Proceedings of the 2009 IEEE International Conference on Systems, Man, and Cybernetics
San Antonio, TX, USA - October 2009

Rolling element bearing fault classification using soft computing techniques

Marco Cococcioni, Beatrice Lazzerini, Sara Lioba Volpi
Department of Information Engineering, Via Diotisalvi 2
University of Pisa
Pisa, Italy
{m.cococcioni, b.lazzerini, sara.volpi}@iet.unipi.it

***Abstract*—This paper presents a method, based on classification techniques, for automatically detecting and diagnosing various types of defects which may occur on a rolling element bearing. In the experiments we have used vibration signals coming from a mechanical device including more than ten rolling element bearings monitored by means of four accelerometers: the signals have been collected both with all faultless bearings and substituting one faultless bearing with an artificially damaged one: four different defects have been taken into account. The proposed technique considers all the aspects of classification: feature selection, different base classifiers (two statistical classifiers, namely LDC and QDC, and MLP neural networks) and classifier fusion. Experiments, performed on the vibration signals represented in the frequency domain, have shown that the proposed classification method is highly sensitive to different types of defects and to different severity degrees of the defects.**

***Keywords*—automatic fault diagnosis, fault classification, multi-layer perceptron, statistical classifiers, classifier fusion**

I. INTRODUCTION

Breakdowns in industrial manufacturing systems can have serious consequences on people, machinery and environment [1]-[4]. For this reason machine condition monitoring and fault diagnostics have became an integral part of industrial systems with the aim of reducing costly machine downtime and ensuring production quality. There are mainly three types of maintenance: corrective, preventive and condition-based maintenance (CBM) [1].

Corrective maintenance consists of repairing faults after they have occurred, while preventive maintenance aims to prevent future faults by performing periodic inspections to identify conditions that would cause breakdowns, and correcting these conditions. Both these kinds of maintenance have serious drawbacks: the former does not prevent any faults, the latter can waste time and money since many controls may result useless, besides no guarantee can be given regarding the proper work of the system between two subsequent checks. To overcome these drawbacks CBM has been introduced.

CBM consists of monitoring the system state and performing appropriate maintenance actions only when necessary. One field in which CBM has been widely applied is rotating machinery maintenance [3,4]. Actually, rotating machines are present in most manufacturing and production industries, and real-time monitoring and diagnostics are needed to guarantee a continuous and reliable production process. Most rotating machines operate by means of bearings which may develop several types of faults. These faults may cause machine breakdown and decrease the performance level. Different methods for detection and diagnosis of faults in bearings have been developed.

Traditional techniques for bearing performance analysis include time-domain [5,6] and frequency-domain analysis [4] used separately or together [7,8]. Time-domain analysis is based on performance indexes such as RMS (Root Mean Square), Crest Factor and Kurtosis, while frequency-domain analysis is based on the Fourier Transform technique. Frequency-domain analysis is the most popular one, perhaps because characteristics of vibration signals are more easily noticed in the frequency domain rather than in the time domain [3,4,8].

This paper aims to achieve the following objectives: given a mechanical object containing rolling bearings, i) to detect the presence of a defect, ii) to recognize the specific kind of defect, iii) to recognize the severity of the defect. To this aim, we have dealt with the problem as a classification problem, adopting two statistical classifiers, namely the Linear Discriminant Classifier (LDC) and the Quadratic Discriminant Classifier (QDC), and Multi-Layer Perceptron (MLP) neural networks. LDC is a minimum-error (Bayes) classifier for normally distributed classes with equal covariance matrices [9-12],. QDC is a minimum-error (Bayes) classifier for normally distributed classes with class-specific covariance matrices [9,10,11]. MLPs [9] are characterized by one or more hidden layers and non-linear transfer functions. In the experiments we used one hidden layer and logarithmic sigmoid transfer functions.

In particular, we use LDC and QDC to perform both feature selection and classification, whereas MLP neural networks perform classification of signals represented by means of the features selected by LDC and QDC.

Finally, to solve particularly difficult classification problems, we have adopted *classifier fusion* [9]. Indeed, it is well-known that the appropriate combination of a set of different classifiers designed for a given classification task may achieve higher performance than any of the classifiers considered individually [9,13].

978-1-4244-2794-9/09/$25.00 ©2009 IEEE

In the experiments we used vibration signals coming from bearings monitored in the time domain by four accelerometers, the second and the third accelerometers closer than the others to the bearing to be analyzed. For each experiment we used the PRTools software in a Matlab environment [14].

II. DEFECTS

In this section we briefly describe the types of defects analyzed.

In the experiments we use vibration signals coming from a mechanical device including more than ten rolling element bearings monitored by means of four accelerometers: the signals have been collected both with all faultless bearings and after substituting one faultless bearing with a damaged one.

The bearings were artificially damaged. Four types of damages have been considered. Experimental data were collected before and after each damage. Therefore the data can be classified into five classes:

*C*1: faultless bearing,

*C*2: bearing with an indentation on the inner raceway,

*C*3: bearing with an indentation on the roll,

*C*4: bearing with sandblasting of the inner raceway,

*C*5: bearing with unbalanced cage.

The fault of bearings of class *C*2 consists of a 450 μm indentation on the inner raceway, whereas bearings of class *C*3 can be divided into three subclasses depending on the severity of the damage (light, medium or high):

*C*3.1: bearing with a 450 μm indentation on the roll (light),

*C*3.2: bearing with a 1.1 mm indentation on the roll (medium),

*C*3.3: bearing with a 1.29 mm indentation on the roll (high).

In the following we will refer to the complete set of all damaged bearings (classes *C*2, *C*3, *C*4, *C*5) as class *C*6.

III. EXPERIMENTAL DATA

The data used in the experiments are the vibration signals recorded by four accelerometers. The data were recorded for time intervals of ten minutes. We considered a data set consisting of one-second signals and including 2890 signals for class *C*1, 1770 for class *C*2, 4790 for class *C*3, 1520 for class *C*4 and 1770 for class *C*5 (Table I).

TABLE I. SIGNALS PER CLASS

Class	*C*1	*C*2	*C*3	*C*4	*C*5
Number of signals (sec)	2890	1770	4790	1520	1770

Class *C*3 is subdivided in the following way: 1770 signals (sec) for class *C*3.1, 1250 for class *C*3.2 and 1770 for class *C*3.3 (Table II). Fig. 1 shows an example of the time signals and the corresponding FFT for classes *C*1 and *C*2.

TABLE II. AMOUNT OF DATA FOR CLASS C3

Class	*C*3.1	*C*3.2	*C*3.3
Number of signals (sec)	1770	1250	1770

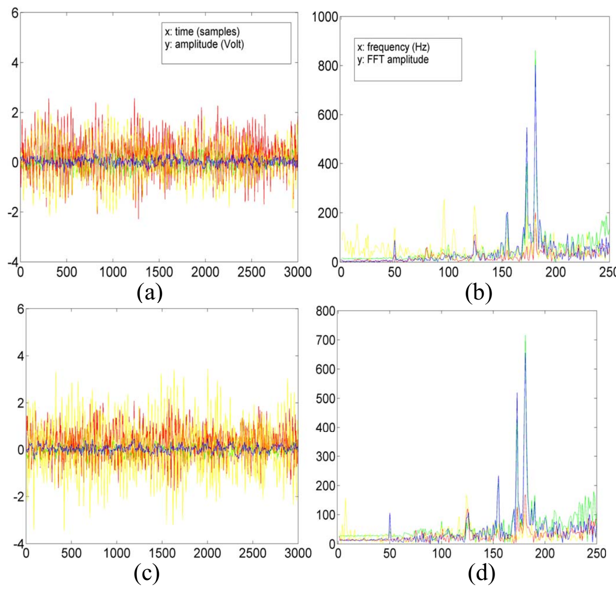

Figure 1. Examples of time signals and corresponding FFT. (a) time signal for *C*1, (b) FFT for *C*1, (c) time signal for C2, (d) FFT for C2. Accelerometer 1 (green), accelerometer 2 (red), accelerometer 3 (yellow), accelerometer 4 (blue)

We worked in the frequency domain by transforming the signals by the Fast Fourier Transform (FFT). Unlike the classical approach, which identifies specific characteristic frequencies associated with given defects, we tried to find out the frequencies able to discriminate among the different defects taken into consideration.

Thus, we considered the frequency interval [1, 250] Hz (the interval does not contain the continuous component), sampled every 1 Hz. Therefore, each signal is represented by 250 frequency samples. As there are four accelerometers the total number of frequency samples for each element to be analyzed is $250\times4=1000$. In other words, each signal is represented in $\Re^{1000}$. The 1000 frequency samples (referred to as *features* in the following) are obtained by concatenating the four groups of 250 frequency samples (i.e., features) relative to the four accelerometers (Fig. 2).

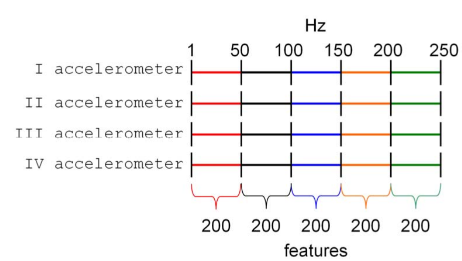

Figure 2. Organization of the features considering the four accelerometers and the five frequency ranges

As a final remark, we point out that for each experiment the data have been balanced using a random technique so that each class involved in the experiment contains the same number of samples as the least numerous one. Then the training set has been built by randomly choosing 70% of the total data, while the remaining data have been used as the test set.

A. *First series of experiments: classification of C1 and C6*

The goal of these experiments is to classify the signals into two classes: faultless bearings ($C1$) and damaged bearings ($C2$, $C3$, $C4$, $C5$). As said before, the set of all the damaged bearings will be referred to as class $C6$.

Since each signal is represented in $\Re^{1000}$, we need to decrease the space dimension. To this aim, we first divide the frequency interval [1, 250] Hz into five sub-intervals consisting, respectively, of the first 50 frequencies, the second 50 frequencies, etc. In each sub-interval, each signal is represented by 200 features obtained by concatenating the four groups of 50 features associated with the four accelerometers (Fig. 2).

For each sub-interval, we look for the best *discriminating frequencies* (DFs), i.e., the frequencies that are able to provide the best accuracy when used to represent the signals to be classified. In this way, besides decreasing the space dimension, we also identify the most significant frequency (sub-)interval for classification purposes. This step is performed using the forward feature selection (FFS), based on the *featself* function of PRTools. We chose to use FFS because it is a reasonable compromise between exhaustive search and random search. We adopted LDC and QDC to perform both feature selection and classification of the signals represented through the selected features. This choice stems from the fact that LDC and QDC are fast trainable classifiers with only one parameter r, called *regularization parameter* (one degree of freedom). However we will consider r fixed to 0 (the default value in PRTools) for both the LDC and QDC classifiers, for all the following experiments.

We use 5 LDCs and 5 QDCs: each LDC/QDC works on a particular range of frequency, namely, the range 1-50 Hz, the range 51-100 Hz, etc. We experimentally verified that each classifier achieves the maximum classification accuracy with less than 200 features. The typical situation is represented in Fig. 3: we can notice that the accuracy increases with the number of features up to a point in which the accuracy remains almost constant and eventually decreases reaching a value that is equal to $1/n$, with n being the number of classes (we recall that we work with balanced classes).

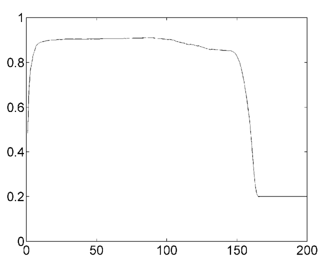

Figure 3. Typical curve representing the classification accuracy (y-axis) versus the number of features (x-axis) for a five-class problem

Considering all the 200 features and repeating the experiment 10 times, LDC and QDC have both selected as the best frequency range for this classification problem the fourth range, i.e., the range 151-200 Hz as in this range we obtained the highest accuracies. In particular, we found that in the fourth frequency range LDC and QDC achieved a maximum accuracy of 99.66 % with 33 features and 99.94 % with 20 features, respectively.

Observing the curve that represents the classification accuracy versus the number of selected features, we noticed that just with the first 6 and 10 DFs, respectively, the two classifiers LDC and QDC achieve a performance close to the maximum in all the frequency ranges. We therefore decided to adopt only 6 and 10 DFs, respectively, to keep computation complexity at an acceptable level. Indeed we can notice that each new added feature brought a negligible improvement after 6 and 10 features, respectively. In this way, the space dimension is reduced from $\Re^{1000}$ to $\Re^{6}$ and $\Re^{10}$, respectively. In the following, we will refer to the discriminating features chosen to reduce the space dimension as *reduced discriminating features* (RDFs). The accuracy obtained by LDC and QDC considering only the RDFs for each frequency range is shown in Table III. Table IV shows the list of the RDFs for the fourth range.

TABLE III. CLASSIFICATION OF $C1$, $C6$. ACCURACY FOR LDC AND QDC IN THE FIVE FREQUENCY RANGES (6 AND 10 FEATURES RESPECTIVELY)

Range	Frequency range	Accuracy of LCD (mean over 10 trails)	Accuracy of QDC (mean over 10 trials)
1	1-50 Hz	86.82 %	85.47 %
2	51-100 Hz	86.82 %	85.47 %
3	101-150 Hz	94.35 %	93.94 %
4	151-200 Hz	**98.45 %**	**99.76 %**
5	201-250 Hz	84.98 %	89.45 %

TABLE IV. CLASSIFICATION OF C1, C6. LIST OF THE RDFS USING THE LDC AND QDC CLASSIFIERS FOR THE FOURTH FREQUENCY RANGE

LDC	181, 23, 131, 138, 39, 32
QDC	181, 23, 131, 31, 118, 44, 5, 187, 190, 8

Fig. 4 shows the signals (both faultless and damaged) around feature 181 of the fourth accelerometer, i.e., the first DF selected by FFS in the fourth frequency range. Fig. 4 shows the good separation of the two classes performed by this DF.

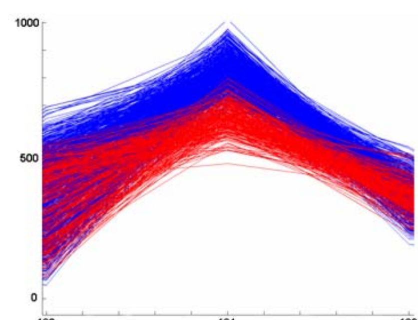

Figure 4. Faultless (blue) and damaged signals (red) around feature 181

B. Second series of experiments: classification of C1, C3.1, C3.2, C3.3

The goal of this series of experiments is to classify the signals into four classes *C*1, *C*3.1, *C*3.2, *C*3.3. These experiments aim to distinguish between faultless and damaged bearings, and to recognize the different levels of severity of the same type of damage.

Repeating the experiment 10 times, once again, both LDC and QDC classifiers, using the FFS algorithm, have selected as the best range for this classification problem the fourth range, i.e., the range [151, 200] Hz as in this range we succeeded in obtaining the highest mean accuracies. In particular, considering all the 200 features, LDC and QDC achieved the maximum accuracy of 99.76 % with 101 features and 99.93 % with 89 features, respectively.

In this case, we considered 10 features as RDFs for both LDC and QDC. Actually, we wish to remark that, in order to adopt a uniform approach in all experiments, and taking into account the plots of accuracy versus number of DFs, we verified that choosing 10 RDFs is a good compromise both in this case and in the following cases. Indeed increasing the number of features brings to negligible improvements of the accuracy.

Table V shows, for each frequency range, the results obtained with the first 10 features (i.e., the RDFs); the accuracy is very close to the maximum one. Table VI shows the list of the 10 RDFs for the fourth frequency range.

TABLE V. CLASSIFICATION OF *C*1, *C*3.1, *C*3.2, *C*3. ACCURACY FOR LDC AND QDC IN THE FIVE FREQUENCY RANGES (10 FEATURES)

Range	Frequency range	Accuracy of LCD (mean over 10 trails)	Accuracy of QDC (mean over 10 trials)
1	1-50 Hz	94.10 %	97.43 %
2	51-100 Hz	90.70 %	92.40 %
3	101-150 Hz	93.07 %	95.50 %
4	151-200 Hz	**99.73 %**	**99.87 %**
5	201-250 Hz	89.70 %	90.96 %

TABLE VI. CLASSIFICATION OF *C*1, *C*3.1, *C*3.2, *C*3. LIST OF THE 10 RDFS USING LDC AND QDC FOR THE FOURTH FREQUENCY RANGE

LDC	131, 182, 181, 31, 23, 73, 130, 32, 123, 138
QDC	131, 181, 123, 182, 73, 31, 120, 138, 48, 132

C. Third series of experiments: classification of C1, C2, C3, C4, C5

The goal of this series of experiments is to classify the signals into five classes *C1*, *C2*, *C3*, *C4*, *C5*. These experiments aim to recognize the different types of damage regardless of their severity.

Repeating the experiment 10 times, once again, the LDC and QDC classifiers have both selected as the best frequency range for this classification problem the fourth range. Considering all the 200 features, we found that, in the fourth frequency range, LDC achieved a maximum accuracy of 94.30 % using 86 features, while QDC obtained a maximum accuracy of 95.00 % using 102 features. The accuracy obtained by LDC and QDC considering only the RDFs (10 also in this case) for each frequency range is shown in Table VII.

TABLE VII. CLASSIFICATION OF *C*1, *C*2, *C*3, *C*4, *C*5. ACCURACY FOR LDC AND QDC IN THE FIVE FREQUENCY RANGES (10 FEATURES)

Range	Frequency range	Accuracy of LCD (mean over 10 trails)	Accuracy of QDC (mean over 10 trials)
1	1-50 Hz	61.72 %	64.74 %
2	51-100 Hz	63.79 %	63.90 %
3	101-150 Hz	56.21 %	57.41 %
4	151-200 Hz	**91.01 %**	**92.41 %**
5	201-250 Hz	58.88 %	761.73 %

The list of the 10 RDFs and an example of the related confusion matrices are shown, respectively, in Tables VIII-X. From Tables IX and X we notice that the main part of the error (48.25 % for the QDC classifier, which achieves the best classification accuracy) is due to the misclassification of class *C*3, which is often recognized as *C*2 and vice versa. The following experiment will allow us to understand where this error is exactly placed, in other words we will expand the class C3 in its subclasses and then we will search the subclass(es) which account for most error (we wonder if *C*2 is misclassified with all the elements of class *C*3 or perhaps only, or mainly, with the elements of a subclass, i.e., *C*3.1, *C*3.2, *C*3.3).

TABLE VIII. CLASSIFICATION OF *C*1, *C*2, *C*3, *C*4, *C*5. LIST OF THE 10 RDFS USING LDC AND QDC FOR THE FOURTH FREQUENCY RANGE

LDC	181, 182, 123, 137, 77, 131, 81, 82, 131, 105
QDC	181, 182, 123, 137, 77, 81, 131, 82, 132, 26

TABLE IX. CLASSIFICATION OF *C*1, *C*2, *C*3, *C*4, *C*5. LDC CONFUSION MATRIX FOR THE TEST SET (10 FEATURES)

		Estimated labels					Total
		C1	***C2***	***C3***	***C4***	***C5***	
True labels	***C1***	439	8	3	0	6	456
	C2	3	375	**60**	18	0	456
	C3	1	**44**	385	26	0	456
	C4	0	34	27	395	0	456
	C5	3	2	1	0	450	456
Total		446	463	476	439	456	2280

TABLE X. CLASSIFICATION OF *C*1, *C*2, *C*3, *C*4, *C*5. QDC CONFUSION MATRIX FOR THE TEST SET (10 FEATURES)

		Estimated labels					Total
		C1	***C2***	***C3***	***C4***	***C5***	
True labels	***C1***	446	1	2	0	7	456
	C2	0	408	**33**	15	0	456
	C3	1	**64**	347	40	0	456
	C4	2	24	10	420	0	456
	C5	1	0	1	0	454	456
Total		450	497	393	475	461	2280

D. Fourth series of experiments: classification of C1, C2, C3.1, C3.2, C3.3, C4, C5

The goal of these experiments is to classify the signals into seven classes *C*1, *C*2, *C*3.1, *C*3.2, *C*3.3, *C*4, *C*5. These experiments aim to recognize not only the different types of fault but also the different degrees of severity.

LDC and QDC, using the FFS algorithm and repeating the experiment 10 trials, achieved the maximum performance on the fourth frequency range. When using all the 200 features, LDC and QDC achieved the maximum accuracy of 95.30 % with 110 features and 97.88 % with 73 features, respectively. With 10 RDFs, the LDC and QDC classifiers achieved the accuracy shown in Table XI. We chose 10 RDFs to keep the complexity at an acceptable level. The 10 RDFs and the related confusion matrices are shown, respectively, in Tables XII-XIV.

TABLE XI. CLASSIFICATION OF *C*1, *C*2, *C*3.1, *C*3.2, *C*3.3, *C*4, *C*5. ACCURACY FOR LDC AND QDC IN THE FIVE FREQUENCY RANGES

Range	Frequency range	Accuracy of LCD (mean over 10 trails)	Accuracy of QDC (mean over 10 trials)
1	1-50 Hz	69.84 %	74.08 %
2	51-100 Hz	65.64 %	69.45 %
3	101-150 Hz	65.68 %	68.30 %
4	151-200 Hz	**91.10 %**	**94.38 %**
5	201-250 Hz	60.93 %	63.81 %

TABLE XII. CLASSIFICATION OF *C*1, *C*2, *C*3.1, *C*3.2, *C*3.3, *C*4, *C*5. LIST OF THE 10 RDFS USING LDC AND QDC FOR THE FOURTH FREQUENCY RANGE

LDC	131, 123,181, 77, 182, 137, 81, 136, 73, 82
QDC	131, 123, 181, 77, 182, 81, 82, 31, 32, 173

TABLE XIII. CLASSIFICATION OF *C*1, *C*2, *C*3.1, *C*3.2, *C*3.3, *C*4, *C*5. LDC CONFUSION MATRIX FOR THE TEST SET (10 FEATURES)

		Estimated labels							Total
		C1	*C2*	*C3.1*	*C3.2*	*C3.3*	*C4*	*C5*	
True labels	*C1*	354	13	2	1	0	0	5	375
	C2	2	297	**56**	0	0	20	0	375
	C3.1	0	**60**	287	4	0	24	0	375
	C3.2	1	0	2	367	0	0	5	375
	C3.3	0	0	0	0	375	0	0	375
	C4	0	20	13	0	0	342	0	375
	C5	4	1	2	13	0	0	355	375
Total		361	391	362	385	375	386	365	2625

From Tables XIII and XIV, we can observe that the main part of the error (48.07 % for the QDC classifier) is due to the misclassification of class *C*3.1, which is sometimes recognized as *C*2 (39.10 %), and vice-versa (8.97 %). This means that the classification system cannot distinguish correctly between indentation on the inner raceway and light indentation on the roll.

On the other hand, the fourth series of experiments resulted in a better accuracy than the third series of experiments. This suggests that one could achieve the objective of the third series of experiments by appropriately exploiting the fourth series of experiments. More precisely, we can classify the data into seven classes and then put together *C*3.1, *C*3.2 and *C*3.3 to obtain *C*3, thus returning to the five-class problem. In this way, considering the results obtained by the QDC classifier, the confusion matrix for the five-class problem becomes the one in Table XV and the accuracy becomes 94.06 %. This accuracy is higher than the one obtained during the third series of experiments (92.41 %). We remark that we still use the same number of RDFs.

TABLE XIV. CLASSIFICATION OF *C*1, *C*2, *C*3.1, *C*3.2, *C*3.3, *C*4, *C*5. QDC CONFUSION MATRIX FOR THE TEST SET (10 FEATURES)

		Estimated labels							Total
		C1	*C2*	*C3.1*	*C3.2*	*C3.3*	*C4*	*C5*	
True labels	*C1*	369	2	1	1	0	1	1	375
	C2	1	341	**14**	0	0	18	1	375
	C3.1	0	**61**	298	0	0	16	0	375
	C3.2	0	0	0	372	0	1	2	375
	C3.3	0	0	0	0	375	0	0	375
	C4	1	8	19	0	1	346	0	375
	C5	4	1	0	2	0	0	368	375
Total		375	413	332	375	376	382	372	2625

TABLE XV. CLASSIFICATION OF *C*1, *C*2, *C*3, *C*4, *C*5. QDC CONFUSION MATRIX FOR THE TEST SET (10 FEATURS).

		Estimated labels					Total
		C1	*C2*	*C3*	*C4*	*C5*	
True labels	*C1*	369	2	2	1	1	375
	C2	1	341	14	18	1	375
	C3	0	61	1045	17	2	1125
	C4	1	8	20	346	0	375
	C5	4	1	2	0	368	375
Total		375	413	1082	382	383	2625

E. Fifth series of experiments: classification of C2, C3.1

Once identified where the main part of the error is (misclassification of *C*2 with *C*3,1 and vice versa rather than *C*2 with *C*3) we tried to cope with this problem with a dedicated classifier. The goal of this series of experiments is thus to classify the signals into two classes *C*2 and *C*3.1, so as to solve the main problem met in the previous series of experiments.

Repeating the experiment 10 times, once again, LDC and QDC achieved the maximum performance in the fourth frequency range. When used with 10 RDFs, the LDC and QDC classifiers achieved the accuracy shown in Table XVI.

To improve the result obtained by the LDC and QDC classifiers, we resort to *classifier fusion* [9,13]. More precisely, we use different classifiers and then appropriately combine their responses. We used nine classifiers (Table XVII). The MLPs used are characterized by one hidden layer and logarithmic sigmoid transfer functions.

TABLE XVI. CLASSIFICATION OF *C*2, *C*3.1. ACCURACY FOR LDC AND QDC IN THE FIVE FREQUENCY RANGES (10 FEATURES)

Range	Frequency range	Accuracy of LCD (mean over 10 trails)	Accuracy of QDC (mean over 10 trials)
1	1-50 Hz	71.84 %	73.21 %
2	51-100 Hz	65.81 %	68.88 %
3	101-150 Hz	68.36 %	67.14 %
4	151-200 Hz	**91.85 %**	**91.95 %**
5	201-250 Hz	83.00 %	83.99 %

In particular, we also introduced another method of feature selection, IFS (Individual Features Selection). This method takes into account the accuracy achieved by features used singularly and not combined together (like in forward feature selection). The nine classifiers were combined by means of the majority rule achieving an accuracy of 94.35 % (Table XVII) (mean over 10 trials). Table XVIII shows an example of the related confusion matrix.

TABLE XVII. CLASSIFICATION OF *C*2, *C*3.1. CLASSIFIER FUSION

Classifier	Neurons in the hidden layer	Number of features	Feature selection	Accuracy (mean)
LDC	---	20	---	91.97 %
LDC	---	10	---	91.85 %
QDC	---	20	---	92.16 %
QDC	---	10	---	91.95 %
MLP	20	20	FFS (LDC)	89.52 %
MLP	20	20	FFS (QDC)	90.86 %
MLP	20	15	FFS (QDC)	90.19 %
MLP	40	10	IFS (LDC)	89.78 %
MLP	40	10	IFS (QDC)	88.88 %
9-classifier Combiner	**---**	**---**	**---**	**94.35 %**

TABLE XVIII. CLASSIFICATION OF *C*2, *C*3.1. CLASSIFIER FUSION. CONFUSION MATRIX FOR THE TEST SET

		Estimated labels		Total
		***C*2**	***C*3.1**	
True labels	***C*2**	489	42	531
	***C*3.1**	18	513	531
Total		507	555	1062

We wish to point out that the obtained accuracy is higher than that of the best of the nine classifiers and furthermore, in this way, we can also significantly increase the robustness of the resulting classification system.

IV. CONCLUSIONS

In this paper we have presented an automatic method, based on classification techniques and classifier fusion, for diagnosing defects of rolling element bearings.

The proposed method has been applied to experimental data, registered by four accelerometers, and related to four different defects with different severities on rolling element bearings. The method has proved to be highly sensitive both to different defects and to different degrees of severity for the considered defects. We achieved an accuracy on the test set greater than 94 % for all the classification cases taken into consideration (sometimes reaching almost 100 % accuracy).

ACKNOWLEDGMENT

We wish to acknowledge Avio Propulsione Aerospaziale, via I Maggio, 99, Rivalta di Torino, Italy, for having provided the set of experimental data used for the present paper.

REFERENCES

[1] A. H. C. Tsang, "Condition-based maintenance: tools and decision making," Journal of Quality in Maintenance Engineering, vol. 1, no. 3, pp. 3-17, 1995.

[2] J. H. Williams, "Condition-based maintenance machine diagnostics," Kluwer Academic Publisher, London UK, 1994.

[3] B. Li, M.-Y Chow, Y. Tipsuwan, J. C. Hung, "Neural-Network-Based Motor Rolling Bearing Fault Diagnosis," IEEE Transactions on Industrial Electronics, vol. 47, no. 5, pp. 1060-1069, 2000.

[4] M. Cococcioni, P. Forte, S. Manconi, C. Sacchi, "Rolling bearing monitoring using classification techniques," 8th International Conference on Vibrations in Rotating Machines (SIRM'09), Vienna, Austria, February 2009.

[5] B. Samanta, K. R. Al-Balushi, S.A. Al-Araimi, "Artificial neural networks and genetic algorithm for bearing fault detection," Soft Computing, vol. 10, pp.264-271, 2006.

[6] B. Samanta, K. R. Al-Balushi, S.A. Al-Araimi, "Bearing fault detection using artificail neural networks and genetic algorithm," Journal on Applied Signal Processing, vol. 3, pp. 366-377, 2004.

[7] N. Nguyen, Lee H., "Bearing fault diagnosis using adaptive network based fuzzy inference system," International Symposium on Electrical & Electronics Engineering, HCM City, Vietnam, 2007.

[8] N. Tandon, A. Choudhury, "A review of vibration and acoustic measurement methods for the detection of defects in rolling element bearings," Tribology International, vol. 32, no. 8, pp. 469-480, 1999.

[9] L. I. Kuncheva, Combining Pattern Classifiers: Methods and Algorithms, Wiley Interscience, New Jersey, USA, 2004.

[10] R.O. Duda, P.E. Hart, and D.G. Stork, Pattern classification, 2nd edition, John Wiley and Sons, New York, 2001.

[11] A. Webb, Statistical Pattern Recognition, 2nd edition, John Wiley & Sons, New York, 2002.

[12] C. Liu and H. Wechsler, "Robust Coding Schemes for Indexing and Retrieval from Large Face Databases," IEEE Transactions on Image Processing, vol. 9, no. 1, pp. 132-136, 2000.

[13] T. G. Dietterich, "Ensemble methods in machine learing,". in J. Kittler and F. Roli, editors, Multiple Classifier Systems, Lecture Notes in Computer Science, vol. 1857, pp. 1-15, Springer, Cagliari, Italy, 2000.

[14] R. P. W. Duin, P. Juszczak, P. Paclik, PRTools4: A Matlab Toolbox for Pattern Recognition, version 4.1, Delft Pattern Recognition Research, Delft University of Technology, Delft, The Netherlands, 2007.