Ultrasonics 73 (2017) 1–8

Contents lists available at ScienceDirect

Ultrasonics

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

Efficient feature selection for neural network based detection of flaws in steel welded joints using ultrasound testing

F.C. Cruz [a,b], E.F. Simas Filho [b,*], M.C.S. Albuquerque [c], I.C. Silva [c], C.T.T. Farias [c], L.L. Gouvêa [b]

[a] Exact and Technology Sciences Department, State University of Santa Cruz, Ilhéus, Brazil
[b] Electrical Engineering Program, Federal University of Bahia, Salvador, Brazil
[c] Ultrasound Testing Laboratory, Federal Institute for Science, Education and Technology of Bahia, Salvador, Brazil

ARTICLE INFO

Article history:
Received 20 January 2016
Received in revised form 7 August 2016
Accepted 23 August 2016
Available online 24 August 2016

Keywords:
Ultrasonic evaluation
Welded joints
Feature extraction
Neural networks
PCA

ABSTRACT

This work studies methods for efficient extraction and selection of features in the context of a decision support system based on neural networks. The data comes from ultrasonic testing of steel welded joints, in which are found three types of flaws. The discrete Fourier, wavelet and cosine transforms are applied for feature extraction. Statistical techniques such as principal component analysis and the Wilcoxon-Mann-Whitney test are used for optimal feature selection. Two different artificial neural network architectures are used for automatic classification. Through the proposed approach, it is achieved a high discrimination efficiency by using only 20 features to feed the classifier, instead of the original 2500 A-scan sample points.

1. Introduction

Welding may be defined as a process of joining of materials (usually metals or thermoplastics) by bringing their surfaces close enough together for atomic bonding to occur [1]. In modern industries such as aerospace, oil and gas, manufacturing and construction, welding processes are critically important. Non-destructive evaluation (NDE) techniques have been applied to verify the integrity of welded components. Among them there are ultrasound testing, X-ray imaging, and also some optical, electrical and magnetic measurements [2]. The ultrasound testing is widely used for weld evaluation as it combines high-efficiency with simple and safe execution procedure.

The accuracy of an ultrasound testing procedure usually relies on the operator experience. In this context, digital signal processing techniques (such as signal transformations [3] and neural networks [4]) have been proposed for ultrasound testing problems in order to provide decision support information. For example, both Fisher linear discriminant and three-layered neural network are applied in [5] for the detection of welding defects in steel plates. The input parameters for the neural network based classifier are pre-selected from the Fisher linear discriminant analysis. The combination of wavelet decomposition and an artificial neural network classifier is used in [6] for the detection of flaws in thin welded steel plates. Different defects such as porosity, tungsten inclusion and lack of fusion are considered and high classification efficiency (approximately 94%) is achieved. Other applications of neural networks for automatic ultrasound evaluation of welded joints can be found in [7,8]. Ultrasonic guided waves are used in [9] for the detection of small notch cuts in ASTM-A53-F steel pipes. Wavelet, Hilbert and Fourier transforms are applied for feature extraction and an artificial neural network is used for automatic classification.

Classifiers based on artificial neural networks have been also widely applied for different purposes such as condition assessment of electrical equipment [10], on-line event selection in particle detectors [11], command of a wheelchair using a brain-computer interface [12], detection of oil spill in the sea using synthetic aperture radar information [13] and decision support for bio-medical diagnostics [14].

An important aspect to be considered in automatic classification systems is the proper choice of the input features set [4,15]. Statistical signal processing techniques are usually employed for this purpose [16], as the presence of non-relevant or redundant information may contribute to decrease the classification efficiency. The neural network based classifiers usually benefit from compact and relevant set of input features.

This paper proposes the application of an artificial neural network based classifier [4] for decision support in ultrasound testing of steel welded joints. Three welding defects are considered: lack

* Corresponding author at: Rua Aristides Novis, n.02 Federação, Salvador, Brazil.
E-mail address: eduardo.simas@ufba.br (E.F. Simas Filho).

http://dx.doi.org/10.1016/j.ultras.2016.08.017
0041-624X/

of fusion, porosity and slag inclusion. One of the main contributions of this work is the use of different signal processing techniques such as Fourier, wavelet and cosine transforms to extract discriminant information from the A-scan (time-domain) signals. The discrimination efficiency obtained by feeding the classifier from these features is compared. Another important contribution is the application of statistical feature selection methods, such as principal component analysis [17] and significance tests [15], to select a compact set of relevant and discriminating features. Additionally, two different classifier architectures are proposed for the flaw detection problem.

This work is organized as follows: the used experimental setup is presented in Section 2; the proposed signal processing chain is defined in Section 3; the experimental results are described in Section 4; and finally, the conclusions are presented in Section 5.

2. Problem definition

The test piece used in this work is a carbon steel 1020 plate (400 mm × 310 mm × 19 mm) welded through both tungsten inert gas (TIG) and shielded metal arc welding (SMAW) processes. Welding defects (lack of penetration, porosity and slag inclusion) were generated during the manufacturing process and their precise locations were confirmed through X-ray testing. The classification problem consists thus on assigning, for each signal, one of the four classes of interest: normal condition (ND - no defect), lack of penetration (LP), slag inclusion (SI) and porosity (PO). Although the used test piece did not present cracks, which are serious welding defects, the considered flaws may generate a crack when the weld is tensioned. Moreover, weld cracks are efficiently detected using different NDE methods (i.e. magnetic-particle or liquid penetrant, in the case of superficial cracks, and ultrasound or X-ray, for internal cracks), and the proposed system can also be easily adjusted to allow crack identification.

In this work, an experimental setup composed of Krautkramer USM-25 ultrasonic signal generator and 5 MHz transducers is used for signal acquisition (see Fig. 1). For system design, approximately 100 A-scan signals (pulse-echo configuration) for each class of interest were recorded. For this, the transducers were placed at several positions along two lines parallel to the weld bead (the adopted perpendicular distances from the weld bead center line were 43.3 mm and 37.5 mm).

The analog to digital conversion is performed using an oscilloscope with maximum sampling rate of 500 MHz and 8 bits for data quantization. The digital signal processing routines are executed in a personal computer, which is connected to the oscilloscope through an USB interface. The signal generator provided the triggering information for the oscilloscope in order to initialize the data acquisition. For each acquired signal 2048 time-samples are generated.

3. Proposed method

The digital signal processing chain proposed in this work comprises three steps (see Fig. 2). Initially, feature extraction is performed over the measured signals. In the following, the most relevant features are selected as inputs to the classifier system, which is the third processing step. The proposed signal processing chain is described in the following sub-sections.

3.1. Feature extraction

In this work, feature extraction is performed using different algorithms such as the discrete Fourier transform (DFT), the discrete cosine transform (DCT) and the discrete wavelet transform (DWT), which are described briefly in this section.

The DFT is a traditional method to obtain the frequency-domain information of a given discrete-time signal. It is widely applied for feature extraction in ultrasound testing problems (see for example [18,19]). The DFT for a finite-length discrete-time signal $x[n]$ can be obtained from Eq. (1) [3]:

$$X\left(e^{\frac{j2\pi k}{N}}\right) = \sum_{n=0}^{N-1} x[n] e^{\frac{-j2\pi kn}{N}}, \quad 0 \leqslant k \leqslant N-1. \tag{1}$$

The DCT is a linear transformation that expresses a finite-length discrete-time sequence $x[n]$ as a linear combination of cosine functions. It is widely applied in image processing for information compaction (see for example [20,21]), as it concentrates the signal energy into a small number of coefficients. The DCT can be defined as in Eq. (2) [3]:

$$C(k) = \alpha(k) \sum_{n=0}^{N-1} x[n] \cos\left[\frac{\pi(n+0.5)k}{N}\right], \quad 0 \leqslant k \leqslant N-1, \tag{2}$$

$$\text{where } \begin{cases} \alpha(k) = \sqrt{1/N}, & k = 0; \\ \alpha(k) = \sqrt{2/N}, & 1 \leqslant k \leqslant N-1. \end{cases}$$

Different from the two previous transformations, the wavelet analysis makes use of finite-length functions $\psi(t)$ (called

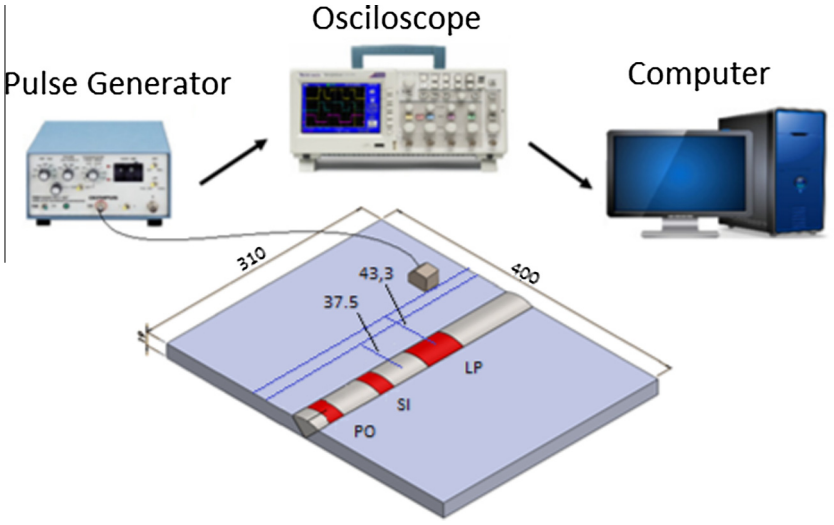

Fig. 1. Diagram of the used experimental setup, the defects locations are indicated in the test-piece.

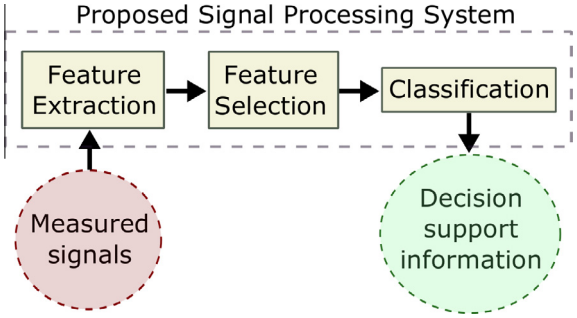

Fig. 2. Diagram of the proposed signal processing system.

mother-wavelets), instead of sinusoidal functions, to create a representation for a given time-domain signal. Scaled and shifted versions of the mother wavelet function are summed in order to approximate the signals of interest. For a continuous-time signal $f(t)$ the wavelet transform $F(a,b)$ is defined in Eq. (3) [22]:

$$F(a,b) = \int_{-\infty}^{\infty} f(t)\psi_{a,b}(t)dt, \tag{3}$$

where $\psi_{a,b}(t) = \frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right)$, a and b are scale and shift parameters, respectively. The mother-wavelet function ($\psi(t)$) must satisfy Eq. (4):

$$\int_{-\infty}^{\infty} \psi(t)dt = 0. \tag{4}$$

The DWT is a discrete-time counterpart of the wavelet transform and it can be implemented through a set of hierarchical filters [3]. Low-pass filters are used to obtain the approximation coefficients ($A_k[n]$) and high-pass filters produce the detail coefficients ($D_k[n]$). At each decomposition level (k) the signal is down-sampled by a factor of 2.

The main issues in the design of a DWT-based signal processing method are related to the choice of mother-wavelet function and decomposition level which better describe the available data-set. In this work, these choices are made experimentally, as described in Section 4.

3.2. Feature selection

The process of selecting features for a given pattern recognition problem is important to the final system performance. It is known that the dimension of the features vector (i.e. the number of used input features) has a direct impact on the performance of the classification system [23,24,16]. If the used dimension is too small, relevant information may be neglected, contributing to decrease the system efficiency. In contrast, by using a too large number of input features greater computational resources are needed. Unfortunately, this does not always ensure an increase in the classification accuracy. In fact, if among the used features there is irrelevant and redundant information, this may contribute to disturb the convergence of the training process.

In this work different methods are used for feature selection. In order to achieve optimal information compaction (in the mean-square error sense), Principal Component Analysis (PCA) [17] is applied. Statistical significance tests such as t-test and Wilcoxon-Mann-Whitney (WMW) test [15] are also used in order to select class-discriminant features.

PCA projects the original feature vector $\mathbf{x} = [x_1, \ldots, x_N]^T$ into principal components (see Eq. (5)):

$$z_i = \mathbf{v}_i^T\mathbf{x} = \sum_{k=1}^{N} v_{ki}x_k, \quad i = 1, \ldots, N. \tag{5}$$

The obtained z_i are mutually uncorrelated and ordered (decreasingly) by their explained variance.

Assuming that $\mathcal{E}\{\mathbf{x}\} = 0$, where $\mathcal{E}\{\cdot\}$ is the expectation operator, then $\mathcal{E}\{z_i\} = 0$. In this case, the variance of the projection z_i can be computed through $\mathcal{E}\{z_i^2\}$. Considering this, the direction of maximum variance $\mathbf{v}_1$ may be found from the maximization of J_1^{PCA}, as defined in Eq. (6) [25]:

$$J_1^{PCA}(\mathbf{v}_1) = \mathcal{E}\{z_1^2\} = \mathcal{E}\{(\mathbf{v}_1^T\mathbf{x})^2\} = \mathbf{v}_1^T\mathcal{E}\{\mathbf{x}\mathbf{x}^T\}\mathbf{v}_1 = \mathbf{v}_1^T\mathbf{C}_\mathbf{x}\mathbf{v}_1, \tag{6}$$

where $\mathbf{C}_\mathbf{x} = \mathcal{E}\{\mathbf{x}\mathbf{x}^T\}$ is the $N \times N$ covariance matrix of $\mathbf{x}$.

The maximization of Eq. (6) can be solved using linear algebra [26] in terms of the decomposition of $\mathbf{C}_\mathbf{x}$ into its eigenvalues λ_i and eigenvectors $\mathbf{e}_i$ ($i = 1, \ldots, N$). The order of the eigenvectors is such that the associated eigenvalues satisfy: $\lambda_1 > \lambda_2 > \cdots > \lambda_N$. Thus, in PCA (see Eq. (7)):

$$\mathbf{v}_i = \mathbf{e}_i, \quad 1 \leqslant i \leqslant N. \tag{7}$$

Alternatively to PCA, statistical feature selection methods aim to estimate the features relevance for classification tasks. In this context, t-test assumes that the feature probability distribution function is Gaussian. Unfortunately, this is not the case for our particular application, as it will be shown in Section 4. Considering this, the t-test cannot be applied in this work for feature selection.

The Wilcoxon-Mann-Whitney rank test is non-parametric. It means that no previous assumptions are made on the features probability distribution function, which is an interesting characteristic for this application. The null hypothesis is that class A and class B samples for a given feature x_i came from the same population. The alternative hypothesis is that there is a shift between the distribution of the two classes (in this case the feature x_i may be useful for class discrimination). To compute the WMW test, the values of x_i are ranked. For the smallest observation is assigned rank 1 and for the highest observation is assigned rank N (where N is the number of x_i samples).

Defining W_A as the sum of the ranks of the observations from x_i for class A, and considering that there are m_A samples of x_i for class A and m_B samples of x_i for class B, the U statistics can be calculated using Eq. (8) [27]:

$$U = m_A m_B + \frac{m_A + m_B + 1}{2} - W_A \tag{8}$$

The value of U is then compared to the critical value (which is obtained for given m_A and m_B) in order to accept or reject the null hypothesis.

3.3. Used classifier architectures

For signal classification two different classifier architectures are proposed in this work. In both, a multilayer perceptron (MLP) neural network comprising two layers of neurons (hidden and output) is used. The number of hidden neurons is defined after exhaustively testing the classification performance for different network configurations. The hyperbolic tangent activation function is used for all neurons, and the Levenberg-Marquardt algorithm [28] is applied for training.

A diagram of the proposed classifier architecture is shown in Fig. 3. The output of the j-th neuron from layer l is computed from Eq. (9):

$$v_{lj} = \varphi\left(\sum_{i=1}^{N} \omega_{lij}x_i + b_{lj}\right) \tag{9}$$

where $\varphi(\cdot)$ is the activation function, ω_{lij} are the neuron weights, b_{lj} is the bias, and x_i are the inputs.

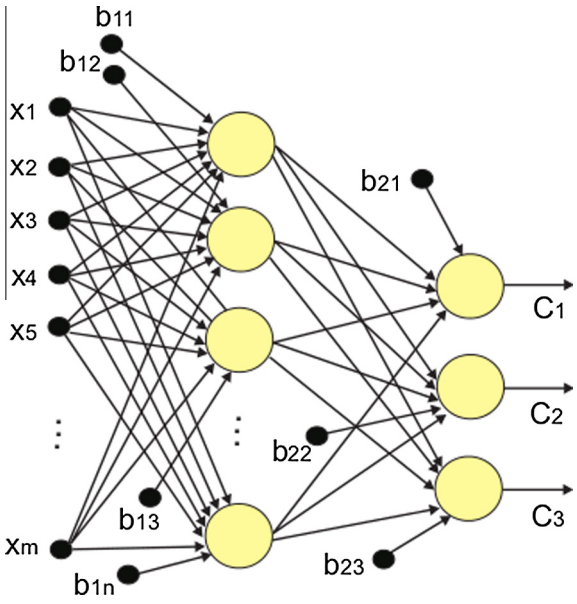

Fig. 3. Diagram of the used MLP architecture.

Different steps for signal classification are executed according to the applied methodology, as it is described in the following:

- Methodology 1: In this case, a four-class classifier is used to identify the four classes of interest (no-defect, lack of penetration, porosity and slag inclusion). For this, four neurons are used in the output layer. The target outputs for class i signals are $y_1 = [1,-1,-1,-1]^T$, $y_2 = [-1,1,-1,-1]^T$, $y_3 = [-1,-1,1,-1]^T$ and $y_4 = [-1,-1,-1,1]^T$. In the classifier operation, the neuron which presents the highest output value is associated to the detected class.
- Methodology 2: Here is proposed the use of a two-stage classification system. The first stage (classifier 1) is designed to separate the signatures into two groups: no-defect and defect. For this, a binary MLP classifier is applied. The second stage (classifier 2) is trained to identify the type of detect: porosity; slag inclusion and lack of penetration (see Fig. 4). The classifier 2 is similar to the one used in Methodology 1, except that, in this case, there are only three classes of interest.

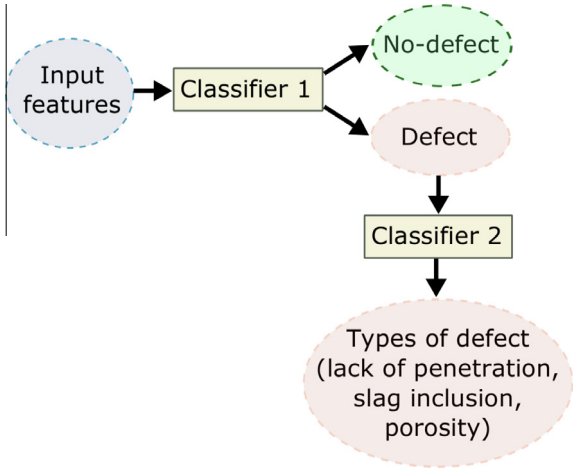

Fig. 4. Diagram of the proposed classifier architecture (methodology 2).

3.4. Performance evaluation and comparison

There are several ways to evaluate the performance of a classification system, for more details see [29–31]. In this work efficiencies product (EP) and Receiver Operating Characteristics (ROC) curve are applied. These parameters are computed from the testing set. The EP is defined as:

$$\mathrm{EP} = \sqrt[N]{\prod_{i=1}^{N} Ef_i}, \tag{10}$$

where Ef_i is discrimination efficiency of class i and N is the total number of classes. When $\mathrm{EP} = 1$, the classifier presents 100% of accuracy.

The ROC is used to visualize and evaluate the performance of binary classifiers as it shows how both, the detection and the false alarm probabilities vary as the decision threshold is adjusted [31].

4. Experimental results

The used database comprises 400 A-scan signals, 100 for each class of interest: no-defect (ND); lack of penetration (LP); slag inclusion (SI) and porosity (sPO). The available data is divided into training, validation and testing sets (using 50%, 20% and 30% of the available signatures, respectively). Fig. 5 illustrates typical A-scan signatures (each one comprised 2500 time-domain samples). It can be observed that identifying the differences among these signatures in the time-domain is a difficult task. This fact reinforces the need for an automatic decision support system to assist the operator.

4.1. Results - methodology 1

Using the methodology 1 a MLP classifier is trained using as inputs the A-scan signatures. In this case a poor average classification efficiency is obtained (EP ~ 53.1%). This result motivates the use of preprocessed information to feed the classifiers.

Considering the DFT and DCT as preprocessing methods, it is observed that most of the information is contained in the first

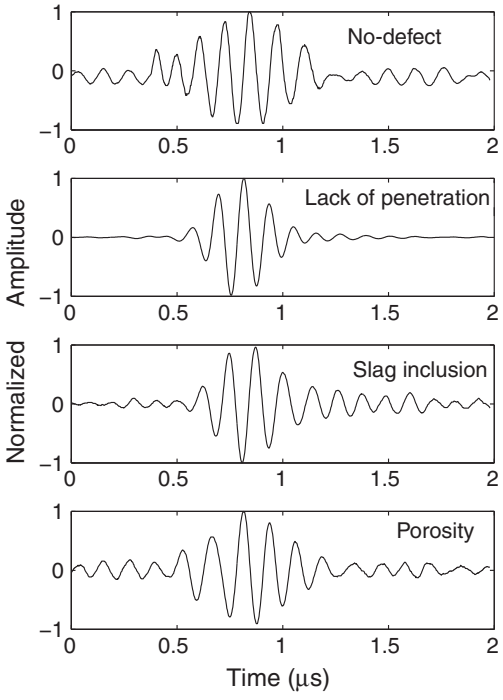

Fig. 5. Typical A-scan signals from the four classes of interest.

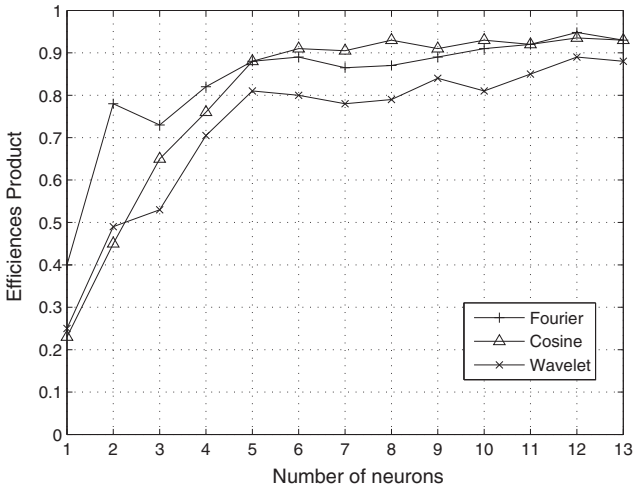

Fig. 6. Efficiencies product as a function of the number of neurons in the classifier hidden layer (methodology 1).

Table 1
Discrimination efficiencies (in %) for the classes of interest, and the efficiencies product (EP), considering different input features for the classification methodology 1.

Feat. ext.	ND	LP	SI	PO	EP
None (A-Scan)	66.7	42.4	44.1	63.6	53.1
DFT	100.0	100.0	85.0	95.0	94.8
DCT	100.0	100.0	85.0	90.0	93.5
DWT	85.0	90.0	90.0	95.0	89.9

coefficients. In view of this, to reduce the amount of information, only the first 100 coefficients are used. This provides a considerable dimensionality reduction (from 2500 time-samples to 100 coefficients).

For the wavelet analysis, after exhaustively testing of different combinations of decomposition levels and wavelet functions (e.g. Daubechies-3 and 4 and Coiflet-3), it is used the third decomposition level (316 coefficients) and the Daubechies-3 (db3) mother-wavelet [22]. The approximation coefficients are preferred as they represent compact and smooth version of the original signal.

During the MLP classifier design procedure an important aspect is to determine the number of neurons in the hidden layer. In this work, this is accomplished by testing different numbers of neurons and evaluating the obtained EP. As it can be observed in Fig. 6, the highest discrimination performance is achieved for 12 hidden neurons.

The classification efficiency obtained by feeding a 12 hidden neurons MLP classifier from different features is shown in Table 1. It can be observed that the discrimination performance considerably increases when it is compared to the classifier which uses as input the A-scan information. By using DFT it is achieved EP = 94.8%. In this case, the signatures from no-defect and lack of penetration are identified without errors. A disadvantage of this classifier is that ~7% of the defects signatures are incorrectly classified as having no-defect, which represents a problem for the application. For DCT and DWT, the average efficiencies are 93.5% and 89.9%, respectively.

In order to reduce the number of input features, PCA is applied using different levels of data compression. It can be observed from Fig. 7 that, for DFT and DCT, the highest discrimination performance is achieved while using only 20 principal components (instead of the original 100 coefficients). Alternatively, for DWT, 79 components (from 316 coefficients) are used.

The classification results obtained by using PCA are summarized in Table 2. It can be observed that the average discrimination efficiencies are improved for DFT (from EP = 94.8% to EP = 97.5%) and DCT (from EP = 93.5% to EP = 96.2%). Considering the DWT, the EP suffers a tiny reduction, from 89.9% to 89.5%.

4.2. Results - methodology 2

The second classification methodology comprises two cascaded classification stages. The first one is used to separate no-defect from defect signatures (in this case, all three types of defects are considered as a single class). The second stage is used to identify the type of defect. The same algorithms applied in methodology 1 for feature extraction (DFT, DCT and DWT) were used. For feature selection are applied both, PCA and WMW test.

Table 2
Discrimination efficiencies (in %) for the classes of interest, and the efficiencies product (EP), considering different input features after PCA for the classification methodology 1.

Feat. ext.	ND	LP	SI	PO	EP
DFT + PCA	95.0	100.0	95.0	100.0	97.5
DCT + PCA	95.0	100.0	95.0	95.0	96.2
DWT + PCA	95.0	90.0	80.0	94.0	89.5

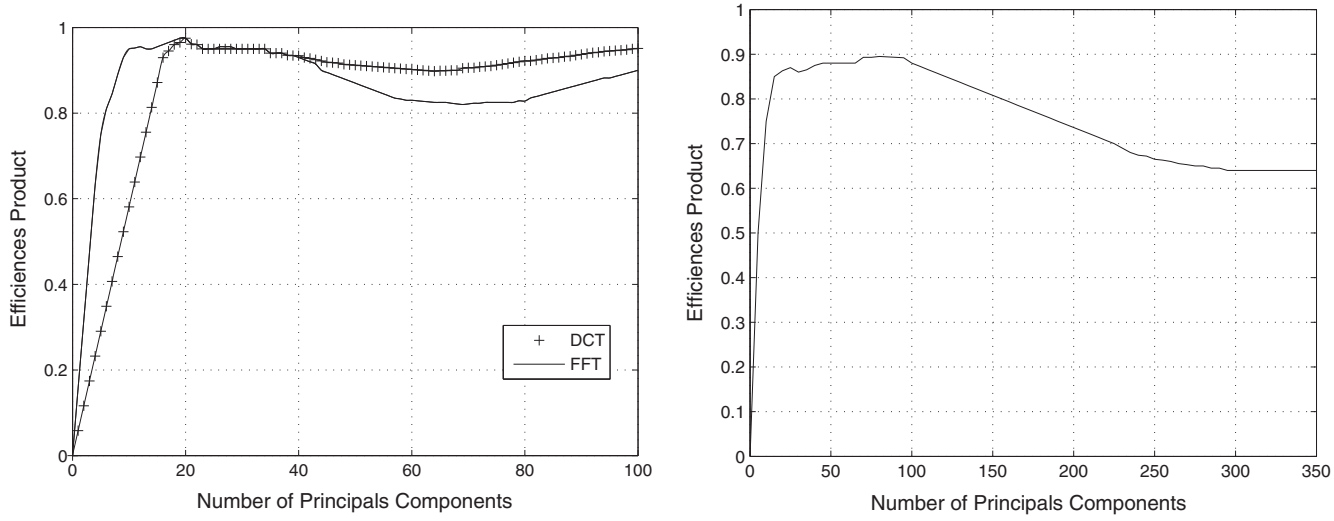

Fig. 7. Efficiencies product versus the number of retained principal components.

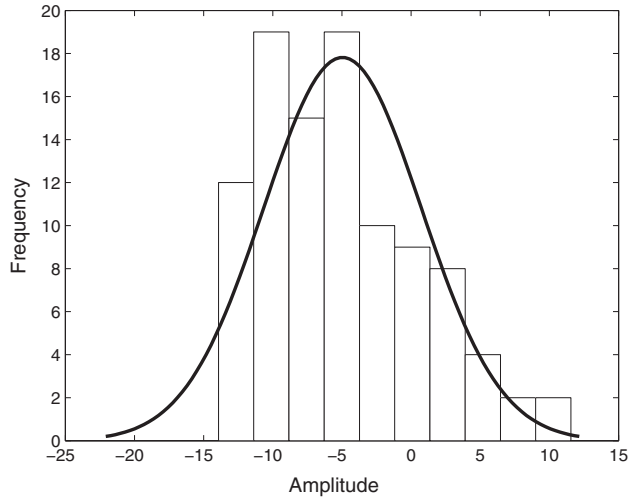

Fig. 8. Histogram of a non-Gaussian feature obtained after PCA preprocessing of DFT (level of significance for a Gaussian fit: $p = 1.2 \times 10^{-4}$%).

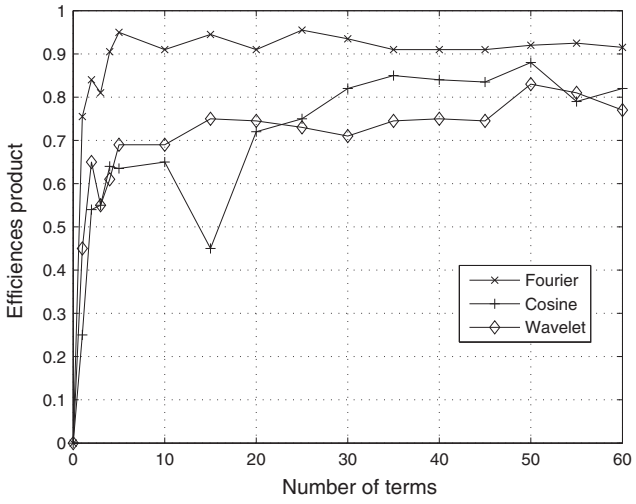

Fig. 9. Efficiencies product for the first classification stage (methodology two) considering different number of features selected by the WMW test.

For our problem most of the features did not passed the chi-square test for Gaussian distributions (their significance levels p are usually bellow 1%), see Fig. 8, for example, and in view of this the t-test could not be applied.

Considering this, the WMW test is used in this work as an alternative for PCA to select the most relevant features. Initially, the features are ranked according to their significance, and the choice of the number of features used to feed the classifier is done through an iterative procedure. It can be observed from Fig. 9, that for DFT the highest discrimination performance is achieved using 25 input features (EP = 95.5%). It is important to note that a quite similar result is obtained for only five features (EP = 95.0%), contributing to considerably reduce the computational complexity. For both, DCT and DWT, 50 features are required.

The first classification stage is trained using as inputs the features selected from both WMW and PCA. It can be observed in Fig. 10, from the ROC curves, that the feature selection with PCA produced higher discrimination performance.

Fig. 11 shows the discrimination performance as a function of the number of hidden neurons for the second classification stage. Table 3 presents a summary of the obtained results. It can be observed that the DWT achieves the highest efficiency for the first classifier and DFT for the second one. It is worth noting that DFT features presents high discrimination efficiency for both classifiers and thus the combined results are better in this case.

4.3. Analysis of the obtained results

Considering the results obtained from methodology 1 (see Table 4), it can be noted that by feeding the classifier from the A-scan (raw time-domain) signals a poor discrimination efficiency is achieved. This classifier has also the disadvantage of requiring considerable greater computational power to deal with 2500 coefficients. Considering the classifiers using preprocessed information, DFT achieves higher discrimination performance. The benefits of using PCA for feature selection are also evident.

For methodology 2 (see Table 5) it can be observed that the DFT features also produce higher discrimination performance. It is interesting to point out that both feature selection algorithms (PCA and WMW test) are able to combine information compaction and discrimination efficiency. An interesting result obtained from methodology 2 is that, using the first classification stage it is possible to detect the signatures from defects using only 5 DFT features selected after the WMW test.

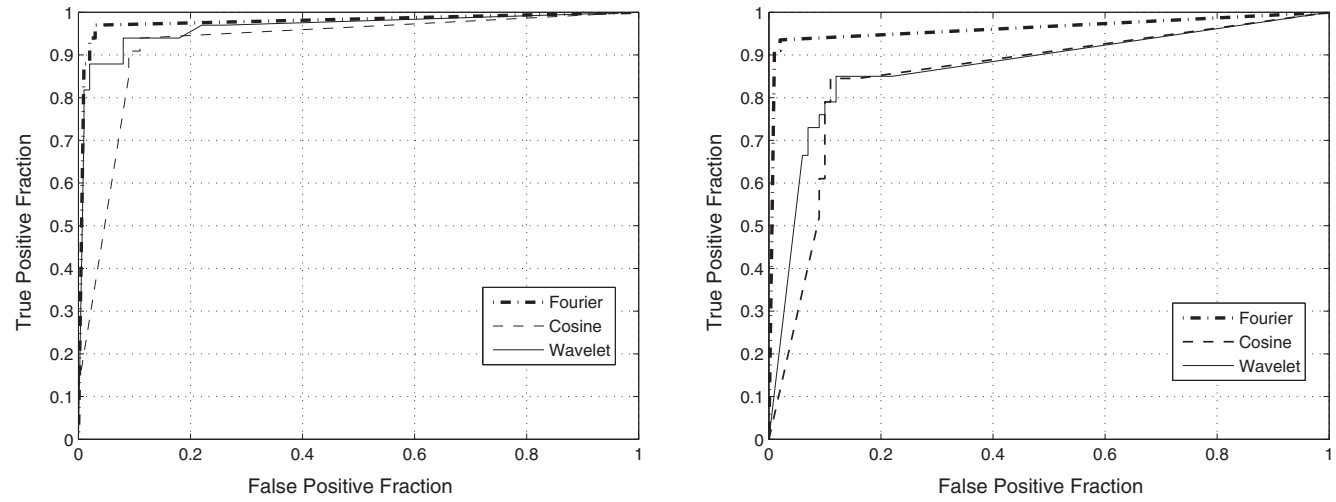

Fig. 10. ROC curves for the classifier 1 (methodology 2) using as feature selection method PCA (left) and WMW test (right).

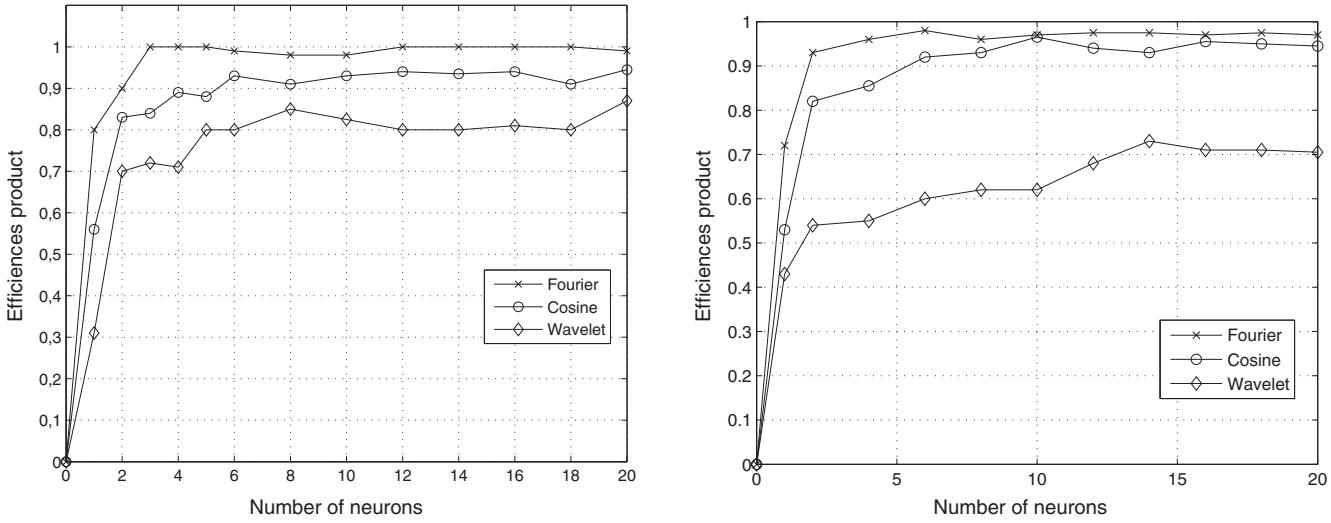

Fig. 11. EP as a function of the number of hidden neurons for the second classification stage, methodology 2, fed from PCA (left) and WMW test (right).

Table 3
EP (in %) for classifiers (CLF) 1 and 2 and for the combination of both classifiers (CLF 1 & 2) for different feature extraction methods.

Feat. ext.	CLF1	CLF2	CLF 1 & 2
DFT + PCA	93.4	100.0	96.9
DCT + PCA	90.9	94.9	85.9
DWT + PCA	95.5	84.9	84.8
DFT + WMW	95.9	97.9	94.5
DCT + WMW	86.9	95.9	85.0
DWT + WMW	82.2	73.8	68.9

Table 4
Summary of the results from methodology I considering: the used feature extraction chain, the discrimination efficiency (EP %), the number of required features (NFeat.) and the probability of error in the classification of a defect (D.E.).

Feat. ext.	EP (%)	NFeat.	D.E. (%)
A-scan	53.1	2500	18.9
DFT	94.8	100	6.7
DFT + PCA	97.5	20	1.7
DCT	93.5	100	8.3
DCT + PCA	96.2	20	3.3
DWT	85.0	316	3.3
DWT + PCA	73.9	79	3.3

Table 5
Summary of the results from methodology II considering: the used feature extraction chain, the discrimination efficiency (EP %), the number of required features (NFeat.) and the probability of error in the classification of a defect (D.E.).

Feat. ext.	EP (%)	NFeat.	D.E. (%)
DFT + PCA	96.9	20	2.9
DFT + WMW	95.4	25	2.0
DCT + PCA	85.9	20	8.9
DCT + WMW	85.0	50	10.9
DWT + PCA	84.8	79	3.0
DWT + WMW	68.9	50	6.0

An important aspect to be observed in NDE problems is the classification error for signatures of defects, as it represents that a potential problematic situation is not detected. It can be observed that this aspect was reduced for DFT + PCA in methodology I (1.7%) and for DFT + WMW in methodology II (2.0%).

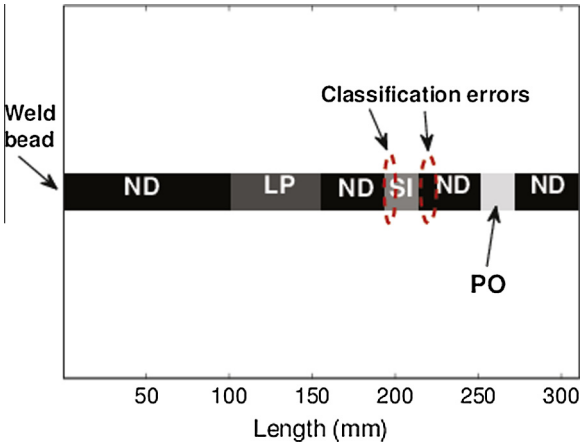

Fig. 12. Identification of the defect regions in the weld bead.

Using the DFT + PCA/methodology I classifier it is possible to identify the location and the size of the defects in the weld bead with high accuracy. As indicated in Fig. 12, the classification errors occur only in two thin regions (in SI and ND areas).

5. Conclusions

Ultrasound testing has been widely applied in different problems. A limitation of this method is that its accuracy relies on the operator previous experience. This work proposes an automatic pattern recognition system which provides decision support information for the operator in the detection of flaws in steel welded joints. Different feature extraction and classification methods are applied to achieve a high detection efficiency. Two different classification methodologies based on MLP neural networks are proposed and the most important input features are selected from PCA and WMW test. The obtained results indicate that the proposed method efficiently identifies the regions of occurrence of defects. For type-I classifier only 20 coefficients, from the original 2500 A-scan time-samples, are required for classification. Using the second classification methodology it is possible to achieve high efficiency in the detection of defect signatures by using only 5 input features. As the required memory and computational complexity may be reduced, a version of the proposed system

may be embedded in a dedicated digital platform to provide on-line information support to the operator.

Acknowledgements

The authors would like to thank FAPESB – Brazil for the financial support.

References

[1] R.W. Messler Jr., Joining of Materials and Structures: From Pragmatic Process to Enabling Technology, first ed., Elsevier, Burlington, USA, 2004.
[2] J.D.N. Cheeke, Fundamentals and Applications of Ultrasonic Waves, second ed., CRC Press, Boca Raton, USA, 2002.
[3] P.S.R. Diniz, E.A.B. da Silva, S.L. Netto, Digital Signal Processing: System Analysis and Design, Cambridge University Press, Cambridge, UK, 2010.
[4] S.O. Haykin, Neural Networks and Learning Machines, third ed., Prentice Hall, New Jersey, USA, 2008.
[5] A. Masnata, M. Sunseri, Neural network classification of flaws detected by ultrasonic means, NDT & E Int. 29 (2) (1996) 87–93.
[6] S. Sambath, P. Nagaraj, N. Selvakumar, Automatic defect classification in ultrasonic NDT using artificial intelligence, J. Nondestruct. Eval. 30 (1) (2011) 20–28.
[7] J.L.B.C. Veiga, A.A. de Carvalho, I.C. da Silva, J.M.A. Rebello, The use of artificial neural network in the classification of pulse-echo and TOFD ultra-sonic signals, J. Br. Soc. Mech. Sci. Eng. 27 (4) (2005) 1–16.
[8] S. Seyedtabaii, Performance evaluation of neural network based pulse-echo weld defect classifiers, Meas. Sci. Rev. 12 (5) (2012) 168–174.
[9] P. Rizzo, I. Bartoli, A. Marzani, F.L. di Scalea, Defect classification in pipes by neural networks using multiple guided ultrasonic wave features extracted after wavelet processing, J. Press. Vessel Technol. 127 (3) (2005) 294–303.
[10] E.F. Simas Filho, A.C.C. Lima, L.A.L. Almeida, Vibration monitoring of on-load tap changers using a genetic algorithm, in: Proceedings of the IEEE Instrumentation and Measurement Technology Conference, Ottawa, Canada, 2005, pp. 2288–2293.
[11] E.F. Simas Filho, J.M. Seixas, L.P. Caloba, Modified post-nonlinear ICA model for online neural discrimination, Neurocomputing 73 (16–18) (2010) 2820–2828.
[12] R. Chai, S.H. Ling, G. Hunter, Y. Tran, H. Nguyen, Brain computer interface classifier for wheelchair commands using neural network with fuzzy particle swarm optimization, IEEE J. Biomed. Health Inform. 18 (5) (2014) 1614–1624.
[13] O. Garcia-Pineda, I. MacDonald, X. Li, C. Jackson, W. Pichel, Oil spill mapping and measurement in the Gulf of Mexico with textural classifier neural network algorithm (TCNNA), IEEE J. Sel. Top. Appl. Earth Obser. Remote Sens. 6 (6) (2013) 2517–2525.
[14] D. Bibicu, L. Moraru, Cardiac cycle phase estimation in 2-D echocardiographic images using an artificial neural network, IEEE Trans. Biomed. Eng. 60 (5) (2013) 1273–1279.
[15] I. Guyon et al., Feature Extraction: Foundations and Applications, first ed., Springer, New York, USA, 2006.
[16] E.F. Simas Filho, J.M. Seixas, Unsupervised statistical learning applied to experimental high-energy physics and related areas, Int. J. Modern Phys. C 27 (1630002) (2016) 1–16.
[17] I.T. Jolliffe, Principal Component Analysis, second ed., Springer, New York, USA, 2002.
[18] E.F. Simas Filho et al., Decision support system for ultrasound inspection of fiber metal laminates using statistical signal processing and neural networks, Ultrasonics 53 (6) (2013) 1104–1111.
[19] C. Cerrillo et al., New contributions to granite characterization by ultrasonic testing, Ultrasonics 54 (1) (2014) 156–167.
[20] J.-J. Ding et al., Two-dimensional orthogonal DCT expansion in trapezoid and triangular blocks and modified JPEG image compression, IEEE Trans. Image Process. 22 (9) (2013) 3664–3675.
[21] J.-Y. Zhu, Z.-Y. Wang, R. Zhong, S.-M. Qu, Dictionary based surveillance image compression, J. Vis. Commun. Image Rep. 31 (2015) 225–230.
[22] S. Mallat, A Wavelet Tour of Signal Processing: The Sparse Way, third ed., Academic Press, Burlington, USA, 2008.
[23] M. Pal, G. Foody, Feature selection for classification of hyperspectral data by SVM, IEEE Trans. Geosci. Remote Sens. 48 (5) (2010) 2297–2307.
[24] M. Kesharaju, R. Nagarajah, Feature selection for neural network based defect classification of ceramic components using high frequency ultrasound, Ultrasonics 62 (2015) 271–277.
[25] A. Hyvarinen, J. Karhunen, E. Oja, Independent Component Analysis, Wiley, New York, USA, 2001.
[26] D. Poole, Linear Algebra: A Modern Introduction, Cengage Learning, Stamford, USA, 2010.
[27] R.L. Ott, M.T. Longnecker, An Introduction to Statistical Methods and Data Analysis, seventh ed., Brooks Cole, Belmont-CA, USA, 2015.
[28] B.M. Wilamowski, H. Yu, Improved computation for Levenberg-Marquardt training, IEEE Trans. Neural Netw. 21 (6) (2010) 930–937.
[29] C. Gigliarano, S. Figini, P. Muliere, Making classifier performance comparisons when ROC curves intersect, Comput. Stat. Data Anal. 77 (2014) 300–312.
[30] J.A. Saez, J. Luengo, F. Herrera, Evaluating the classifier behavior with noisy data considering performance and robustness: the equalized loss of accuracy measure, Neurocomputing 176 (2016) 26–35.
[31] H.L.V. Trees, Detection, Estimation, and Modulation Theory, Part I, Wiley, New York, USA, 2001.