRROC ceases to improve. For more details on the IBC technique, please refer to Algorithms 1 to 3 in Khreich et al. (2010).

4. Simulation results and discussions

Two main experiments are performed. In the first experiment, each database $DB_{(nc)}(1 \leqslant nc \leqslant 6)$ is employed in the generation of a baseline system $S_{(nc)}$. For each $DB_{(nc)}$:

- *trn* is used to train n different classifiers c_i, $1 \leqslant i \leqslant n$, by employing different SVM parameters;
- *vld* is used to validate each individual classifier c_i, by means of ROC curves, and select that one with the highest AUC. The select classifier is called $S_{(nc)}$;
- *tst* is used to test the performance of $S_{(nc)}$.

In the second experiment, the IBC technique (Khreich et al., 2010) is used to combine the best classifier of each noise configuration.

4.1. Experiment 1

The goal of the first experiment was to obtain the best baseline system for each one of the noise configurations defined in Table 5. For each database $DB_{(nc)}(1 \leqslant nc \leqslant 6)$, several SVMs were trained using the grid search technique (Chang and Lin, 2001), so that

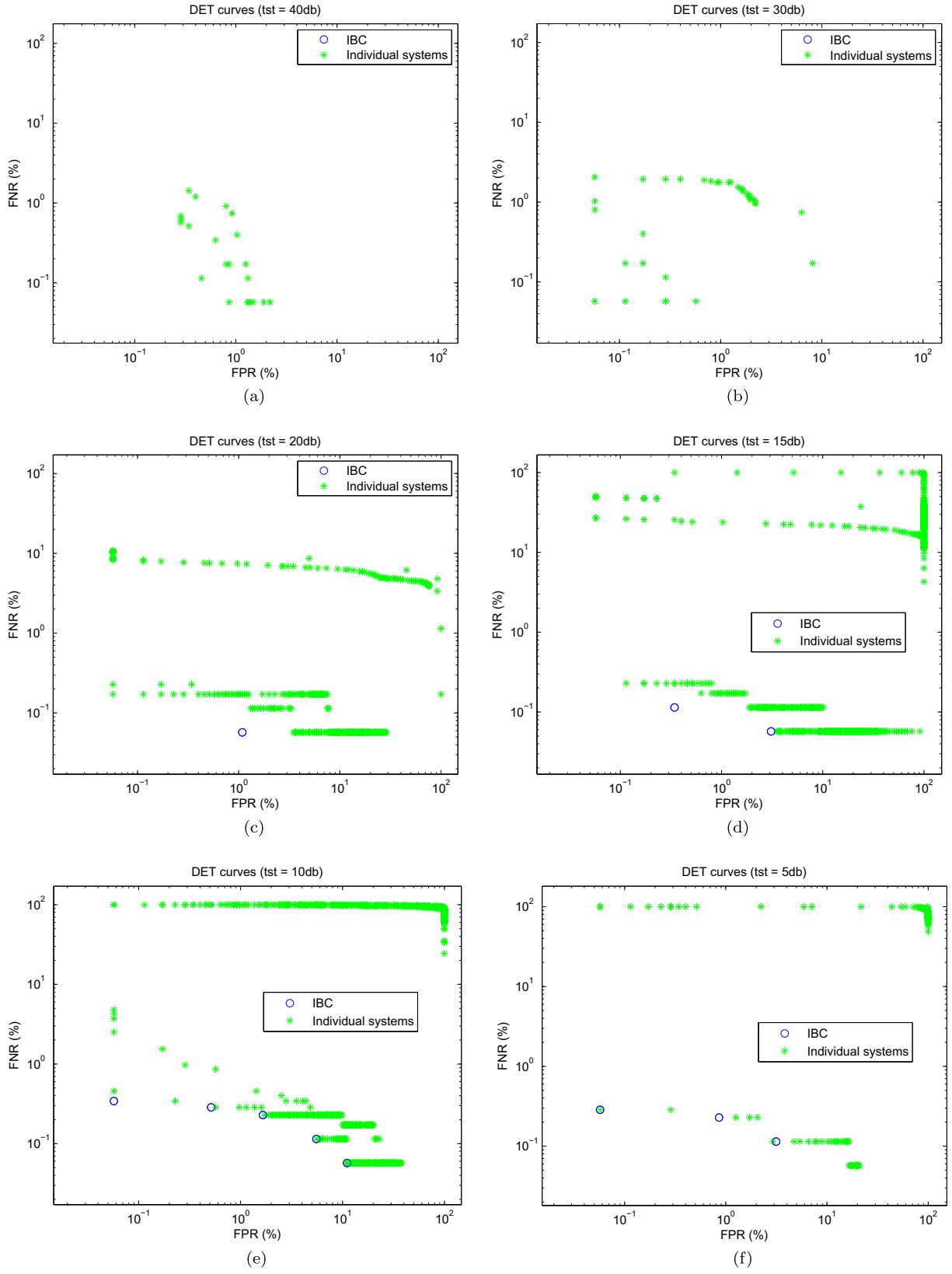

Fig. 6. DET curve obtained with IBC using the test sets (*tst*). The DET curves of the 6 individual systems are shown as well.

the SVM providing the highest AUC is selected. To train the SVMs with RBF kernel, the following values were employed: $\gamma = (\{2^{-4},2^{-3},2^{-2},2^{-1},2^{0}\}$ and $C = \{2^{-5},2^{-4},2^{-3},2^{-2},2^{-1},2^{0},2^{1},2^{2},2^{3},2^{4},2^{5}\}$.

Since the obtained ROC curves reached AUC close to 1, as indicated in Table 8, DET curves on a log–log scale are presented instead (see Fig. 3 (a)). Note that the curve representing system

Table 12
Average EER $\pm\sigma$ (%) on test data over 10 trials.

tst	IBC technique	Majority vote	Single best ($S_{(nc=6)}$)
40 db	0.04 ± 0.04	0.06 ± 0.06	0.57 ± 0.27
30 db	0.00	0.01 ± 0.02	0.32 ± 0.16
20 db	0.11 ± 0.06	0.06 ± 0.06	0.10 ± 0.06
15 db	0.11 ± 0.04	0.10 ± 0.04	0.15 ± 0.06
10 db	0.29 ± 0.09	ʁ	0.68 ± 0.26
5 db	0.33 ± 0.07	ʁ	0.36 ± 0.08

Table 13
Additional error rates $\pm\sigma$ (%) obtained with IBC over 10 trials.

tst	FNR (%)	FPR (%)	Average (%)
Expected FPR = 1%			
40 db	0.02 ± 0.02	10.70 ± 7.66	5.36
30 db	0.01 ± 0.01	8.61 ± 5.09	4.31
20 db	0.13 ± 0.15	5.34 ± 3.92	2.73
15 db	0.16 ± 0.09	6.48 ± 3.50	3.32
10 db	0.18 ± 0.11	8.94 ± 2.78	4.56
5 db	0.09 ± 0.10	21.87 ± 8.68	10.98
Expected FPR = 0.1%			
40 db	0.03 ± 0.03	0.37 ± 0.65	0.40
30 db	0.01 ± 0.02	0.15 ± 0.33	0.08
20 db	0.18 ± 0.13	0.01 ± 0.01	0.09
15 db	0.23 ± 0.10	0.09 ± 0.12	0.16
10 db	0.36 ± 0.13	1.35 ± 0.62	0.85
5 db	0.15 ± 0.08	5.74 ± 1.21	2.94
Expected FPR = 0.01%			
40 db	0.05 ± 0.04	0.10 ± 0.27	0.07
30 db	0.03 ± 0.04	0.03 ± 0.07	0.03
20 db	0.50 ± 0.31	0.00	0.02
15 db	0.56 ± 0.36	0.02 ± 0.03	0.29
10 db	1.60 ± 1.14	0.14 ± 0.10	0.87
5 db	0.28 ± 0.13	0.95 ± 0.44	0.61
Expected FPR = 0.001%			
40 db	0.09 ± 0.09	0.11 ± 0.26	0.10
30 db	0.07 ± 0.14	0.03 ± 0.07	0.05
20 db	0.63 ± 0.38	0.00	0.31
15 db	0.94 ± 0.66	0.01 ± 0.02	0.47
10 db	3.38 ± 3.20	0.06 ± 0.07	1.72
5 db	0.48 ± 0.29	0.51 ± 0.53	0.49
Expected FPR = 0.0001%			
40 db	0.09 ± 0.10	0.11 ± 0.26	0.10
30 db	0.10 ± 0.18	0.03 ± 0.07	0.06
20 db	0.66 ± 0.37	0.00	0.33
15 db	0.96 ± 0.68	0.01 ± 0.02	0.48
10 db	3.68 ± 3.56	0.06 ± 0.07	1.87
5 db	0.52 ± 0.34	0.46 ± 0.55	0.49

$S_{(nc=1)}$ does not appear in the graphic because a complete separation of both classes was obtained.

Fig. 4 shows the DET curves obtained on test data (*tst*) using the validation operating points. Observe that DET curves plotted in a same graphic are the results of a same system on different test data. Therefore, these curves are useful in order to analyse the robustness of each system regarding individual noise levels. It is worth noting that system $S_{(nc=1)}$ provided a complete class separation for 40 db (that's why the corresponding DET curve does not appear in the graphic), and, in a similar way, $S_{(nc=2)}$ and $S_{(nc=3)}$ provided a complete class separation for 40 db and 30 db.

Table 9 presents the average EER – as well as the standard deviation, σ – obtained for each noise level during test, over 10 trials. The symbol 'ʁ' indicates that the system has a random (or worse than random) behaviour for a given test set. A similar situation was observed in the work ofLazzerini and Volpi (2011), where classification accuracies of 50% or less were obtained for high levels of noise. As expected, the systems become more robust to higher noise levels as they are gradually incorporated to the training phase.

4.2. Experiment 2

In the second experiment, IBC was used to combine the best classifier of each noise configuration, found in the first experiment. For all classifiers, a same validation set containing all noise levels (i.e., 40, 30, 20, 15, 10 and 5 db) was employed. Since a high number of combinations is performed, the number thresholds per curve was limited to 500 (in the previous experiment, all validation scores were employed as thresholds).

Fig. 5 shows the DET curve obtained with IBC, along with the DET curves of the six systems employed during the combination process. Note that IBC improved the Maximum Realizable DET (MRDET) curve of the individual systems.

The operating points falling on the IBC curve are presented on Table 10. Each point is the result of a Boolean combination of different individual classifiers. For instance, the operating point 5, which gives the EER, corresponds to a boolean combination (*BC*) of all 6 classifiers (c_j,$1 \leqslant j \leqslant 6$), that is, $BC_{\{EER\}} = (c_1 \wedge c_2 \wedge c_3 \wedge c_4 \wedge c_5 \wedge c_6)$, using the decision thresholds indicated in Table 11.

It is worth noting that the AND rule emerges most of the time in the IBC curve of Fig. 5. In the ideal case, when the classifiers are conditionally independent, and their ROC/DET curves are proper and convex, the AND and OR combinations are proven to be optimal, providing a higher performance than the original ROC curves (Khreich et al., 2010). Indeed, the datasets employed to design the proposed system are independent, randomly generated by using the simulator BEAT.

Fig. 6 shows the DET curves obtained on test data using the IBC points indicated in Fig. 5, and Table 12 presents the average EER (over 10 trials) obtained with IBC, Majority vote and with the best single classifier. The Majority vote rule reached very low EER with respect to 40, 30, 20 and 15 db noise levels. On the other hand, a random behaviour was observed for 10 and 5 db noisy data. The reason is due to the fact that the majority of the individual classifers presents a random behaviour for high levels of noise.

Observe that IBC provided an improvement for almost all test datasets with respect to the single best classifier obtained in the previous experiment.

Finally, Table 13 presents additional results of IBC on test data, when the threshold is set in order to reach FPR (%) = {1,0.1,0.010.001,0.0001}. These intermediate points are obtained by using interpolation (Scott et al., 1998). Note that the FPR decreases at the expense of an FNR increasing. In practice, the trade-off between FPR and FNR can be adjusted by the operators according to the current error costs.

5. Conclusion

In this paper, a new system based on the fusion of classifiers in the ROC space was proposed in order to detect the presence of absence of bearing defects in noisy environments. Noise robustness was achieved through the incorporation of noisy vibration signals (ranging from 40 to 5 db) during the training phase, along with the Iterative Boolean Combination (IBC) of different SVMs, each one designed to deal with a specific noise configuration. In order to generate enough vibration signals, considering as well different defect dimensions, the BEAring Toolbox (BEAT) was employed.

Experiments performed using time- and frequency- domain indicators (i.e., RMS, peak, Kurtosis, crest factor, impulse factor, shape factor, BPFO, BPFI, 2BSF, and hamonics) indicated that the proposed system can significantly reduce the error rates, even in the presence of high levels of noise. Future work consist of validating the proposed strategy with real vibration signals.

References

Abbasion, S., Rafsanjani, A., Farshidianfar, A., & Irani, N. (2007). Rolling element bearings multi-fault classification based on the wavelet denoising and support vector machine. *Mechanical Systems and Signal Processing, 21*(7), 2933–2945.

Alguindigue, I., Loskiewicz-Buczak, A., & Uhrig, R. (1993). Monitoring and diagnosis of rolling element bearings using artificial neural networks. *IEEE Transactions on Industrial Electronics, 40*(2), 209–217.

Bhavaraju, K., Kankar, P., Sharma, S., & Harsha, S. (2010). A comparative study on bearings faults classification by artificial neural networks and self-organizing maps using wavelets (vol. 2, no. 5, pp. 1001–1008).

Case Western Reserve University, Bearing Data Center. <http://csegroups.case.edu/bearingdatacenter/>.

Chang, C., & Lin, C. (2001). LIBSVM: a library for Support Vector Machines. In <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.

Cococcioni, M., Lazzerini, B., & Volpi, S. (2009a). Automatic diagnosis of defects of rolling element bearings based on computational intelligence techniques. *International Conference on Intelligent Systems Design and Applications*, 970–975.

Cococcioni, M., Lazzerini, B., & Volpi S. (2009b). Rolling element bearing fault classification using soft computing techniques. In *IEEE international conference on systems, man and cybernetics, 2009* (pp. 4926–4931).

Elmaleeh, M., & Saad, N. (2008). Acoustic emission techniques for early detection of bearing faults using LabVIEW, in: *Fifth international symposium on mechatronics and its applications* (pp. 1–5).

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters, 27*, 861–874. ISSN 0167-8655.

Guo, H., Jack, L., & Nandi, A. (2005). Feature generation using genetic programming with application to fault classification. *IEEE Transactions on Systems, Man, and Cybernetics, Part B, 35*(1), 89–99.

Jack, L., & Nandi, A. (2001). Support vector machines for detection and characterization of rolling element bearing faults. *Journal of Mechanical Engineering Science, 215*(9), 1065–1074.

Jack, L., & Nandi, A. (2002). Fault detection using support vector machines and artificial neural networks, augmented by genetic algorithms. *Mechanical Systems and Signal Processing, 16*(2–3), 373–390.

Kankar, P., Sharma, S., & Harsha, S. (2011). Fault diagnosis of ball bearings using continuous wavelet transform. *Applied Soft Computing, 11*, 2300–2312.

Khreich, W., Granger, E., Miri, A., & Sabourin, R. (2010). Iterative Boolean Combination of classifiers in the ROC space: An application to anomaly detection with HMMs. *Pattern Recognition, 43*, 2732–2752. ISSN 0031-320.

Lazzerini, B., & Volpi, S. (2011). Classifier ensembles to improve the robustness to noise of bearing fault diagnosis. In *Pattern Analysis and Applications* (pp. 1–17).

Lei, Y., He, Z., Zi, Y., & Hu, Q. (2008). Fault diagnosis of rotating machinery based on a new hybrid clustering algorithm. *The International Journal of Advanced Manufacturing Technology, 35*, 968–977. ISSN 0268-3768.

Liang, S., Hecker, R., & Landers, R. (2004). Machining process monitoring and control: The state-of-the-art. *Journal of Manufacturing Science and Engineering, 126*(2), 297–310.

Lou, X., & Loparo, K. (2004). Bearing fault diagnosis based on wavelet transform and fuzzy inference. *Mechanical Systems and Signal Processing, 18*(5), 1077–1095.

Purushotham, V., Narayanan, S., & Prasad, S. A. (2005). Multi-fault diagnosis of rolling bearing elements using wavelet analysis and hidden Markov model based fault recognition. *NDT & E International, 38*(8), 654–664.

Rojas, A., & Nandi, A. (2006). Practical scheme for fast detection and classification of rolling-element bearing faults using support vector machines. *Mechanical Systems and Signal Processing, 20*(7), 1523–1536.

Samanta, B., & Al-Balushi, K. (2003). Artificial neural network based fault diagnostics of rolling element bearings using time-domain features. *Mechanical Systems and Signal Processing, 17*(2), 317–328.

Samanta, B., Al-Balushi, K., & Al-Araimi, S. (2003). Artificial neural networks and support vector machines with genetic algorithm for bearing fault detection. *Engineering Applications of Artificial Intelligence, 16*(7-8), 657–665.

Samanta, B., Al-Balushi, K., & Al-Araimi, S. (2004). Bearing fault detection using artificial neural networks and genetic algorithm. *EURASIP Journal on Applied Signal Processing*, 366–377.

Sassi, S., Badri, B., & Thomas, M. (2007). A numerical model to predict damaged bearing vibrations. *Journal of Vibration and Control, 13*(11), 1603–1628.

Sassi, S., Badri, B., & Thomas, M. (2008). Tracking surface degradation of ball bearings by means of new time domain scalar descriptors. *International Journal of COMADEM, 11*(3), 36–45.

Scott, M., Niranjan, M., & Prager, R. (1998). Realisable classifiers: Improving operating performance on variable cost problems.

Sreejith, B., Verma, A., & Srividya, A. (2008). Fault diagnosis of rolling element bearing using time-domain features and neural networks. In *Third international conference on industrial and information systems* (pp. 1–6).

Sugumaran, V., Muralidharan, V., & Ramachandran, K. (2007). Feature selection using decision tree and classification through proximal support vector machine for fault diagnostics of roller bearing. *Mechanical Systems and Signal Processing, 21*(2), 930–942.

Sugumaran, V., Sabareesh, G., & Ramachandran, K. (2008). Fault diagnostics of roller bearing using kernel based neighborhood score multi-class support vector machine. *Expert Systems with Applications, 34*(4), 3090–3098.

Tandon, N., & Choudhury, A. (1999). A review of vibration and acoustic measurement methods for the detection of defects in rolling element bearings. *Tribology International, 32*(8), 469–480.

Teotrakool, K., Devaney, M., & Eren, L. (2008). Bearing fault detection in adjustable speed drives via a support vector machine with feature selection using a genetic algorithm. In *IEEE instrumentation and measurement technology conference* (pp. 1129 –1133).

Thomas, M. (2011). *Fiabilité, maintenance prédictive et vibration des machines*. 9782760533578. Presses de l'Université du Québec (D3357).

Tumer, K., & Ghosh, J. (1996). Analysis of decision boundaries in linearly combined neural classifiers. *Pattern Recognition, 29*(2), 341–348.

Volpi, S., Cococcioni, M., Lazzerini, B., & Stefanescu, D. (2010). Rolling element bearing diagnosis using convex hull. In *International joint conference on neural networks* (pp. 1–8).

Widodo, A., Kim, E., Son, J., Yang, B., Tan, A., Gu, D., et al. (2009). Fault diagnosis of low speed bearing based on relevance vector machine and support vector machine. *Expert Systems with Applications, 36*(3, Part 2), 7252–7261.