

Heterogeneous Feature Models and Feature Selection Applied to Bearing Fault Diagnosis

Thomas W. Rauber, *Member, IEEE*, Francisco de Assis Boldt, and Flávio Miguel Varejão

Abstract—Distinct feature extraction methods are simultaneously used to describe bearing faults. This approach produces a large number of heterogeneous features that augment discriminative information but, at the same time, create irrelevant and redundant information. A subsequent feature selection phase filters out the most discriminative features. The feature models are based on the complex envelope spectrum, statistical time- and frequency-domain parameters, and wavelet packet analysis. Feature selection is achieved by conventional search of the feature space by greedy methods. For the final fault diagnosis, the k -nearest neighbor classifier, feedforward net, and support vector machine are used. Performance criteria are the estimated error rate and the area under the receiver operating characteristic curve (AUC-ROC). Experimental results are shown for the Case Western Reserve University Bearing Data. The main contribution of this paper is the strategy to use several different feature models in a single pool, together with feature selection to optimize the fault diagnosis system. Moreover, robust performance estimation techniques usually not encountered in the context of engineering are employed.

Index Terms—Case Western Reserve University (CWRU) Bearing Fault Database, fault diagnosis, feature extraction, feature selection.

I. INTRODUCTION

THE detection of bearing faults is an important engineering task, particularly in the context of electric machines [1]. A common engineering approach in model-free [2] diagnosis is to define a single signal feature extraction model, together with a classifier methodology and to present cross-validation experimental results. The considered applied research in this field frequently lacks soundness in the pattern recognition part of the system. To motivate this work and to support this hypothesis, papers were analyzed that use the Case Western Reserve University (CWRU) Bearing Data [3] as a testbed.

Statistical features are exclusively used in [4]–[8]. The work of Sun *et al.* in [9] and Sreejith *et al.* in [10] use complex envelope features together with the wavelet packet analysis, which constitutes the principal feature model in the following publications: [10]–[16]. It is postulated here that there is no necessity to restrict the discriminatory description of a fault

diagnosis system to a single feature model, thus avoiding the risk to miss important aspects of the data. It is better to produce as much information as possible about the process condition, even contaminated with noise and filter out the best features in a posterior information processing step.

Feature selection [17] is such an important technique in computational intelligence. Its aim is to reduce dimensionality and eventually improve performance parameters of the system such as accuracy. No feature selection at all is used in [5], [6], [9], [10], [12]–[16], [18], and [19] or only univariate methods, i.e., ranking features individually, such as, e.g., [11] (factor analysis) and [4] (Fisher score). Here, search algorithms are applied that are multivariate, analyzing a *set* of candidate features, thus considerably avoiding redundancy in the final feature model. The nature of the features does not play an important role, as long as they are powerful separators of the fault classes. Therefore, a candidate pool can be assembled where features can originate from quite distinct models such as, for instance, in this work, statistical, wavelets, and envelopes.

Considering the final attribution of the vibration pattern to a machine condition, it can be observed that no classifier at all is used in [5], [12], [13], and [18]; only the visual inspection of peaks in frequency graphs suggests the discriminative power of the method. In general, feedforward neural nets and support vector machines (SVMs) are the dominant techniques in contemporary work.

Training, validation, and test data splitting is very important to obtain reliable performance criteria for the fault classifier. Simple one-time train-test splitting or no information at all about the pattern division can be observed in [4], [6]–[11], [14]–[16], and [18]. None of these papers use a statistically more significant model validation, such as leave-one-out (LOO) or K-fold cross validation (CV). Moreover, the estimated accuracy is the only performance criterion used, although other criteria are interesting, for instance, the area under the receiver operating characteristic curve (AUC-ROC) [20], particularly when the classes are considerably unbalanced with respect to the number of their training samples.

The following research on model-free bearing fault detection does not use the CWRU data but is similar to this work with respect to the objectives and methodologies. In [21], statistical features describe the bearing fault; linear discriminant analysis is used to rank the features in a preprocessing step in a univariate, bivariate, and trivariate combination; a curvilinear component analysis is used for feature extraction; and a multilayer perceptron (MLP) provides the final process condition. The faults of a wound-rotor induction machine are characterized by wavelet coefficients in [22], and a time–frequency feature

Manuscript received October 1, 2013; revised January 10, 2014 and April 3, 2014; accepted May 3, 2014. Date of publication May 30, 2014; date of current version December 19, 2014.

The authors are with the Computer Science Department, Universidade Federal do Espírito Santo, 29075-91 Vitória, Brazil (e-mail: thomas@inf.ufes.br; fassis@inf.ufes.br; fvarejao@inf.ufes.br).

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/TIE.2014.2327589

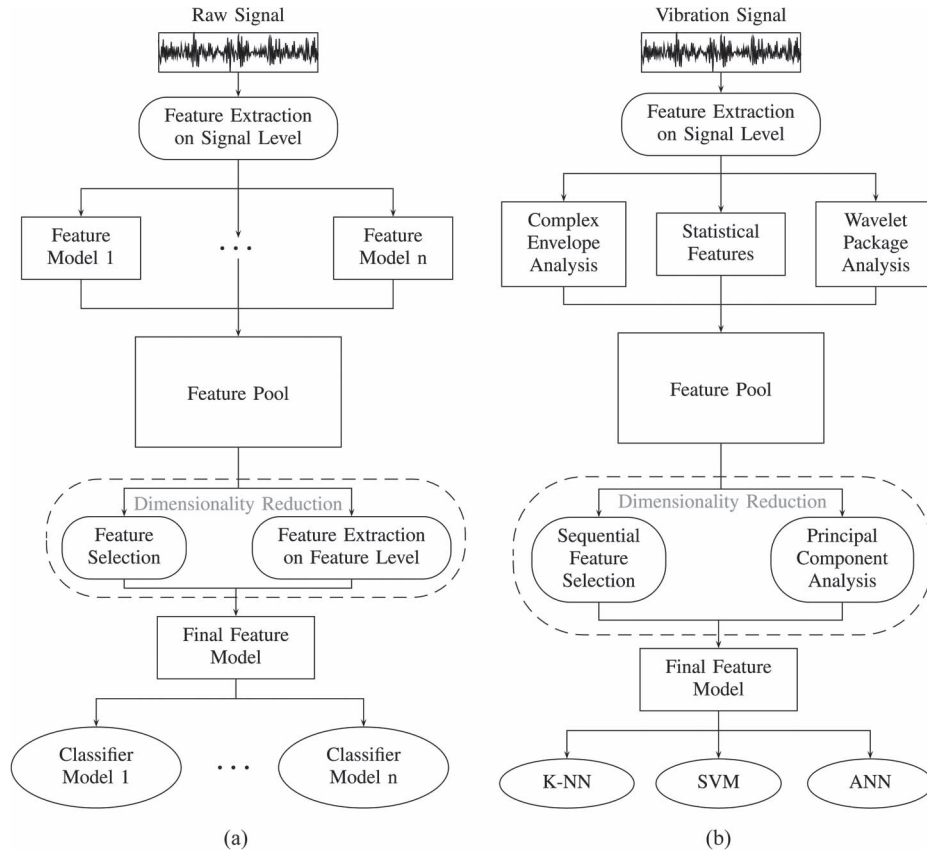


Fig. 1. Computational intelligence framework for bearing fault diagnosis. (a) Generic model. (b) Instantiated model.

model is chosen in [23] by designing particular kernels for the given classification task. Moreover, sequential forward feature selection is used. The multidimensional selection criterion is the average pairwise mutual correlation among candidate features and the already selected features. He *et al.* in [24] propose a two-stage system for bearing fault diagnosis. First, complex envelope features are used for outer race separation; then, empirical mode decomposition by the Hilbert–Huang transform generates features that are used for the remaining fault classes.

The innovative idea of this paper is to consider the final feature model as the result of an information processing sequence. It does not matter where the discriminative information comes from, as long as it improves the classification of the machine conditions. Traditional fault diagnosis systems limit the process description to a single feature model, for instance, wavelet coefficients. More than one simultaneous feature model is not considered in general.

Obviously, this approach does not work, if the good information is not filtered out by the subsequent feature selection stage. The combination of several feature models plus feature selection seems a quite obvious method to improve classification performance but is not being employed in practice.

The rest of this paper is organized in the following manner: Section II presents the proposed framework to perform the bearing fault diagnosis and describes the three feature models used in this research that are pooled into the global candidate set. In Section III, the important step of feature selection is discussed. Section IV presents the classifier models and performance estimation methods used in this work. The description

of the experimental workbench in which the bearing faults are artificially created, the different machine and fault conditions that define the classes to be recognized, together with experimental results for fault diagnosis of the CWRU bearing data are presented in Section V, and finally, conclusions are drawn in Section VI.

II. FEATURE MODELS

The main hypothesis of this work is that higher discriminative power of the fault detection system can be achieved if as much information as possible is extracted from the signal and a subsequent feature selection reduces the final number of features to a reasonable amount. This means that various feature extraction paradigms can be combined in parallel to create a quite heterogeneous pool of candidate features from which the final set is obtained by feature selection. None of the previously analyzed publications propose a fusion of features that stem from completely different signal feature extraction methods. The experimental results suggest that this strategy of pooling, followed by selection in general, improves the performance of the fault diagnosis. Fig. 1 illustrates the general framework to perform the bearing fault diagnosis. The sequence of information processing steps is as follows.

- 1) *Signal feature extraction:* Use the available sensors attached to the machine to acquire the raw signal in the time domain. From the raw signal, considering a feature model, extract the feature vector within this model. Do this for all available models.

- 2) *Feature pooling*: Assemble a global feature vector that contains as much information as possible about the machine condition. Reduce the number of features by a recombination of the existing ones (extraction) or by retaining only a subset (selection), cf., the following two steps.
- 3) *Feature extraction on the feature level*: From the existing feature vector, extract new features. Usually, the dimension is considerably reduced. The new features are abstract descriptions of the machine condition. The most prominent linear method is principal component analysis (PCA).
- 4) *Feature selection*: Either discard or preserve existing features to form a subset of the existing features. The main goal is dimensionality reduction and the increase in discriminatory power of the net feature set.
- 5) *Classification*: Define a single- or a multiple-classifier model and estimate its performance.

The right-hand side in Fig. 1 shows the specialization of the framework used in this work, the specific feature models, selection techniques, and classifiers.

The main contribution of this paper is the simultaneous use of distinct types of features and the posterior phase of feature selection. This strategy ensures that no important information is omitted during the extraction phase and that no irrelevant information from the extraction phase is preserved. Representative feature models mainly observed in the considered literature were used in this work, i.e., statistical features from the time and frequency domains, wavelet packet energy, and complex envelope magnitudes.

A. Statistical Features

As representative sets, those features used in [11] were chosen, cf., Tables I and II. Table I presents the definition of ten statistical features in the time domain: root mean square (RMS), square root of the amplitude (SRA), kurtosis value (KV), skewness value (SV), peak-to-peak value (PPV), crest factor (CF), impulse factor (IF), margin factor (MF), shape factor (SF), and kurtosis factor (KF). Table II presents the definition of three statistical features in the frequency domain: frequency center (FC), RMS frequency (RMSF), and root variance frequency (RVF). The total number of statistical features is $(10 + 3) \times 2 = 26$, i.e., the ten statistical features of the time domain, the three of the frequency domain, taken at both the drive end (DE) and fan end (FE) of the motor housing of the CWRU testbed.

B. Complex Envelope Analysis

There are four characteristic frequencies at which faults can occur [25]. Knowing the shaft rotational frequency F_S , they are the fundamental cage frequency F_C (not present in the CWRU database), ball-pass inner-raceway frequency F_{BPI} , ball-pass outer-raceway frequency F_{BPO} , and the ball-spin frequency F_B . For a detailed description of the fault frequency model, refer, e.g., to [24]. A detailed description of the CWRU testbed is provided in [11].

TABLE I
TIME-DOMAIN STATISTICAL FEATURE SET OF THE VIBRATION SIGNAL
COMPOSED OF N ACCELERATION AMPLITUDES x_i

$X_{\text{rms}} = \left(\frac{1}{N} \sum_{i=1}^N x_i^2 \right)^{1/2}$	$X_{\text{sra}} = \left(\frac{1}{N} \sum_{i=1}^N \sqrt{ x_i } \right)^2$
$X_{\text{kv}} = \frac{1}{N} \sum_{i=1}^N \left(\frac{x_i - \bar{x}}{\sigma} \right)^4$	$X_{\text{sv}} = \frac{1}{N} \sum_{i=1}^N \left(\frac{x_i - \bar{x}}{\sigma} \right)^3$
$X_{\text{ppv}} = \max(x_i) - \min(x_i)$	$X_{\text{cf}} = \frac{\max(x_i)}{\left(\frac{1}{N} \sum_{i=1}^N x_i^2 \right)^{1/2}}$
$X_{\text{if}} = \frac{\max(x_i)}{\frac{1}{N} \sum_{i=1}^N x_i }$	$X_{\text{mf}} = \frac{\max(x_i)}{\left(\frac{1}{N} \sum_{i=1}^N \sqrt{ x_i } \right)^2}$
$X_{\text{sf}} = \frac{\max(x_i)}{\left(\frac{1}{N} \sum_{i=1}^N x_i^2 \right)^{1/2}}$	$X_{\text{kf}} = \frac{\frac{1}{N} \sum_{i=1}^N \left(\frac{x_i - \bar{x}}{\sigma} \right)^4}{\left(\frac{1}{N} \sum_{i=1}^N x_i^2 \right)^2}$

TABLE II
FREQUENCY-DOMAIN STATISTICAL FEATURE SET OF THE VIBRATION
SIGNAL COMPOSED OF N FREQUENCY AMPLITUDES f_i

$X_{\text{fc}} = \frac{1}{N} \sum_{i=1}^N f_i$	$X_{\text{rmsf}} = \left(\frac{1}{N} \sum_{i=1}^N f_i^2 \right)^{1/2}$
$X_{\text{rvf}} = \left(\frac{1}{N} \sum_{i=1}^N (f_i - X_{\text{fc}})^2 \right)^{1/2}$	

The signal envelope can be calculated by virtue of the Hilbert transform. Given a signal $h(t)$ in the time domain, the Hilbert transform is the convolution of $h(t)$ with the signal $1/\pi t$, producing a new signal in the time domain, i.e.,

$$\tilde{h}(t) := \mathcal{H}\{h(t)\} := h(t) * \frac{1}{\pi t} = \frac{1}{\pi} \int_{-\infty}^{\infty} h(\tau) \frac{d\tau}{t - \tau}. \quad (1)$$

The *analytic signal*, i.e., $h_a(t)$, which is a complex signal in the time domain, composed of the original signal $h(t)$ and its Hilbert transform $\tilde{h}(t)$ in quadrature, is defined as

$$h_a(t) := h(t) + i\tilde{h}(t). \quad (2)$$

The analysis steps for the calculus of the bearing defect frequencies spectrum can be resumed as follows:

- 1) high-pass filtering of the raw signal to eliminate the influence of slow vibrations, producing the signal $h(t)$;
- 2) calculus of the analytic signal $h_a(t)$ of the original signal $h(t)$;
- 3) Fourier transform of the analytic signal $\mathcal{F}\{h_a(t)\}$;
- 4) analysis of its spectrum $|\mathcal{F}\{h_a(t)\}|$.

For the CWRU database, the sensor at the DE, although with less confidence, can detect the faults at the FE; hence, the number of features duplicates (cross detection). For each of the characteristic fault frequencies F_{BPI} , F_{BPO} , F_B , and up to the sixth harmonic of each of these frequencies, the 1% narrowband RMS energy (i.e., X_{rmsf} in Table II) was

calculated. Hence, $2 \times 3 \times 2 \times 6 = 72$ envelope features are obtained, i.e., acquisition position (FE and DE) \times characteristic frequencies \times cross detection \times harmonics.

C. Wavelet Packet Analysis

Dual-domain analysis methodologies that extract features from the time–frequency domain are represented in this work by wavelets [26], more specifically by the wavelet packet analysis, which allows a more flexible decomposition guided by information theory. Work describing the CWRU data by wavelet packets is found in [11]–[14] and [16]. Here, the procedure proposed in [11] is used. The mother wavelet is Daubechies 4, and refining is done down to the fourth decomposition level. A 1-D time-domain vibration signal of S samples is considered. With a tree depth of j , 2^j final leaves $W_{j,0}, \dots, W_{j,2^j-1}$ were obtained. Each has approximately $S/2^j$ wavelet coefficients.¹ The features from the final 2^j leaf nodes of the tree are the respective percentages of the energy of each leaf. Let $c_{j,n}^s$, $s = 0, \dots, S/2^j - 1$ be the $S/2^j$ wavelet coefficients of leaf node n at tree depth j , $n = 0, \dots, 2^j - 1$. The energy of the n th node is

$$E_j(n) = \sum_{s=0}^{S/2^j-1} [c_{j,n}^s]^2. \quad (3)$$

Then, the n th wavelet packet feature is

$$x_n = \frac{E_j(n)}{\sum_{m=0}^{2^j-1} E_j(m)}. \quad (4)$$

With a tree depth of $j = 4$, 16 final leaves $W_{4,0}, \dots, W_{4,15}$ were obtained and, consequently, 16 features $\{x_n | n = 0, \dots, 15\}$ with $\sum_{n=0}^{15} x_n = 1$. Since calculation is done at both the DE and the FE (cross detection), this number is duplicated to 32.

D. Feature Pooling and Dimensionality Reduction

After the acquisition of the different feature models, it is possible to merge them into a common pool. The cardinality of the feature pool is 130, composed of 26 statistical features, 72 complex envelope features, and 32 wavelet packet features. The features within this global set contain redundancy and noise. A consequent step is the reduction of the dimension of the feature vector to its final value. At this point, a possible dimensionality reduction technique could be feature extraction on the feature level, e.g., PCA or key performance indicator prediction and diagnosis [27]. In [28], the Tennessee Eastman chemical process simulation software [29] was employed, since it provides a research testbed for reproducible and comparable experiments. This simulator was also used as a benchmark in a model-based fault diagnosis system [30]. The feature extraction techniques in [28] were PCA, partial least squares, independent component analysis, Fisher discriminant analysis, and subspace-aided approach. These techniques belong to the

feature extraction on the feature-level methods in Fig. 1. To distinguish the methodology of this work, a conceptual and experimental comparison with PCA as a representative is included in the experiments.

III. FEATURE SELECTION

A feature selection algorithm is basically composed of a selection criterion and a search strategy. A subset of d from all D available features are preserved. Excellent review papers of feature selection methodologies are [17] or [31]. The *wrapper* approach in feature selection consists in taking the estimated performance of a classifier as the proper feature selection criterion. The performance criterion in the overwhelming part of past and contemporary work is the estimated accuracy of the classifier. This means that those features during the search are labeled as good that minimize the estimated error. In the context of fault diagnosis, an alternative performance criterion, the AUC-ROC in Section IV-B, will also be motivated.

The *filter* approach uses different criteria to judge a feature set or judge the performance of the classifier. Usually, the benefit of a selection filter is its speed, while its drawback is a possible performance inferior compared with that of a wrapper.

Since an exhaustive search is computationally unfeasible, suboptimal search strategies are chosen. The best feature (BF) search is simply evaluating a selection criterion $J(\{x_j\})$ for each feature x_j , $j = 1, \dots, D$, ordering the features in descending order relative to J and setting the selected set X_d to the first d features of the ordered set. BF is fast but ignores the multidimensionality of the problem. In the considered context of the fault classification of the CWRU bearing data, Mahamad and Hiyama in [4] used a univariate, distance-based selection criterion.

Sequential forward selection (SFS) [17] starts with an empty set and then tests each candidate together with the already-selected features. The feature that performed best is included in the already-selected set. The algorithm stops when the desired number of d features is reached.

Sequential backward selection (SBS) [17] starts with all D available features being selected and then discards one feature at a time, until $D - d$ features have been deleted, i.e., d features have been retained as the selected ones. *Floating* techniques [32] allow backtracking for an arbitrary number of times as long as the quality criterion J is improving. As representatives for a complete sequential search strategy algorithm, sequential floating forward search (SFFS) and sequential floating backward search (SFBS) are used.

Recent approaches that try to combine wrapper and filter methods are not considered, since they would deviate attention from the innovative aspects of the proposed methodology.

IV. CLASSIFICATION AND PERFORMANCE ESTIMATION

Frequently, in CWRU-related work, the choice of a particular classifier architecture that decides the machine condition is not sufficiently motivated or its description is incomplete, which does not permit its implementation to compare results. Moreover, the way the data are split to evaluate the performance

¹Exactly, if the original signal length S were a power of two.

frequently leads to overoptimistic results. To maintain scientific rigor and allow comparison, it is necessary to use statistically sound and robust methods [33].

A. CV Techniques

A simple unique division of the available machine condition patterns into $x\%$ training data, and $100 - x\%$ test data potentially introduces a considerable bias for the estimated performance. A good compromise between statistical significance and computational complexity is to use K-fold CV, where the data set is divided into K subsets; each subset is used once for test and $K - 1$ times for training. When the training time of the classifier is not excessive, the LOO CV can be used. LOO is a special case of K-fold CV where the total data set has N patterns; hence, $K = N$. LOO gives a more reliable estimate than simple train-test data division, since implicitly, each available pattern is tested. In the experiments, the LOO and K-fold CV are used, depending on the training costs of the classifier architecture.

B. Performance Criteria

Two criteria, judged as most appropriate for this application, are employed, namely, accuracy and AUC-ROC. The accuracy or, equivalently, its complement, i.e., the error rate, is most easily understandable to a human being. It constitutes the vast majority among the criteria in the considered literature and is also used in this research. When only two classes (positive and negative) are considered and the classifier returns a continuous score, ideally a probability estimate of the two classes, it is possible to use the AUC-ROC [20], [33] as an alternative performance criterion. It can handle unbalanced classes in which negative class examples are usually much more common than positive ones, which is a situation frequently encountered in machine fault diagnosis, where the normal operation condition is dominant.

C. Classifier Architectures

The k -nearest neighbor classifier [34] represents the category of simple nonparametric methods. As a representative classifier from the area of artificial neural networks, MLP is proposed, trained by the scaled conjugate gradient algorithm [35]. The SVM [36] classification architecture has been extensively used during the last decade in many distinct domains, and also in machine fault diagnosis. It is currently considered one of the most powerful methods in machine learning and, therefore, is also represented in this work.

V. EXPERIMENTAL RESULTS

As a testbed, vibration signals provided and publicly available from CWRU [3] were used. This choice is motivated by the fact that a pattern database that is accessible to the research community allows a fair comparison of the performance of the proposed algorithms. The database is organized as MATLAB/Octave processable vibration signal files, with all

TABLE III
CLASS DISTRIBUTION AND DESCRIPTION, WHERE DE MEANS
DRIVE END AND FE MEANS FAN END

Class	Name	Sam- ples	Distri- bution	Description
1	Ball_DE_007	120	5.23%	0.007" ball fault, DE
2	Ball_FE_007	60	2.61%	0.007" ball fault, FE
3	Ball_DE_014	120	5.23%	0.014" ball fault, DE
4	Ball_FE_014	60	2.61%	0.014" ball fault, FE
5	Ball_DE_021	120	5.23%	0.021" ball fault, DE
6	Ball_FE_021	60	2.61%	0.021" ball fault, FE
7	InnerRace_DE_007	120	5.23%	0.007" inner race fault, DE
8	InnerRace_FE_007	60	2.61%	0.007" inner race fault, FE
9	InnerRace_DE_014	120	5.23%	0.014" inner race fault, DE
10	InnerRace_FE_014	60	2.61%	0.014" inner race fault, FE
11	InnerRace_DE_021	120	5.23%	0.021" inner race fault, DE
12	InnerRace_FE_021	60	2.61%	0.021" inner race fault, FE
13	Normal	60	2.61%	Normal.
14	OuterRace_DE_007	360	15.59%	0.007" outer race fault, DE
15	OuterRace_FE_007	180	7.84%	0.007" outer race fault, FE
16	OuterRace_FE_014	75	3.27%	0.014" outer race fault, DE
17	OuterRace_DE_014	120	5.23%	0.014" outer race fault, FE
18	OuterRace_DE_021	360	15.59%	0.021" outer race fault, DE
19	OuterRace_FE_021	60	2.61%	0.021" outer race fault, FE

necessary parameters attached, needed to calculate the feature models described in Section II. The machine condition differs with respect to the fault severity, bearing manufacturer, motor load, sensor position, and acquisition frequency and duration. The main components are the motor with a maximum power of 2 hp, the torque encoder, and a dynamometer for load simulation.

A. Condition Classes

Table III lists the machine condition classes defined for the experiments. Note that the number of classes is much more extensive than that found usually in work related to the CWRU data. It was tried to distinguish among the different bearing fault locations derived from the models in Section II, the severity of the fault (i.e., the diameter of the artificially drilled hole into the material) and the position of the motor bearing (DE or FE). Even the different loads for the same fault were considered as separate classes, which constitutes the most challenging classification problem.

B. Signals to Patterns

The provided descriptor for a particular machine condition is a single high-resolution digital vibration signal with a duration of approximately 10 s. For instance, the file '234.mat', describing an outer race, 0.021 in diameter, 0-hp load fault, contains $n = 122\,426$ samples, acquired at $f_s = 12$ kHz, resulting in a duration of 10.20 s. Since the complete signal represents several periods of a rotational process, it is justifiable to simply chop the total duration into consecutive intervals and consider these as independent patterns acquired at different time instances. Preliminary tests suggested that 15 nonoverlapping intervals is a threshold beyond which performance starts to degrade; hence, this value was fixed in all experiments. Wang in [37] seems to have taken the complete signal to extract the features, obtaining only eight different patterns, Yu in [19] splits

TABLE IV
STATISTICAL FEATURES IN THE TIME AND FREQUENCY DOMAINS
EXTRACTED FROM REPRESENTATIVE MACHINE CONDITIONS. SIGNALS
WERE ACQUIRED AT THE DE OR THE FE. IR = INNER RACE, OR = OUTER
RACE. PERFORATION DIAMETER OF THE HOLES DRILLED INTO THE
BEARING COMPONENTS FOR ALL FAULT TYPES WAS 0.021 in

Feature	Machine Condition						
	Ball DE 021	Ball FE 021	IR DE 021	IR FE 021	Nor- mal	OR DE 021	OR FE 021
DE RMS	0.119	0.113	0.516	0.184	0.073	0.278	0.109
DE SRA	0.08	0.076	0.276	0.116	0.05	0.155	0.071
DE Kurtosis	3.085	3.071	6.561	4.141	2.846	7.188	3.39
DE Skewness	-0.015	-0.007	0.265	-0.03	-0.085	0.1	0.002
DE Peak-to-p.	0.916	0.962	6.071	1.68	0.52	3.233	0.901
DE Crest f.	4.016	4.268	6.526	4.877	3.702	6.344	4.482
DE Impulse f.	5.054	5.366	9.473	6.419	4.614	9.05	5.74
DE Margin f.	5.974	6.336	12.195	7.756	5.439	11.378	6.86
DE Shape f.	1.258	1.257	1.451	1.316	1.246	1.426	1.28
DE Kurtosis	1.5×10^4	1.8×10^4	9.2×10^2	3.5×10^3	9.6×10^4	1.1×10^4	2.3×10^4
DE Freq. cent.	4.363	4.563	17.211	7.071	2.446	12.055	5.019
DE RMS f.freq.	10.799	10.228	46.577	16.583	9.399	25.139	9.777
DE Root variance freq.	9.878	9.153	43.28	14.999	9.074	22.06	8.391

TABLE V
NARROWBAND ENERGY OF THE COMPLEX ENVELOPE SPECTRUM
AT THE DE. ONLY TWO OF SIX HARMONICS ARE SHOWN

Feature	Machine Condition						
	Ball DE 021	Ball FE 021	IR DE 021	IR FE 021	Nor- mal	OR DE 021	OR FE 021
DE 5.4152Harm1	57.495	7.066	210.842	24.758	0.530	4.963	21.110
DE 5.4152Harm2	26.601	6.413	95.183	18.273	0.431	7.736	10.972
DE 3.5848Harm1	13.428	13.168	38.090	39.629	0.546	35.978	19.205
DE 3.5848Harm2	20.563	12.692	27.225	19.043	0.489	12.652	16.939
DE 4.7135Harm1	10.054	7.559	20.102	11.675	0.515	5.949	17.260
DE 4.7135Harm2	7.741	6.913	29.216	15.749	0.243	4.030	9.719
DE 4.9469Harm1	25.616	6.222	46.687	104.864	0.583	3.895	14.417
DE 4.9469Harm2	12.435	6.435	115.286	17.589	0.528	6.474	11.743
DE 3.053Harm1	37.458	14.906	131.837	126.299	0.799	10.041	49.858
DE 3.053Harm2	12.838	7.919	25.093	23.699	0.479	7.281	14.158
DE 3.9874Harm1	18.917	10.348	24.734	111.116	0.604	7.402	18.128
DE 3.9874Harm2	18.315	6.680	26.142	13.945	0.535	3.573	14.185

the signal into 30 parts, and Sreejith *et al.* in [10] divides the signal into 24 parts for the $f_s = 12$ kHz data.

C. Experiment 1: Feature Extraction

Subjecting the CWRU bearing fault data to the feature extraction models defined in Section II, explicit feature vectors for representative machine conditions are provided. Table IV presents the statistical features in the time and frequency domains, defined in Tables I and II of Section II-A for selected condition classes.

The 72 features obtained as the narrowband energy of the complex envelope magnitude described in Section II-B are pointed out in Table V. As an illustrative example, consider the inner-race fault of the DE bearing (6205-2RS JEM SKF). Six harmonics are used, but only the first two are shown. Consider, for instance, the second harmonic. Since the expected frequency is 5.4152 times the running speed of the machine (30 Hz), the focus is on $2 \times 5.4152 \times 30$ Hz, i.e., 324.91 Hz. The 1% narrow band around this frequency is then defined

by the interval [321.66, 328.16 Hz], from which the energy is calculated. In fact, in case of an inner-race fault, the value of 95.183 in the third column is much higher than when the fault is not present, for instance, when a ball fault is present, the value drops to 6.413.

Considering the next feature model, in Fig. 2, the resulting wavelet packet tree is illustrated for two real CWRU examples. In order not to overburden the graph, only the first 0.1 s of the two signals were processed. Each node shows the reconstructed sample, using only the wavelet packet coefficients of that node. Two signals for the same inner-race fault are compared, however with different severities (7 and 21 mils). To elevate the contrast, the 7-mils case is the one acquired without load of the motor, and the 21-mils case is the one at 3-hp load. Qualitatively, it can be observed that the node energies do differ considerably, for instance, the leaves $W_{4,4}$ and $W_{4,12}$ and, consequently, the energy percentages allow a firm distinction between two severities. In Table VI, the feature values for seven representative classes are shown. Each class has two sets of values (acquisition at the DE and the FE). From the table, it can be seen that, e.g., the $W_{4,0}$ energy portion is very elevated for the ‘normal’ class, such that a low value suggests a fault.

D. Experiment 2: Performance Without Feature Selection

Table VII presents the estimated performance of the 1-NN, SVM, and MLP classifiers with data sets extracted by three different feature models and the merged pool of all features, joining the three models. All 130 features were used to train the classifier and run the test, i.e., no feature selection was performed. The 1-nearest neighbor classifier always uses LOO to estimate the performance parameters. For the SVM and MLP classifiers, this method is computationally too expensive; therefore, a tenfold CV was employed. To improve the robustness of the estimation, the mean over ten different of such tenfold runs was taken.

The results of Table VII give a valuable feedback considering the classification task and the feature models. The accuracy (ACC) and AUC-ROC suggest that the CWRU data set is easy to distinguish with respect to the defined classes. The statistical feature model seems to be the least discriminative. The wavelet package energy is by itself the best feature model. The complete pool is not always the best feature model.

An analysis of the results provides a principal motivation of the need for the posterior feature selection. When pooling all feature models, the classification results degrade, since some features contain more noise than information. The subsequent feature selection can improve performance and simultaneously diminish the complexity of the feature model.

E. Experiment 3: Feature Selection

In all feature selection graphs, for the SFS and SFFS algorithms, the horizontal axis represents the number of features selected. For the SBS and SFBS algorithms, the horizontal axis represents the number of features excluded. Fig. 3 shows the estimated accuracy of the 1-NN classifier using LOO CV during the process of feature selection. The algorithm uses the wrapper

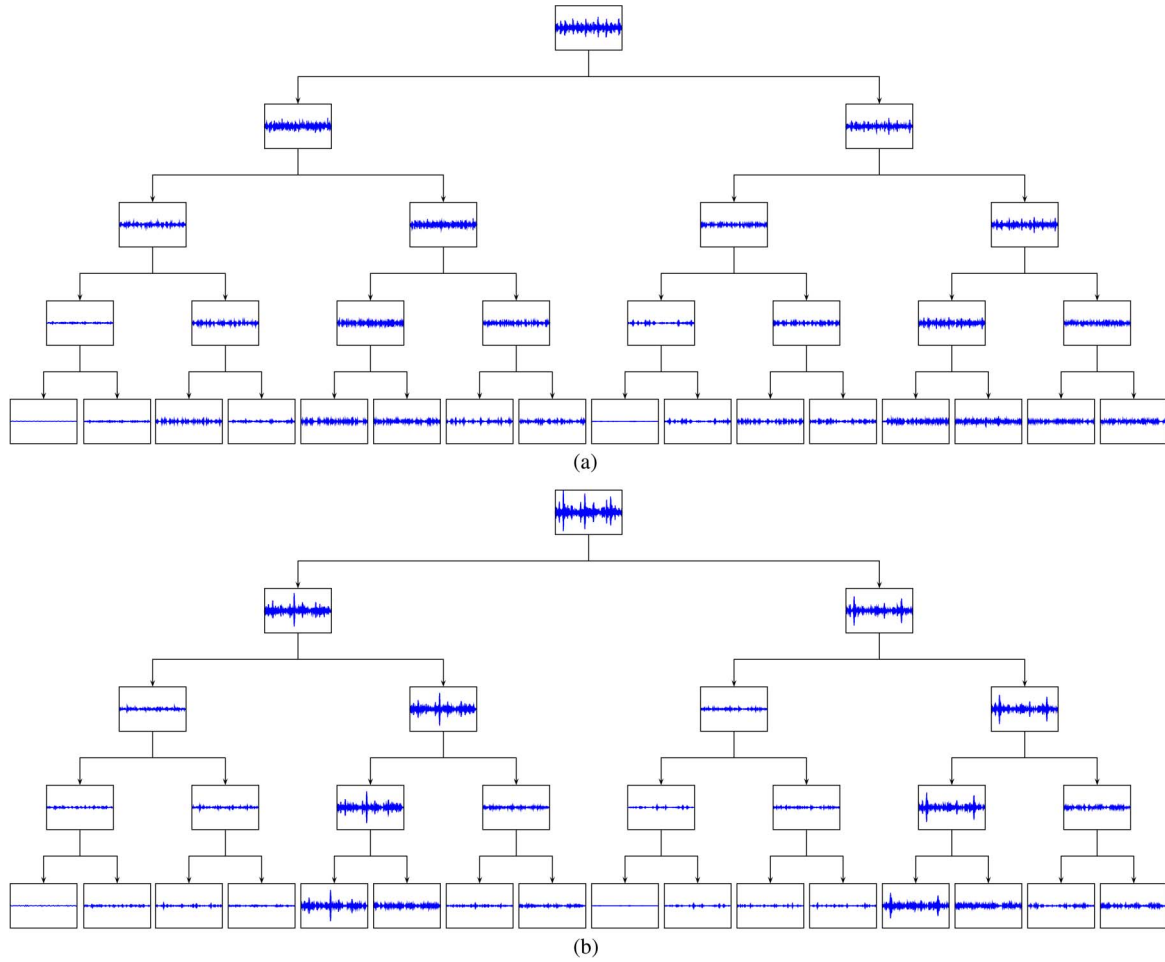


Fig. 2. Reconstructed signal in wavelet packet tree of depth $j = 4$ for two inner-race faults, varying with respect to the fault severity and work load. Only the first 0.1 s of a single sample are processed and shown in order not to overburden the graph. (a) 0.007-in 0-hp inner-race fault signal decomposition. (b) 0.021-in 3-hp inner-race fault signal decomposition.

TABLE VI
WAVELET PACKET TREE FEATURE VECTORS AS THE ENERGY PERCENTAGE OF THE LEAF NODES FOR SEVEN MACHINE CONDITIONS AT THE DE.
LOWER CASE d AND f INDICATE, RESPECTIVELY, THE ACCELEROMETER POSITION (DE, FE)

Machine condition	Feature values (node depth level j , node position s)															
	4, 0	4, 1	4, 2	4, 3	4, 4	4, 5	4, 6	4, 7	4, 8	4, 9	4, 10	4, 11	4, 12	4, 13	4, 14	4, 15
Normal	d 24.89	7.24	11.38	45.39	0.10	1.03	1.63	7.62	0.01	0.01	0.04	0.08	0.04	0.10	0.02	0.35
	f 36.46	4.79	8.09	39.45	0.05	1.21	1.28	7.70	0.01	0.04	0.07	0.10	0.02	0.12	0.10	0.45
Ball DE 007	d 2.70	4.13	1.49	1.76	22.93	10.45	1.12	0.76	0.01	0.03	0.15	0.05	35.33	17.43	0.52	1.06
	f 10.80	8.91	1.62	8.99	1.37	2.92	1.67	2.60	0.04	2.10	4.31	6.85	1.57	4.21	19.32	22.65
Ball DE 021	d 4.69	4.30	1.82	2.49	23.88	8.59	1.74	0.88	0.01	0.14	1.17	0.26	33.20	13.08	2.63	1.05
	f 12.93	6.77	1.22	9.20	1.77	2.32	1.67	2.33	0.05	2.11	7.51	5.31	2.55	3.25	22.67	18.27
IR DE 007	d 0.73	3.67	7.93	2.73	16.45	11.77	4.82	4.06	0.01	0.14	0.49	0.26	16.37	23.29	0.72	6.49
	f 2.80	6.84	21.64	8.35	4.33	3.84	19.46	13.66	0.05	0.70	1.83	1.91	3.46	1.54	4.95	4.56
IR DE 021	d 1.09	4.02	0.37	0.96	44.58	5.75	0.70	0.83	0.01	0.03	0.06	0.05	37.20	3.87	0.08	0.33
	f 4.36	15.47	2.72	8.28	14.63	3.11	4.00	5.23	0.08	1.61	3.91	2.85	14.09	5.13	7.20	7.25
OR DE 007	d 0.89	0.85	0.45	0.58	17.33	12.25	0.49	1.18	0.01	0.06	0.31	0.20	27.30	33.25	0.87	3.93
	f 8.11	3.91	1.15	3.80	1.75	1.33	0.85	1.06	0.26	4.25	16.34	31.00	2.82	2.57	10.39	10.34
OR DE 021	d 5.27	2.16	1.18	0.72	24.70	13.26	0.90	1.20	0.01	0.03	0.08	0.04	17.95	28.84	0.32	3.27
	f 8.14	9.25	1.67	6.41	11.26	7.00	1.42	2.33	0.05	1.52	1.76	3.91	12.42	13.69	7.03	12.09

approach with the performance value as the proper selection criterion, i.e., the estimated accuracy in this case. All three feature models are tested, plus the global pool. Only the wavelet

package graph is shown; the statistical and envelope features exhibit a similar behavior. For the global pool, SFFS and SFBS have identical behavior as SFS and SBS, respectively.

TABLE VII
PERFORMANCE (ACCURACY AND AUC-ROC) OF THREE CLASSIFIERS USING DATASETS WITH DIFFERENT FEATURE EXTRACTION MODELS

Feature model	ACC			AUC-ROC		
	1-NN	SVM	MLP	1-NN	SVM	MLP
Statistical	96.25%	97.25%	93.90%	0.9239	0.9976	0.9827
Wavelet Package	99.83%	99.63%	99.04%	1.0000	0.9999	0.9915
Complex Envelope Spectrum	97.65%	96.88%	93.63%	0.9581	0.9998	0.9999
Complete Pool	99.96%	98.13%	99.97%	0.9917	0.9991	1.0000

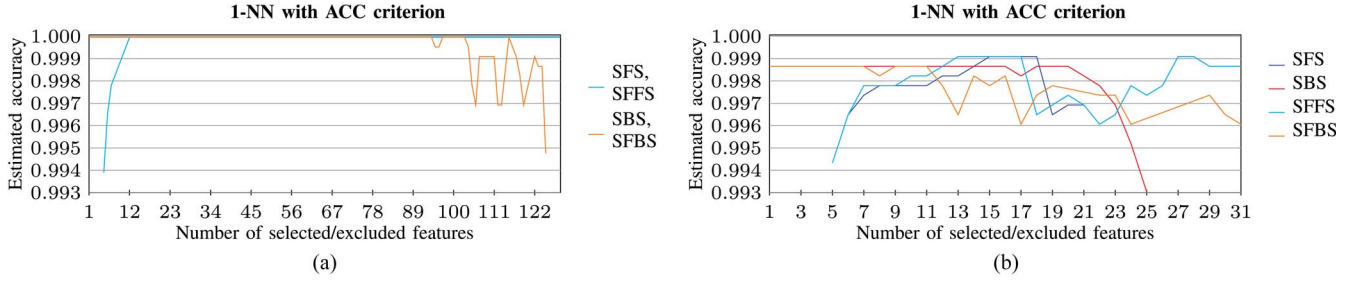


Fig. 3. Estimated accuracy during feature selection for all features and the wavelet package model as a representative of the individual feature models, using the 1-NN classifier. (a) Selection from global feature pool. (b) Selection of wavelet package features.

In general, the floating techniques seem to perform better than the sequential search in the case of the wavelet package model.

The selected subset of all feature models together enters an error-free saturation in a very early stage. The analysis of this result emphasizes the main hypothesis of this work, namely, that selected features from the global feature pool have superior performance compared with the individual feature models.

F. Experiment 4: AUC-ROC for Different Feature Models and Number of Selected Features

As mentioned earlier, the CWRU data set is easy to classify, even with a large number of classes. This implies that the feature selection curve enters a maximum performance score saturation area in an early stage. This is most evidently seen in the first graph in Fig. 3 where the global feature pool is used. To produce more readable ROC curves, the difficulty in separating the classes was artificially raised. A 12-class set was considered, varying severity and work load of the ball fault in the FE, since it showed the worst AUC-ROC in preliminary experiments. Moreover, the signal sampling resolution was reduced, by augmenting the number of patterns obtained from a single signal from 15 to 50, which makes it harder to discern the patterns. Finally, only the first 2 s were used to sample the 50 patterns, instead of the original 10-s signals. A detailed description of the classes is pointed out in Table VIII. The second one is considered as the positive class. This fault of the ball at the FE was identified as the hardest to be distinguished in preliminary tests. The remaining eleven classes are merged into the negative class.

Figs. 4 and 5 show the ROC curves for the global feature pool and the wavelet package model as the best individual feature model, respectively. The AUC-ROC performance criterion increases during feature selection, and additionally, the global feature pool performs better than all the other feature models

TABLE VIII
CLASSES OF THE SPECIAL AUC-ROC DATASET, VARYING FAULT SEVERITY AND LOAD FOR THE FAN END BALL FAULT

Class	Name	Sam- ples	Distri- bution	Description
1	Ball_FE_0_007	50	8.33%	0.007", 0 hp load
2	Ball_FE_1_007	50	8.33%	0.007", 1 hp load
3	Ball_FE_2_007	50	8.33%	0.007", 2 hp load
4	Ball_FE_3_007	50	8.33%	0.007", 3 hp load
5	Ball_FE_0_014	50	8.33%	0.014", 0 hp load
6	Ball_FE_1_014	50	8.33%	0.014", 1 hp load
7	Ball_FE_2_014	50	8.33%	0.014", 2 hp load
8	Ball_FE_3_014	50	8.33%	0.014", 3 hp load
9	Ball_FE_0_021	50	8.33%	0.021", 0 hp load
10	Ball_FE_1_021	50	8.33%	0.021", 1 hp load
11	Ball_FE_2_021	50	8.33%	0.021", 2 hp load
12	Ball_FE_3_021	50	8.33%	0.021", 3 hp load

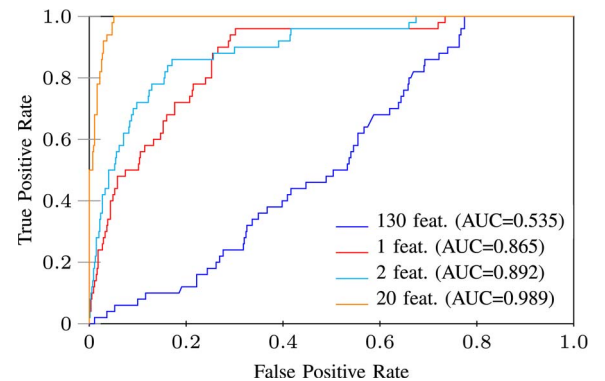


Fig. 4. SVM classifier AUC-ROC for SFS with the global feature pool. Number of all features is 130, best set has 20 features.

individually. The graphs show the result for one feature, two, all, and the number with best performance.

The analysis of this experimental result corroborates the main hypothesis of this work, namely, that using simultaneously more than one feature model, in conjunction with feature selection, increases performance of the fault diagnosis system.

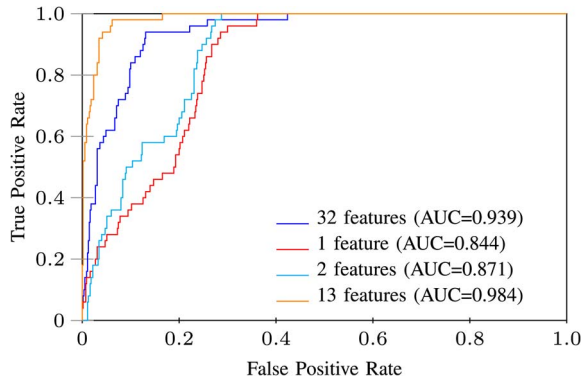


Fig. 5. SVM classifier AUC-ROC for SFS with wavelet package feature model. Number of all features is 32; best set has 13 features.

TABLE IX

BALL BE FAULT CLASSIFICATION PERFORMANCE. CRITERION SHOWN TOGETHER WITH OPTIMAL NUMBER OF FEATURES OR PRINCIPAL COMPONENTS

		Complex Envelope	Pool	Statistical Features	Wavelet Package
Accuracy					
KNN	SFS	36: 0.553	54: 0.758	06: 0.485	21: 0.668
	PCA	65: 0.470	96: 0.652	13: 0.300	16: 0.610
SVM	SFS	32: 0.554	12: 0.792	09: 0.573	13: 0.758
	PCA	16: 0.499	24: 0.721	06: 0.333	15: 0.663
AUC-ROC of second class of table VIII					
KNN	SFS	35: 0.786	54: 0.941	08: 0.718	13: 0.888
	PCA	69: 0.655	06: 0.744	04: 0.640	13: 0.795
SVM	SFS	25: 0.954	24: 0.991	05: 0.877	16: 0.981
	PCA	03: 0.866	38: 0.975	06: 0.836	26: 0.970

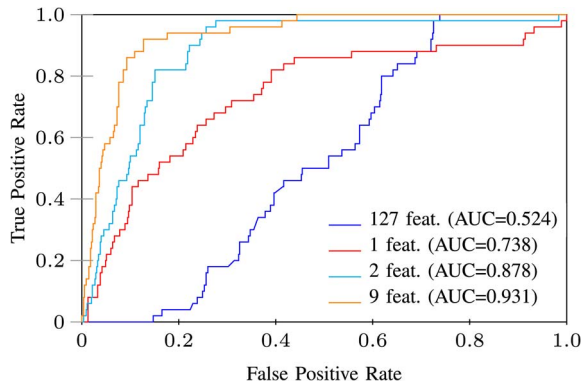


Fig. 6. SVM classifier AUC-ROC for PCA with the global feature pool. Number of all nonzero components is 127; best set has nine principal components. Positive class is the second in Table VIII.

G. Experiment 5: Comparison of Feature Selection and Feature Extraction by PCA

Paying tribute to a consolidated dimensionality reduction technique in fault diagnosis, see, e.g., [28] and [38]; a comparison of feature *selection* to feature *extraction* is done, represented by its most prominent linear method, i.e., PCA. The basic difference is that PCA produces new features as a recombination of the existing ones, whereas feature selection leaves the original features unchanged. The new PC features are ranked by their variance. Since PCA is an unsupervised method, the maximization of variance does not necessarily augment class discernability.

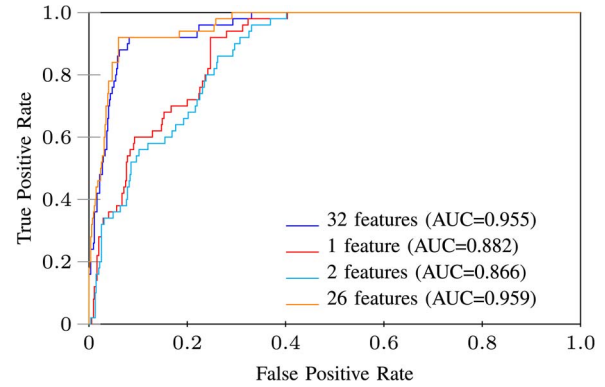


Fig. 7. SVM classifier AUC-ROC for PCA with wavelet package feature model. Number of all components is 32; best set has 26 principal components. Positive class is the second in Table VIII.

Table IX shows the performance criteria for the KNN and SVM classifiers, comparing selection (SFS) and extraction (PCA). For accuracy estimation, all classes are separately considered. For AUC-ROC estimation, all classes in Table VIII are merged, except the second one in the table, which is considered as the positive class, cf., the previous ROC experiment in Section V-G. Selection performs better in all cases. The corresponding ROC curves obtained by PCA with the global feature pool are shown in Fig. 6 and with only the wavelet package model as the best individual feature model in Fig. 7.

VI. CONCLUSION

The three particular feature models used in this work are well known, so are the classifier architectures and performance criteria. There is, however, a novelty approach in the information processing part. The process description is taken to a higher abstraction level by characterizing the process conditions by multiple heterogeneous feature models. This approach by itself would have a low impact on the discriminative power of the diagnosis system. Only if a posterior information filter in the form of feature selection is used, the general performance can be expected to improve. A considerable amount of publications try to compensate a poor feature model by an oversophisticated classifier model. In this paper, a different approach is taken. If a good process description is available, a simple classifier is sufficient. In this case, the description appears in the form of a selection of many available features, stemming from quite distinct sources. Of course, this work cannot formally prove that the idea of this paper has general value. The experiments, however, suggest that this approach is a promising methodology, extensible to other areas of industrial applications.

REFERENCES

- [1] S. Nandi, S. Choi, and H. Meshgin-Kelk, *Electric Machines Modeling: Condition Monitoring and Fault Diagnosis*. Boca Raton, FL, USA: CRC Press, 2013.
- [2] V. Devedzic, "Knowledge modeling—State of the art," *Integr. Comput.-Aid. Eng.*, vol. 8, no. 3, pp. 257–281, Aug. 2001.
- [3] Case Western Reserve University, Bearing Data Center, 2014, accessed: 2014-05-15. [Online]. Available: <http://csegroups.case.edu/bearingdatacenter>

- [4] A. K. Mahamad and T. Hiyama, "Fault classification based artificial intelligent methods of induction motor bearing," *Int. J. Innov. Comput., Inf. Control*, vol. 7, no. 9, pp. 5477–5494, Sep. 2011.
- [5] S. Sassi and B. B. M. Thomas, "A numerical model to predict damaged bearing vibrations," *J. Vib., Control*, vol. 13, no. 11, pp. 1603–1628, Nov. 2007.
- [6] J. Chebil, M. Hrairi, and N. Abushikhah, "Signal analysis of vibration measurements for condition monitoring of bearings," *Australian J. Basic Appl. Sci.*, vol. 5, no. 1, pp. 70–78, 2011.
- [7] S.-D. Wu, C.-W. Wu, T.-Y. Wu, and C.-C. Wang, "Multi-scale analysis based ball bearing defect diagnostics using Mahalanobis distance and support vector machine," *Entropy*, vol. 15, no. 2, pp. 416–433, Jan. 2013.
- [8] J.-F. Zhang and Z.-C. Huang, "Kernel Fisher discriminant analysis for bearing fault diagnosis," in *Proc. IEEE Mach. Learn. Cybern.*, Aug. 2005, vol. 5, pp. 3216–3220.
- [9] W. Sun, G. Yang, Q. Chen, A. Palazoglu, and K. Feng, "Fault diagnosis of rolling bearing based on wavelet transform and envelope spectrum correlation," *J. Vib., Control*, vol. 19, no. 6, pp. 924–941, Apr. 2013.
- [10] B. Sreejith, A. Verma, and A. Srividya, "Fault diagnosis of rolling element bearing using time-domain features and neural networks," in *Proc. 3rd IEEE ICIS Region 10*, Dec. 2008, pp. 1–6.
- [11] Z. Xia, S. Xia, L. Wan, and S. Cai, "Spectral regression based fault feature extraction for bearing accelerometer sensor signals," *Sensors*, vol. 12, no. 10, pp. 13 694–13 719, Oct. 2012.
- [12] J. Liu, "Shannon wavelet spectrum analysis on truncated vibration signals for machine incipient fault detection," *J. Meas. Sci. Technol.*, vol. 23, no. 5, p. 055604, May 2012.
- [13] J. Luo, D. Yu, and M. Liang, "A kurtosis-guided adaptive demodulation technique for bearing fault detection based on tunable-Q wavelet transform," *J. Meas. Sci. Technol.*, vol. 24, no. 5, pp. 1–11, May 2013.
- [14] Z. Wei, J. Gao, X. Zhong, Z. Jiang, and B. Ma, "Incipient fault diagnosis of rolling element bearing based on wavelet packet transform and energy operator," *WSEAS Trans. Syst.*, vol. 10, no. 3, pp. 81–90, Mar. 2011.
- [15] K. F. Al-Raheem and W. Abdul-Karem, "Rolling bearing fault diagnostics using artificial neural networks based on Laplace wavelet analysis," *Int. J. Eng., Sci. Technol.*, vol. 2, no. 6, pp. 278–290, 2010.
- [16] J. Chebil, G. Noel, M. Mesbah, and M. Deriche, "Wavelet decomposition for the detection and diagnosis of faults in rolling element bearings," *Jordan J. Mech. Ind. Eng.*, vol. 3, no. 4, pp. 260–267, Dec. 2009.
- [17] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *J. Mach. Learn. Res.*, vol. 3, pp. 1157–1182, Mar. 2003.
- [18] S.-D. Wu, P.-H. Wu, C.-W. Wu, J.-J. Ding, and C.-C. Wang, "Bearing fault diagnosis based on multiscale permutation entropy and support vector machine," *Entropy*, vol. 14, no. 8, pp. 1343–1356, Jul. 2012.
- [19] J.-B. Yu, "Bearing performance degradation assessment using locality preserving projections," *Expert Syst. Appl.*, vol. 38, no. 6, pp. 7440–7450, Jun. 2011.
- [20] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognit. Lett.*, vol. 27, no. 8, pp. 861–874, Jun. 2006.
- [21] M. Delgado Prieto, G. Cirrincione, A. Garcia Espinosa, J. Ortega, and H. Henao, "Bearing faults detection by a novel condition monitoring scheme based on statistical-time features and neural networks," *IEEE Trans. Ind. Electron.*, vol. 60, no. 8, pp. 3398–3407, Aug. 2013.
- [22] Y. Gritli *et al.*, "Advanced diagnosis of electrical faults in wound rotor induction machines," *IEEE Trans. Ind. Electron.*, vol. 60, no. 9, pp. 4012–4024, Sep. 2013.
- [23] T. Boukra, A. Lebaroud, and G. Clerc, "Statistical and neural-network approaches for the classification of induction machine faults using the ambiguity plane representation," *IEEE Trans. Ind. Electron.*, vol. 60, no. 9, pp. 4034–4042, Sep. 2013.
- [24] D. He, R. Li, and J. Zhu, "Plastic bearing fault diagnosis based on a two-step data mining approach," *IEEE Trans. Ind. Electron.*, vol. 60, no. 8, pp. 3429–3440, Aug. 2013.
- [25] R. K. Mobley, *Root Cause Failure Analysis (Plant Engineering Maintenance Series)*. Woburn, MA, USA: Butterworth-Heinemann, 1999.
- [26] R. X. Gao and R. Yan, *Wavelets. Theory and Applications for Manufacturing*. New York, NY, USA: Springer-Verlag, 2011.
- [27] S. Ding, S. Yin, K. Peng, H. Hao, and B. Shen, "A novel scheme for key performance indicator prediction and diagnosis with application to an industrial hot strip mill," *IEEE Trans. Ind. Informat.*, vol. 9, no. 4, pp. 2239–2247, Nov. 2013.
- [28] S. Yin, S. X. Ding, A. Haghani, H. Hao, and P. Zhang, "A comparison study of basic data-driven fault diagnosis and process monitoring methods on the benchmark Tennessee Eastman process," *J. Process Control*, vol. 22, no. 9, pp. 1567–1581, Oct. 2012.
- [29] J. J. Downs and E. F. Vogel, "A plant-wide industrial process control problem," *Comput. Chem. Eng.*, vol. 17, no. 3, pp. 245–255, Mar. 1993.
- [30] S. Yin, H. Luo, and S. Ding, "Real-time implementation of fault-tolerant control systems with performance optimization," *IEEE Trans. Ind. Electron.*, vol. 61, no. 5, pp. 2402–2411, May 2014.
- [31] M. Kudo and J. Sklansky, "Comparison of algorithms that select features for pattern classifiers," *Pattern Recognit.*, vol. 33, no. 1, pp. 25–41, Jan. 2000.
- [32] 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.
- [33] N. Japkowicz and M. Shah, *Evaluating Learning Algorithms: A Classification Perspective*. New York, NY, USA: Cambridge Univ. Press, 2011.
- [34] T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE Inf. Theory*, vol. 13, no. 1, pp. 21–27, Jan. 1967.
- [35] M. F. Möller, "A scaled conjugate gradient algorithm for fast supervised learning," *Neural Netw.*, vol. 6, no. 4, pp. 525–533, 1993.
- [36] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, Sep. 1995.
- [37] Y. Wang *et al.*, "Classification of fault location and the degree of performance degradation of a rolling bearing based on an improved hypersphere-structured multi-class support vector machine," *Mech. Syst. Signal Process.*, vol. 29, pp. 404–414, May 2012.
- [38] C.-M. Vong, P.-K. Wong, and W.-F. Ip, "A new framework of simultaneous-fault diagnosis using pairwise probabilistic multi-label classification for time-dependent patterns," *IEEE Trans. Ind. Electron.*, vol. 60, no. 8, pp. 3372–3385, Aug. 2013.



Thomas W. Rauber (M'09) received the Diploma degree from the Fakultät für Informatik, Universität Karlsruhe (TH), Karlsruhe, Germany, in 1989 and the Ph.D. degree in electrical engineering from the Department de Engenharia Electrotécnica, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Lisbon, Portugal, in 1995.

He conducted postdoctoral research at AG ROSY/FB Informatik/TU Kaiserslautern, Germany in 2005/2006 and at Ludwig-Maximilian Universität (LMU), Munich, Germany, in 2012. Since 1996, he has been a Professor of computer science with the Department de Informática, Universidade Federal do Espírito Santo, Vitória, Brazil. His current research interests include pattern recognition, artificial neural networks, and fault diagnosis in industrial processes.



Francisco de Assis Boldt received the P.Tech. degree from Centro Universitário do Espírito Santo, Colatina, Brazil, in 1998 and the M.Sc. degree in computer science from the Universidade Federal do Espírito Santo, Vitória, Brazil, in 2008, where he is currently working toward the Ph.D. degree in computer science.

His research interests include computational intelligence applied to fault diagnosis.



Flávio Miguel Varejão received the B.E. degree in civil engineering from the Universidade Federal do Espírito Santo, Vitória, Brazil, in 1987 and the M.Sc. and Ph.D. degrees in computer science from the Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, Brazil, in 1991 and 1999, respectively.

Since 1990, he has been a Professor of computer science with the Department de Informática, Universidade Federal do Espírito Santo. His research interests include pattern recognition and data-driven fault diagnosis.