Wavelet and Artificial Neural Network Comparison Results in Classification of Power Quality Disturbances

A. Rodríguez, J. Aguado, J.J. López, F. Martín, F. Muñoz, E. Ruiz, J. Muñoz and M. Medina

Departamento de Ingeniería Eléctrica
Universidad de Málaga
Plaza El Ejido s/n, 29014 Málaga (España)
Telef:+34 952 134195, fax:+34 952 131091, e-mail: ale@uma.es; fimartin@uma.es

Resumen. This paper presents classification power quality disturbances comparison of results. This is a supervised classification based on multirresolution Wavelet analysis for feature extraction. The comparison is made in two levels: first, using different wavelet mother, to determinate whether the classification performance depends on the choice of wavelet mother, analyzing the noise influence and the filter length. The second level using two different Artificial Neural Network (ANN): backpropagation (BP) and probabilistic (PNN).
Besides two different patterns are obtained, the first is based on a feature extraction, the energy of wavelet detail levels with no very good results, and a second pattern based on the first one adding a feature selection obtained from the original signal, the maximun and minimun Root Mean Square (RMS) value, with very goods and robust results.
It is found that the classification performance depends majorly on the suitable features and not on the wavelet mother. Also, the performance for each ANN depends on the pattern used, obtaining better results with BP than PNN.

Palabras llave

Power Quality, classification, Wavelet transform, neural network.

1. Introduction

Power quality analysis is a classical problem in electrical engineering since the voltage signal contains transients disturbances that can be caused for diverse reasons: switching of loads, unbalanced power system, distributed generation with renewable energy, by the use of loads with electronic control, etc.
It is necessary to analyze efficiently and to understand deeply these disturbances One way is to establish an automated classification of these disturbances.
In supervised classification, or simply classification, the mapping from a set of input data vectors to a finite set of discrete class labels is modelled in terms of some mathematical function [1]. The values of these parameters are determined (optimized) by an inductive learning algorithm (also named inducer), whose aim is to minimize an empirical functional risk (related to an inductive principle) on a finite data set of input–output examples [2][3]. When the inducer reaches convergence or terminates, an induced classifier is generated [4].
One of the keys in the classification scheme is the pattern design. A suitable selection can greatly reduce the workload and simplify the design process that followed. The pattern design consists in to identify characteristics that facilitate joint distinctive patterns belonging to each group, immune to noise and easy to obtain and interpret. Obtaining these features is done in two ways: selection and feature extraction. The selection is to elect directly or calculated characteristics from the original data. Moreover extraction features need some transformations used to generate useful features. In power quality disturbances feature extraction techniques include frequency transform such as Fourier Transform (FT) and the time-frequency transform, such as Wavelet Transform (WT) and S-Transform (ST). In this work has beeb used WT.
Another important aspect is the selection or design of the classifier. There are three different approaches [5]. The first approach is the simplest and the most intuitive, and is based on the concept of similarity. The second one is a probabilistic approach and includes methods based on Bayes decision rule. The third approach is to construct decision boundaries directly by optimizing certain error criterion. Examples of the latest are multilayer perceptrons, decision trees and support vector machines.
Once the algorithm classifier is selected or designed it has to be trained. At this stage, the classification algorithm is adjusted so that the function that assigns data to its corresponding class has the lowest possible error based on a particular set of training data. In this way the system learns to match the output of the function with the tag or class for each data. Once the classifier has been adjusted results are verified, in order to check the effectiveness of the whole system.
In this work, the norm EN 50160 [6] has been considered to generate 7000 signals, 1000 per each power quality disturbance, including voltage sag, interruption, voltage swell, flicker, oscillatory transient and voltage harmonics, moreover normal voltage.

The 95% of these signals, 6650, 950 for each class, have been used for training, and the rest, 350, 50 for each class, to verify the effectiveness of system.
The signals have been simulated using MATLAB [7], according to power disturbance signal models [8], with sampling frequency of 3.2 kHz, which is equal to 64 samples per cycle, and a length of five cycles or 100 ms. The fundamental frequency is 50 Hz and the virtual value of voltage is 1.
A random white noise of zero mean and the signal to noise ratio varying from 40dB to 20dB has been added to the signals. A typical SNR value of 30 dB is equivalent to a peak noise magnitude of nearly 3.5% of the voltage signal.
In addition to this introduction, the article is organized as follows. The second part is devoted to a concise development of multirresolution wavelet analysis, necessary for the feature extraction. Section 3 presents classification results using a pattern obtaining from the wavelet analysis. Different wavelet mothers and noise levels, and two classifiers, BP and PNN, are tried. Section 4 presents classification results with a second pattern, based in the first one adding selected parameters of signal, RMS values, with different wavelet mothers, noise levels and the same two classifiers. At last the conclusions are presented in order to compare the results.

2. Wavelet transform

A wavelet is a small wave which has its energy concentrated in time and is a valuable tool for the analysis of transient non-stationary or time-varying phenomena [9].
A function *f(t)* can be expressed as a linear decomposition as follows:

$$f(t) = \sum c_l \Psi_l(t) \tag{1}$$

where l is an integer index , c_l is a real coefficient, and $\psi_l(t)$ is a set of functions called the expansion set. If the expansion set is orthogonal, then:

$$\left\langle \Psi_k(t), \Psi_l(t) \right\rangle = 0 \qquad k \neq l \tag{2}$$

where < > denotes the inner product.
Equation (1) can be written as:

$$f(t) = \sum_{j \in J} \sum_{k \in K} c_{j,k} \Psi_{j,k}(t) \tag{3}$$

where J and K are sets of integer indices and $\Psi_{j,k}(t)$ are the wavelet expansion functions. From (3), $f(t)$ can be expressed as:

$$f(t) = \sum_k \sum_j c_{j,k} \Psi_{j,k}(t) = \sum_k a_{J_0,k} \varphi_{J_0,k}(t) + \sum_{j=J_0}^{\infty} \sum_k d_{j,k} \Psi_{j,k}(t), \quad t \in R \tag{4}$$

where J_0 is an integer. Equation (4) is a linear combination of wavelet coefficients, ($a_{Jo,k}, d_{j,k}$), a set of functions $\varphi_{Jo,k}(t)$, called scalling function and $\Psi_{j,k}(t)$. Coefficients $a_{Jo,k}$ and $d_{j,k}$ are the Discrete Wavelet Transform (DWT) of $f(t)$, and can be calculated as:

$$a_{J_0,k} = \langle f(t), \varphi_{J_0,k}(t) \rangle \tag{5}$$

$$d_{j,k} = \langle f(t), \Psi_{j,k}(t) \rangle \tag{6}$$

Equation (4) can be truncated for $j=J-1$, obtaining:

$$f(t) = \sum_{k=0}^{2^{J_0}-1} a_{J_0,k} \varphi_{J_0,k}(t) + \sum_{j=J_0}^{J-1} \sum_{k=0}^{2^j-1} d_{j,k} \Psi_{j,k}(t), \quad t \in R \tag{7}$$

The first summation in (7) is a broad representation of $f(t)$ that has been expressed as a linear combination of 2^{J_0} translations of the scaling function, $\varphi_{Jo,0}$. The second summation contains the details of $f(t)$. For each level j, a linear combination of 2^j translations of the wavelet function, ψ_{j0}, are added to obtain a more accurate approximation of $f(t)$.
The Mallat algorithm [9] has been used in the practical implementation of DWT.
The DWT acts as two FIR (Finite Impulse Response) quadrature filters defined by two sequences $h(n)$ and $g(n)$. $h(n)$ is a high frequency filter and g(n) is a low frequency filter. Both filter the same cut frequency $f_N/2$, where f_N is the Nyquist frequency. Therefore the function $f(n)$ is split in two parts, the high frequency part d_1 that contains the higher octave and is called detail function, and the low frequency part a_1, that contains the frequencies lower than $f_N/2$, and is called smoothed function. Decimation by *2* is done for eliminating redundant information.
The algorithm is iterated for a_1, obtaining a second level detail function d_2 and a second level smoothed function a_2, that is again splitted, obtaining a series of detail and broad functions. The original function f(n) is split into a series of detail functions d_1, d_2, …, d_k, and a smoothed function a_k, that correspond to the frequencies:
$d_1 : f_N - f_N/2$; d_2: $f_N/2 - f_N/4$; ….. d_k: $f_N/2^n - f_N/2^{k+1}$;
and a_n contains the frequencies lower than $f_N/2^{k+1}$.

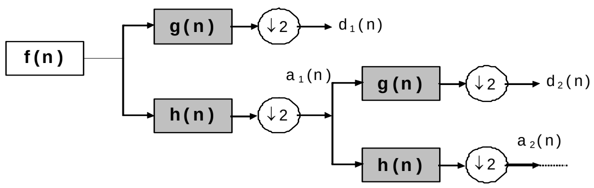

Fig. 3. Mallat algorithm schematics.

In this work, different wavelet mother has been used, performing 5 levels of decomposition. A voltage sag signal is shown in fig. 4, and it wavelet transform decomposition, using db5, can be seen in fig. 5.

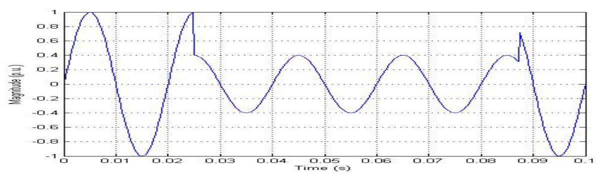

Fig. 4.Voltage sag with 60% depth and 400 samples length.

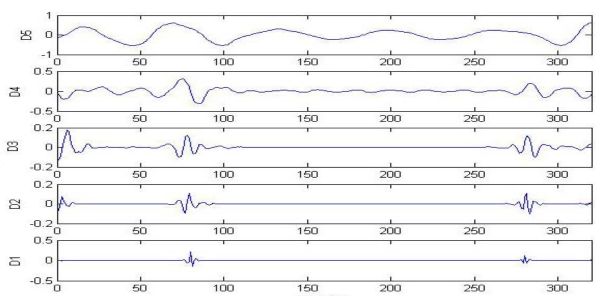

Fig. 5. Voltage sag wavelet analysis.

3. Artificial neural network

In this work the ANNs have been selected as algorithms classifiers. An ANN is composed of very simple elements operating in parallel. These elements are inspired by biological nervous systems. As in nature, the network function is determined largely by the connections between elements. An ANN can be trained to perform a particular function by adjusting the values of the connections between elements. In this paper two different ANNs have been used.

A. BP

Standard BP is a gradient descent algorithm, in which the network weights are moved along the negative of the gradient of the performance function. The term backpropagation refers to the manner in which the gradient is computed for nonlinear multilayer networks.
In this work the BP has been set with only two layers. The number of nodes in the hidden layer has been chosen as a function of number of inputs [10], according on the expression (2n+1), where n is number of inputs. So when the first pattern is used, build up for five features, the hidden layers has 11 nodes. For the second pattern, with seven features, the hidden layer has 15 nodes. In both cases, the output layer has one node. The network has been trained by the Levenberg-Marquardt algorithm. The transfer function for the hidden layer and output layer is tansigmoidal and lineal, respectively. The learning ratio is 0.1, and the epoch is 1000.

B. PNN

The basic principle of PNN is that is implemented using the probabilistic model, such as Bayesian classifiers. The training examples are classified according to their values of probabilistic density function. When an input is presented, the first layer computes distances from the input vector to the training input vectors, and produces a vector whose elements indicate how close the input is to a training input. The second layer sums these contributions for each class of inputs to produce a vector of probabilities as output. Finally, a complete transfer function on the output of the second layer picks the maximum of these probabilities [11].

4. Results with based energy pattern

The first pattern is based in the energy of different wavelet decomposition levels, so the pattern is made up of five values. These values have been normalized with reference to the values obtained from an ideal sinusoidal signal.

A. Results for different wavelet mother

This pattern has been calculated for the 7000 available signals, using 10 wavelet mothers indicated in Table I.
This pattern calculated with every wavelet mother has been used as input to the classifier, BP that is trained and subsequently verified. This process is carried out five times for every wavelet mother and the best result is always chosen as the representative. Likewise, the process has been repeated with different levels of noise, no noise, 40, 30 and 20dB.

TABLE I. – Wavelet mothers used

Wavelet family	Wavelet mother	Filter length
coiflets	coif1	6
	coif2	12
daubechies	db3	6
	db4	8
	db5	10
symlets	sym3	6
	sym4	8
	sym5	10
biorthogonal	bior3.7	4
	bior6.8	11

As it can be seen in Fig. 6, all the wavelet mothers give a figure definitely better than the rest, checking that all are affected by noise.

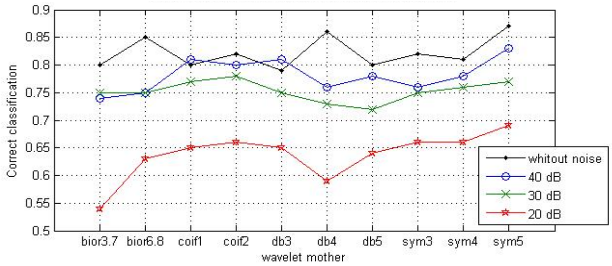

Fig. 6. Classification results for different wavelet mothers.

B. Results for daubechies mother with different filter lenght

The Fig. 7 shows the results of classification obtained by using the daubechies wavelet mother, db1 to db5, with filter length ranging from 2 to 10 (table 2). This shows that the filter length is not a relevant factor in the success of disturbances discrimination. It is also noted that noise affects different wavelet regardless of their length.

TABLE II. – Wavelet mothers used of daubechies family.

Wavelet family	Wavelet mother	Filter length
daubechies	db1	2
	db2	4
	db3	6
	db4	8
	db5	10

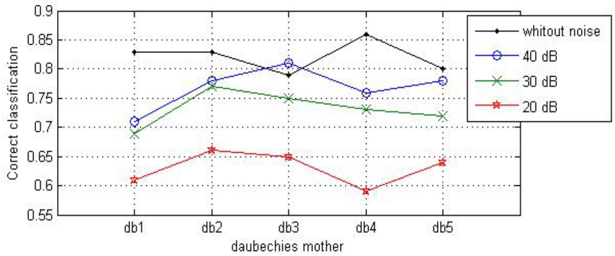

Fig. 7. Classification results for daubechies mother with different filter length

C. *Comparison of classification results for BP and PNN*

The following table shows a comparison of results using different wavelets mothers, signal with different noise level, and two different classifiers, BP and PNN.
In Table 3, the best results for each classifier are in bold for each level of noise and wavelet mother. In addition, the best results achieved for each level of noise is shown in white on black background. For example, BP performed better than PNN for signals without any noise for all wavelet mothers, except in the coif1 case. The best results are achieved with the sym5 wavelet mother. If noise is present, the PNN performs better than BP. The best results are obtained with sym5, bior6.8 and db3, with 40, 30 and 20 dB, respectively.

TABLE III. – Comparison of results using BP and PNN with based energy pattern.

	without noise		40 dB		30 dB		20 dB	
	PNN	**BP**	**PNN**	**BP**	**PNN**	**BP**	**PNN**	**BP**
bior3.7	0.43	**0.80**	**0.85**	0.74	**0.80**	0.75	**0.70**	0.54
bior6.8	0.64	**0.85**	**0.89**	0.75	**0.88**	0.75	**0.80**	0.63
coif1	**0.88**	0.80	**0.88**	0.81	**0.86**	0.77	**0.79**	0.65
coif2	0.76	**0.82**	**0.87**	0.80	**0.87**	0.78	**0.84**	0.66
db3	0.69	**0.79**	**0.86**	0.81	**0.87**	0.75	**0.84**	0.65
db4	0.50	**0.86**	**0.89**	0.76	**0.87**	0.73	**0.79**	0.59
db5	0.57	**0.80**	**0.85**	0.78	**0.84**	0.72	**0.8**	0.64
sym3	0.69	**0.82**	**0.86**	0.76	**0.87**	0.75	**0.84**	0.66
sym4	0.78	**0.81**	**0.88**	0.78	**0.86**	0.76	**0.83**	0.66
sym5	0.86	**0.87**	**0.90**	0.83	**0.86**	0.77	**0.78**	0.67

Figure 8 summarises the results shown in Table 3. A comparison between BP-PNN for all wavelet and two levels of noise (40 and 20dB, low and high respectively) is provided. The results obtained with PNN, regardless of the mother wavelet used, are noticeable better.

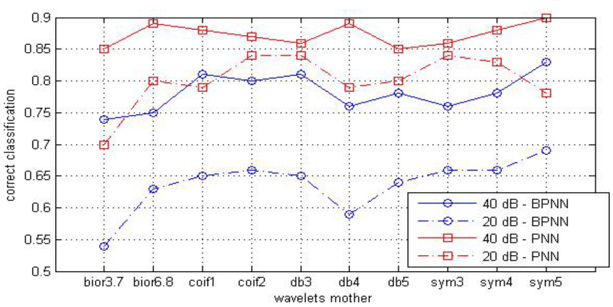

Fig. 8. Comparison of results PNN-BP for different wavelet mothers and 40 and 20dB noise levels.

D. *Results for disturbances class*

Results are presented in tables, where classification accuracy for each class of disturbance is given. The best results obtained with the first pattern for signals without any added noise are shown, are shown in Table IV, with coif1 as wavelet mother and PNN as classifier. The overall accuracy is 0.88.
Table V presents the best results obtained with the first pattern, 40dB added noise, sym5 as wavelet mother and a PNN classifier. The overall accuracy is 0.90. The worst results were found in the discrimination between the signal without disturbance and flicker.

TABLE IV. – Results for each disturbance class for the first pattern using coif1, non-noisy signals and PNN.

	sin	**sag**	**int**	**swell**	**flick**	**HF**	**harm**
sin	**0.9**	0	0	0.10	0	0	0
sag	0	**0.94**	0.06	0	0	0	0
int	0	0,16	**0.82**	0	0	0.2	0
swell	0	0	0	**1**	0	0	0
flick	0.40	0	0	0	**0.60**	0	0
HF	0.1	0	0	0	0	**0.9**	0
harm	0	0	0	0	0	0	**1**

Overall accuracy 0.88

TABLE V**. –** Results for disturbance class, using based energy pattern and PNN, with sym5 and 40 dB added noise.

	sin	**sag**	**int**	**swell**	**flick**	**HF**	**harm**
sin	**0,82**	0	0	0	0,18	0	0
sag	0	**0,96**	0,04	0	0	0	0
int	0	0,04	**0,96**	0	0	0	0
swell	0	0	0	**0,96**	0,04	0	0
flick	0,38	0	0	0	**0,62**	0	0
HF	0	0	0	0	0	**1**	0
harm	0	0	0	0	0	0	**1**

Overall accuracy 0.90

5. Results with based energy pattern and RMS

An energy-based pattern is not completely discriminatory because the energy of a magnitude disturbance depends on the depth of the disturbance and its duration in time. Thus a signal of magnitude 0.91 with a length of 4.5 cycles has less energy than a voltage sag of magnitude 0.8 but lasting 1 cycle (Fig. 9 and Table VI).

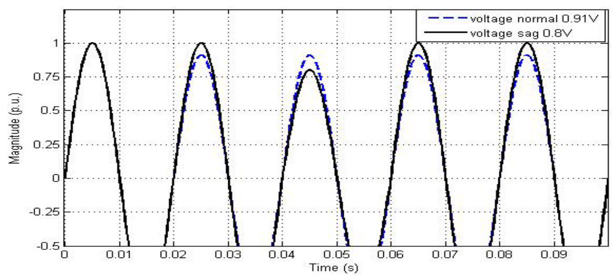

Fig. 9. Signals with different magnitude and duration.

TABLE VI. **–** Comparison of energy signals with different magnitude and duration.

	D1	D2	D3	D4	D5
Perfect signal	1	1	1	1	1
Normal 0.91 V	1,0103	0,98464	0,96797	0,90796	0,87283
Sag 0.8 V	1,043	1,0029	0,99816	0,94883	0,92061

Therefore it is necessary to provide a feature based on the magnitude of the signal. The RMS value is a widely accepted tool that provides information on how the magnitude of the voltage changes. It is a fast and simple algorithm that requires little computational resources.

The digital measurement instruments perform the calculation of this amount from the instantaneous values of the samples, choosing a temporary window depending on the frequency of the signal steady state. If the RMS values are updated every time a sample is acquired, the method is called RMS continuum. If the RMS values are updated at certain interval, usually half cycle, then it is called RMS (1/2).

In the present work the RMS has been calculated by selecting the maximum and minimum values of the voltage (see example in Fig. 10).

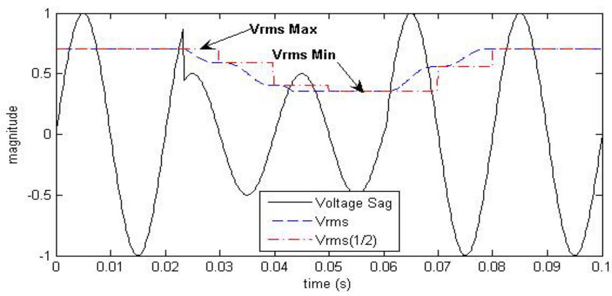

Fig. 10. RMS max and min in a voltage sag.

So seven features from the second pattern: the first five coincide with the first pattern. The two remaining features are the maximum and minimum RMS value (hereinafter RMS-Max and RMS-Min) calculated directly from each signal.

A. Results for different wavelet mothers

Here, the pattern calculated with every wavelet mother and RMS-Max and RMS-Min, has been used to feed the BP classifier. The classifier has been trained and verified five times for every wavelet mother, of which the best result is chosen as representative.

Also, the process has been repeated with different levels of noise: no noise, 40, 30 and 20dB. The results are displayed in graphical form to simplify the data presentation.

In Fig. 11, it can be seen that the results have been improved markedly over the first pattern and virtually all wavelet mothers used present a full success in the classification for 40 and 30dB noise levels. Even for significant levels of noise, 20 dB, success rates are higher than 0.95, except for the biorthogonal wavelet family with success levels around 0.92. Likewise, the results are very similar for all the different wavelet mothers.

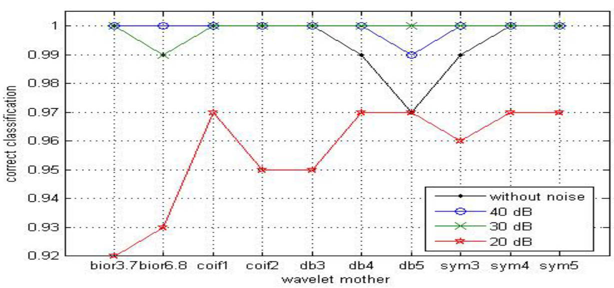

Fig. 11. Comparison results for different wavelet mother.

A. Comparison of results for wavelet mothers with differents filter lenght

Here classification results are compared using different wavelets mother within the family daubechies, from the db1 to db5, filter lengths from 2 to 10, respectively. This shows the relevance of the filter length on the success of the classification.

All wavelet mothers lead to good results with the exception of the db1 when levels of noise are present (20 dB), where the success the success rate is 0.92.

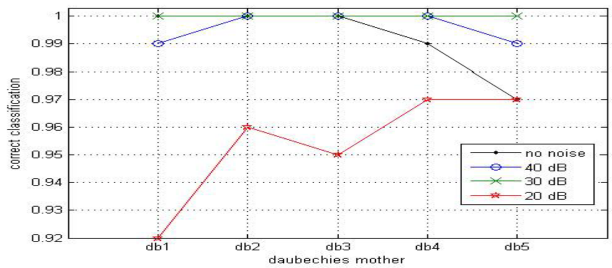

Fig. 12. Comparison results for wavelet mother with different filter length

Figure 13 presents the comparison of the BP-PNN classifiers with the second pattern. The results are given for all wavelets mother and two levels of noise, 40 dB and 20 dB, low and high respectively.

One can see that the results obtained with BP are significantly better regardless of the mother wavelet used. Even the results obtained with this network for signals with high noise level, 20 dB, are better than with PNN and low noise levels.

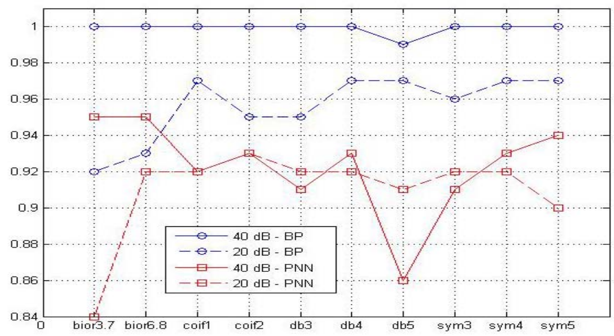

Fig. 13. Comparison of results BP–PNN using different wavelet mothers and 40 and 20dB noise levels.

B. Results for daubechies mother with different filter lenght

Here classification results are compared using different wavelets mother within the family daubechies, from the db1 to db5, filter lengths from 2 to 10, respectively. This shows the relevance of the filter length on the success of the classification.

All wavelet mothers lead to good results with the exception of the db1 when levels of noise are present (20 dB), where the success the success rate is 0.92.

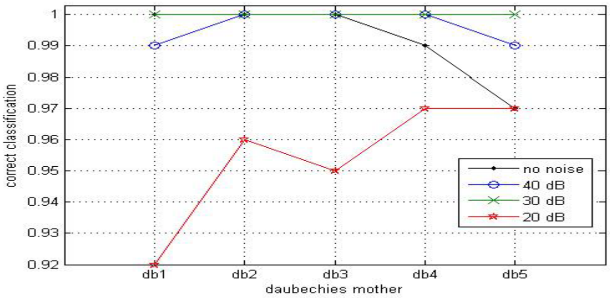

Fig. 12. Comparison of classification results for daubechies mother with different filter length.

C. Comparison of classification results for BP and PNN

Figure 13 presents the comparison of the BP-PNN classifiers with the second pattern. The results are given for all wavelets mother and two levels of noise, 40 dB and 20 dB, low and high respectively.

One can see that the results obtained with BP are significantly better regardless of the mother wavelet used. Even the results obtained with this network for signals with high noise level, 20 dB, are better than with PNN and low noise levels.

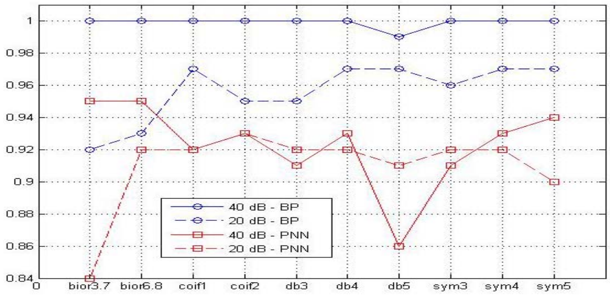

Fig. 13. Comparison of classification results BP–PNN using different wavelets mother and 40 and 20dB noise levels.

7. Conclusions

In this paper, we have proposed the use of DWT for obtaining a pattern based on the energy levels of wavelet decomposition. No wavelet mother gave results significantly better than the others to motivate their choice. One may say that the results are independent of the mother wavelet chosen and the filter length. Therefore it would be advisable to use a wavelet mother with short filter length to reduce computational cost if the goal is to implement a classification in real time.

As energy depends on the magnitude and time of the event, it can give misleading information on disturbance magnitude, so it is appropriate to incorporate RMS to the selected features. A pattern based on energy levels of wavelet decomposition together with the RMS values leads to robust classification results.

The choice of ANN depends on the pattern considered.

Referencias

[1] Rui Xu, and Donald Wunsch II, “Survey of Clustering Algorithms” IEEE Transactions on Neural Networks, Vol. 16, No. 3, May 2005.

[2] C. Bishop, Neural Networks for Pattern Recognition. New York: Oxford Univ. Press, 1995.

[3] V. Cherkassky and F. Mulier, Learning From Data: Concepts, Theory, and Methods. New York: Wiley, 1998.

[4] R. Kohavi, “A study of cross-validation and bootstrap for accuracy estimation and model selection,” in Proc. 14th Int. Joint Conf. Artificial Intelligence, 1995, pp. 338–345.

[5] Meijer, B.R.; “Rules and algorithms for the design of templates for template matching”, Pattern Recognition, 1992. Vol.1. Conference A: Computer Vision and Applications, 11th IAPR International Conference on, pp: 760 – 763, Aug. 1992.

[6] EN 50160, Voltage Characteristic of electricity supplied by public distribution systems, CENELEC, Brussels, Belgium, 1999.

[7] MATLAB. Natick, MA: Math Works, Inc., 2000.

[8] Youssef, A.M. Abdel-Galil, T.K, El-Saadany, E.F., and Salama, M.M.A, “Disturbance classification utilizing dynamic time warping classifier,” IEEE Trans. On Power Delivery, vol. 19, no 1, pp.272-278, 2004.

[9] S. Mallat, A wavelet Tour of signal processing, Academic Press, 1999.G. O. Young, "Synthetic structure of industrial plastics," in *Plastics*, 2nd ed., vol. 3, J. Peters, Ed. New York: McGraw-Hill, 1964, pp. 15-64.

[10] Hecht-Nielsen, R, "Counterpropagation networks", Applied Optics 26: 4979-4984. 1989.

[11] D. F. Specht, “Probabilistic neural networks,” Neural Netw., vol. 3, no. 1, pp. 109–118, 1990.