# PATTERNS NONLINEAR CLASSIFIERS OF WELD DEFECTS IN INDUSTRIAL RADIOGRAPHIES

**Romeu R. da Silva**
Programa de Engenharia Metalúrgica e de Materiais/COPPE/UFRJ
Laboratório de Ensaios Não Destrutivos, Cidade universitária, CT
Bloco I Sala I-243 CEP 21945-970 Rio de Janeiro Brasil
E-mail:romeu@metalmat.ufrj.br
http://www.metalmat.ufrj.br

**Professor Luiz P. Calôba**
Programa de Engenharia Elétrica/COPPE/UFRJ
Laboratório de Processamento de Sinais, Cidade Universitária, CT
Bloco H Sala H-220 CEP 21945-970 Rio de Janeiro
E-mail:caloba@lps.ufrj.br
**http://www.lps.ufrj.br**

**Marcio H. S. Siqueira**
Programa de Engenharia Metalúrgica e de Materiais/COPPE/UFRJ
Laboratório de Ensaios Não Destrutivos, Cidade universitária, CT
Bloco I Sala I-243 CEP 21945-970 Rio de Janeiro Brasil
E-mail:marcioh@metalmat.ufrj.br
http://www.metalmat.ufrj.br

**Professor João M. A. Rebello**
Programa de Engenharia Metalúrgica e de Materiais/COPPE/UFRJ
Laboratório de Ensaios Não Destrutivos, Cidade universitária, CT
Bloco I Sala I-243 CEP 21945-970 Rio de Janeiro Brasil
E-mail:jmarcos@metalmat.ufrj.br
http://www.metalmat.ufrj.br

*Abstract. The progresses in researches for the development of an automatic system to analyze weld defects in radiographic images have been evident in the last years. Such research refers to a detailed study of nonlinear classifiers of patterns implemented by artificial neural networks, in order to classify existing weld defects in radiographed weld joints. To increase the reliability of the obtained results, radiographic patterns of the IIW (International Institute of Welding) were used. Geometrical features of the defect classes are used as inputs of the classifiers. The relevance of such features to classify the studied classes was also evaluated. Techniques of analysis of the principal components of discrimination, which were also developed by neural networks, are described in order to visualize in two dimensions the classification problem, as well as the obtained classification performance with the obtained components. The results proved the efficiency of such technique for the data used in the study.*

*Keywords. Radiographic Patterns, Welding Defects, Criterion of Relevance, Nonlinear Classifier, Artificial Neural Networks.*

## 1 – INTRODUCTION

In our modern and globalized world, more and more industries, such as naval, nuclear, chemical, automobile, aerospace, metallurgical, petrochemical and petroliferous industries, among others, have been turning their attention to the quality of their materials and equipments, a basic requisite to keep competitive in national and international markets. Based on this principle, nondestructive tests have a fundamental role in evaluating the quality of the materials and equipment, either in the manufacturing process or during the operation period.

Among nondestructive tests, the radiographic inspection is one the oldest methods and it has been used for about five decades. Among the advantages obtained with such method, if compared to others, as the ultrasonic test for instance, is the fact that the image projected on the film shows the internal *photograph* of the piece, which no other method can provide. The radiographic test using X-rays or gamma source of welded joint or cast pieces is among the most important applications of such inspection method. In this study, we will especially deal with radiographic images of welded joints that are used to inspect the existence of weld defects caused by several reasons.

In this test, the conventional inspection method requires an excellent radiographic image quality and, for this reason, it is controlled by strict international standards. Besides, it is known that the visual inspection of the radiography done by the inspector is performed subjectively, which requires a large experience, visual acuity and knowledge of the technique used in the inspection. However, even when done properly, many times errors occur concerning the interpretation of defects existing in weld joints, either in an existing defect that was not detected or in the incorrect classification of the detected defect.

Nowadays, there are several research centers in the world directed towards the development of an automatic system of inspection and analysis of radiographic images, using computerized methods, especially aiming at the defect classification. Along the period of 15 to 20 years of researches, many publications have been released about this subject, due to its undeniable relevance and importance to this technological development area. Some of these publications will be discussed along the presentation of this study, in some cases, comparing them to this proposed study.

The automatic analysis system applied to radiographic images of welded joint or cast pieces involves basically the following steps: film scanning, or direct digital acquisition without using the film, image preprocessing, defect identification in the weld joint and classification of defects that were found; sometimes, with an extra step of post-processing the results[1-3]. Nowadays, the step corresponding to the classification of patterns has been one of the most studied ones, in terms of research [4-11].

In this study, we describes one research that was conducted in order to inspect the efficiency of nonlinear pattern classifiers of defects implemented by using artificial neural networks. The features of weld defects are evaluated regarding the relevance to the classification of the following defects: undercutting, lack of penetration, porosity and slag inclusion, the latter subdivided into liner and nonlinear inclusion. To do so, neural relevance criteria are used aiming at reducing the dimensionality of the data, which are necessary to the classification. Besides, with the intention of studying this type of classifier deeper, the technique of PCD (Principal Components of Discrimination) was used, also implemented by neural networks, in order to reduce the dimensionality of the input data and bidimensional display of the regions showing the separations among classes. The results are presented in percent of accuracy in the classification of each class. The probability calculation of correct classification using these types of classifiers is presented as well. Considering that such results are a sequence to previously conducted and published studies [12-14], some summaries of previous results are presented along this text to make it easy to understand the new results.

## 2 - EXPERIMENTAL PROCEDURE

### 2.1 - *Radiographic Films*

The radiographic films used in this study follow the IIW[1] (International Institute of Welding) patterns, which attest the existence of a certain defect class in each film. These patterns were used in such a way to ensure the reliability of the results obtained, because they are well accepted by the professionals who work with radiographic tests. The IIW patterns are radiographic images of welded joints of carbon steel, and each film receives the information concerning the radiographic parameters used in the radiographic inspection (equipment voltage, current used in the study, type of existing weld, focus-film distance, etc), but these factors were not taken into consideration because they are not part of the proposed study scope.

### 2.2 - *Film Scanning*

Industrial radiographic film scanning is allowed, according to the standard ASME V[15], since some requirements included in the standard are observed. The most frequently used method to scan radiographic images is the utilization of specific scanners, which work with a light transmission mode.

In his study, Liao[6,7] relates film scanning using the NDT SCAN scanner. Aoki[4] also used one scanner for the radiographic images in his study.

In this present study, one UMAX scanner was used, model: Mirage II (maximum optical density: 3.3; maximum resolution for films: 2000dpi) to scan the IIW films. The spatial resolution used in the study was 500 dpi (dots per inch), totaling an average image size (due to a small size variation in the patterns) of 2900 pixels (horizontal length) × 950 pixels (vertical length), which resulted in an average pixel size of 50μm. Such resolution was adopted for its possibility of detecting and measuring defects of hundredths of millimeters, which, in practical terms of radiographic inspection, is much higher than the usual cases. The adopted gray scale resolution was 8 bits (256 levels), because this scanner model does not allow a better resolution. However, it is known that visually, the human being is not able to recognize any difference in gray scale over 128 levels [16] and images of 10, 12 or 16 bits occupy a significant memory space.

---

[1] Properly authorized by the IIW.

### 2.3 - *Image Preprocessing*

The radiographic films usually have the presence of noises and deficient contrast, due to intrinsic factors of the inspection technique, such as the non-uniform illumination [3]. The noises in scanned radiographic images are usually characterized as randomly spread pixels, with intensity values that are different from their neighboring pixels [2]. The use of low-pass filters to soften the noises is very common, as well as techniques to extend the gray scale histogram to optimize the contrast [2,4].

After being scanned, the radiographic patterns were preprocessed using the Image Pro Plus 4.0 (Media Cybernetics) program. One median-type filter was used to soften randomly existing noise and extend the gray scale histogram to improve the image contrast. The median filter is usually the most indicated to eliminate noises in radiographic images, according to the present literature[2,4]. Further details about the use of such filters with the IIW patterns can be found in Silva[17].

### 2.4 - *Features Extraction*

One of the most used methods to pattern recognition of weld defects is the one that uses the selection of shape and position features of such defects in the weld joint. Kato [5] extracted 10 features in his study to develop one automatic inspection system for radiographic images of joints. The study that Aoki [4] proposed was to deal with 10 features of weld defects.

In this study, we use four features to build the input data set of nonlinear pattern classifier. Such features are briefly described below (a detailed description can be found in the previous study [12]):

- Position (P=h/H): ratio of the defect distance to the weld joint center to the joint thickness at the defect occurrence point. H aims at normalizing the thickness variation of the joint, which occurs frequently in weld radiographic images.
- Aspect Ratio (a=L/e): it is one of the most known features for the identification of "objects" in pattern recognition. Ratio of the highest axis to the smallest axis of the smallest ellipse of an area which is equivalent to the defect area.
- Ratio e/A: ratio of the smallest axis to the defect area.
- Roundness ($p^2/4\pi r$): ratio that measures the resemblance degree between the defect shape and a circumference.

With these features, it was possible to build the input data set of the neural network (input vector). It had a total of 14 samples of undercutting (UC), 15 samples of lack of penetration (LP), 17 samples of porosity (PO) and 49 samples of slag inclusion (SI), divided into 24 samples of nonlinear slag inclusion (NLSI) and 25 samples of linear slag inclusion (LSI). As the data amount differed from one class to another, some data from the less favored classes were randomly duplicated, until 25 samples were obtained, the greatest available amount of a class. Considering the slag inclusion (SI) class as just a single class, 50 samples were handled. It is necessary to point out that such amount of samples was obtained after preprocessing, in order to eliminate outliners that were more than three standard deviations distant from the average in a gaussian distribution, not to make the network training difficult[12].

### 2.5 - *Nonlinear Pattern Classifiers Using Neural Networks*

The step of weld defect classification using artificial intelligence tools is one of the most updated research lines, which is the subject of several publications [4,5]. Most publications talk about the use of Artificial Neural Networks and Fuzzy Logic [18]. In this study, the neural networks [19,20] are used to implement nonlinear pattern classifiers of weld defects, giving a sequence to the studies presented in Silva [12,13,14]. Having a training set, defect patterns extracted from radiographic images of the IIW, networks with supervised learning were used, being the error retropropagation algorithm [19] used in the neural network training. Nonlinear classifiers were implemented using the neural networks with two layers: one intermediate layer and one output layer. In order to obtain the best number of neurons in the intermediate layer, one variation of neuron numbers in this layer was carried out, following the performance and errors obtained after the training in function of the number of neurons utilized. With this information, it was possible to find the best number of neurons to obtain the best classification performance with the data used in the network learning. The obtained results will be displayed in graphs with the number of neurons versus the classification performance and errors. Since it is a pattern classification network, the number of neurons of the output layer corresponding to the number of studied classes. In this case, 4 neurons in the output layer, considering the slag inclusion as just a single class (SI), and 5 neurons in the output layer, when the slag inclusion was considered as subdivided into two classes: linear slag inclusion (LSI) and nonlinear slag inclusion (NLSI). All neurons were hyperbolic tangent-type and with the presence of bias [19,20].

The classifier input, vector $\underline{x}$ composed of defect features, was variable, depending on the relevance found through a criterion that was previously described in Silva [13,14] and which will be briefly described as follows. Figure 1 shows the neural network architecture used in the implementation of nonlinear classifier.

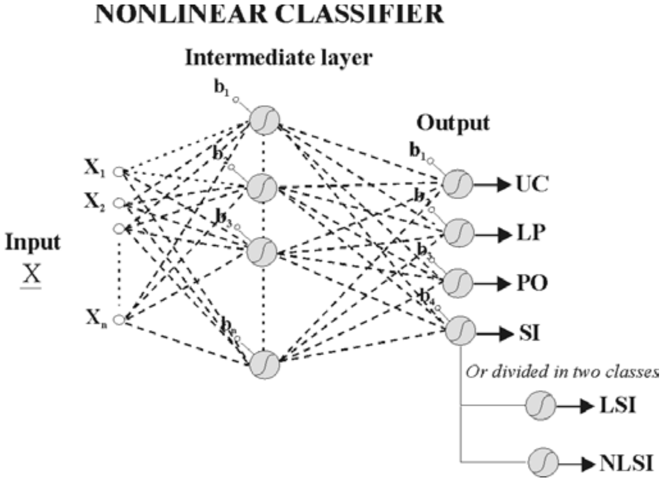



***Figure 1***: Neural network architecture used in the implementation of nonlinear pattern classifier.

**2.6 - *Criterion of Relevance***

The pattern classifier performance depends significantly on the relevance of the features used in the input vector. There are several methods to investigate how relevant one feature is to a class discrimination, all of them aiming at the elimination of irrelevant information in the input vector, which produces a reduction in the system dimensionality and, consequently, in the calculations involved in the process. This study presents the new results obtained through the criterion of relevance described in Silva [13]. This criterion is based on the search for changes in the answer by the network for when one used feature is replaced by its average value. The greater the difference between the answer of the network, the greater the feature relevance [21]. This criterion was used in previous studies [13] to evaluate the six features (also known as characteristic parameters) that were initially used in the study, but, in this case, the classifiers were linear, implemented only with one hyperbolic-tangent type neuron. The criterion of relevance is calculated through the following equation [13,21]:

$$R(x_i) = \frac{1}{N} \sum_{j=1}^{N} \left\| \underline{\tilde{y}}\left(\underline{x}_j\right) - \underline{\tilde{y}}\left(\underline{x}_{j,i}\right) \right\|^2 \qquad (1)$$

$R(x_i)$ : Relevance of Component $\underline{x}_i$ of input vectors $\underline{x}$ ;

$N$ : Number of patterns;

$\underline{\tilde{y}}\left(\underline{x}_j\right)$: Output vector of the neural network for each presented input pattern $x_i$ ;

$\underline{x}_{ji}$: Input vector $\underline{x}_j$ , in which the i-th component was replaced by its average value, which was taken considering all the input vectors;

$\underline{\tilde{y}}\left(\underline{x}_{ji}\right)$: Network output for the input $\underline{x}_{ji}$ .

The new results differ from the previous ones, because now the criterion of relevance was applied using nonlinear classifiers. After calculating the relevance parameters for each feature used in the study, the result is showed in a table.

**2.7 - *Principal Components of Nonlinear Discrimination***

The technique of principal component analysis (PCA) is well known as being one of the most used techniques to reduce the dimensionality of a multivariable data. Originally [19], it is a multidimensional data linear mapping technique in lower dimensions, minimizing the loss of information. However, as it is a mapping linear method, it becomes improper for several engineering problems that are nonlinear, because the smaller components may contain important information that, consequently, shall not be discarded [22]. For this reason, the principal nonlinear component analysis is utilized [22].

One of the ways to develop the principal components is by using the artificial neural networks, as described by Haykin [19] in the self-representation case. In this case, they can be used either to reduce the representation components or discrimination components. In this study, the principal components are used in nonlinear classification, implemented with the neural networks, being developed with the error retropropagation algorithm. Three types of principal component development are evaluated: the components with independent performance and the components with cooperative performance, the latter divided into two types.

As the initial input vector is composed of four components, four defect features, the two nonlinear and independent principal components of classification were used to display the problem of defect class separation in two dimensions. Besides, the principal components found by the three types are used as inputs of nonlinear classifier to evaluate its performance in a situation of reduced dimension. The developed methodology is described as follows.

Suppose that one pattern classification system has a vector $\underline{x}$ of dimension *n* as input, then it is desired to reduce the input vector dimension to a vector $\underline{z}$ of dimension *m*, *m* < *n*, containing only information that is relevant to the data set, so the representation is as follows:

$$\underline{x}\,(\dim n) \rightarrow \underline{z}\,(\dim m) \qquad m<n \tag{2}$$

Thus, considering one pattern classification problem, vector $\underline{z}$ is composed of *m* principal components of classification. If these components are hierarchically developed, the first component is a result of the orthogonal projection of vector $\underline{x}$ in direction $\underline{w}_1$ ($|\underline{w}_1|=1$), which represents the principal direction of the separation of the data in question. Schematically:

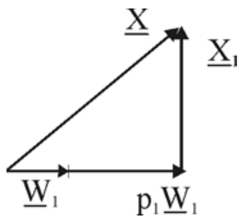



***Figure 2***: Projection $p_1 \underline{w}_1$ of $\underline{x}$ in the principal direction of the discrimination.

Thus, the principal component $p_1$ is the orthogonal projection of $\underline{x}$ in $\underline{w}_1$ and it can be represented as:

$$p_1 = \underline{x}^t \underline{w}_1 \tag{3}$$

Where,

$\underline{x}^t$: Transpose vector of $\underline{x}$;

$\underline{w}_1$: Vector representing the principal direction of the classification.

So, the representation of $\underline{x}$ focusing only one principal component is:

$$\underline{x} \cong p_1 \underline{w}_1 \tag{4}$$

The representation error $\underline{x}_1$ is the following:

$$\underline{x} = p_1 \underline{w}_1 + \underline{x}_1 \tag{5}$$

Vector $\underline{x}_1$ represents non-projected information of $\underline{x}$ in the direction of $\underline{w}_1$. Breaking vector $\underline{x}$ up in *m* components can be represented by:

$$\underline{x} = p_1 \underline{w}_1 + p_2 \underline{w}_2 + p_3 \underline{w}_3 + \cdots\cdots + p_m \underline{w}_m + \underline{\varepsilon}, \tag{6}$$

where $\underline{\varepsilon}$ is the residual error vector [22] for the representation using *m* principal components.

The first nonlinear classification component can be obtained by training a network of three layers in error retropropagation, as shown in Figure 3. The first layer is composed of only one linear-type neuron, while the others are composed of hyperbolic tangent-type neurons.

The adjustment of synaptic vectors of the three layers is done by the descending gradient method using the quadratic average error as objective function [19]. After the training, the vector $\underline{w}_1$ represents the principal direction of nonlinear classification of studied pattern classes.

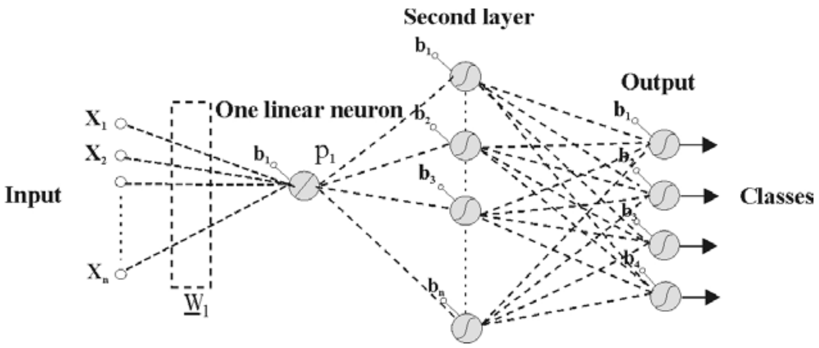



***Figure 3***: Network for the implementation of the first principal component of nonlinear classification.

One special case of development for the remaining components is to make them orthogonal, that is, $\underline{w}_1 \perp \underline{w}_2 \perp \underline{w}_3$ ·and $\perp \underline{w}_m$. For this purpose, the network is trained again after obtaining $\underline{w}_1$ the same way, but the new input to be used will be $\underline{x}_1$, obtained in the equation 5, proceeding successively until the component $m$ (usually $m<n$). Here we call them principal component of independent performance, once each component works with the noise $\underline{x}_j$ of the information that was not used by the previous ones:

$$\underline{x}_j = \underline{x} - \sum_{k=1}^{j-1} p_k \underline{w_k} \quad , \tag{7}$$

Trying to make the classification in an independent way, without counting on the help of the previous components, which had already been extracted, thus now available.

Another way to obtain the components is to consider that the classification is done by utilizing a cooperative performance process, that is, the classification is done by utilizing the component that is being calculated, and all the previously extracted ones. This case can be subdivided into two:

- Type I: after obtaining $\underline{w}_1$, the second component is found by adding a second linear neuron to the first layer. However, $\underline{w}_1$ is made fixed during the training, training only $\underline{w}_2$ and the synaptic weights of the other layers. In this case, the input continues $\underline{x}$, because the components are not independent. The objective in this case is to find the second classification component that is favorable to work in cooperation with the first one. The procedure is similar to obtain the third, fourth or $m$-th component, keeping $\underline{w}_2$, $\underline{w}_3$, and $\underline{w}_{m-1}$ fixed. One way to speed up the process is to use $\underline{x}$ as the input of the first neuron, and $\underline{x}_{j-1}$ as the input of $j$-th linear neuron of the first layer. This is possible because, at the input of each hyperbolic tangent-type neuron of the second layer, the excitation is one linear combination of the network inputs.
- Type II: the two components can also be obtained by training them simultaneously, that is, the two cooperate between themselves during the network training. Similarly, the same can be done for three, four or $m$ components. In this case, one base is created for the reduced and optimized input space for the classification.

In one data set of several dimensions, it is difficult to display the dimension of the separation problem from one class to another, so, by using the two principal components of nonlinear classification, it is possible to have an optimized display of the pattern arrangement. In this study, the two principal components of classification with independent performance were used to display the separation graphs of the four and five classes that were jointly treated. These components were obtained with a neural network, as shown in figure 3, which was trained by using the error retropropagation and the variable learning rate and moment [19,20].

The component $p_1$, as well as the independent set ($p_1 + p_2$), and the two components obtained by the two types of training with cooperative performance were used as input vectors of a nonlinear classifier for the performance evaluation.

## 3 - RESULTS AND DISCUSSIONS

### 3.1 - *Previous Results*

In previous studies [12,13,14], nonhierarchical and hierarchical linear classifiers were evaluated in the aspects of classification performance when six features were used to classify the same classes: UC, LP, PO, NSI and LSI (also the inclusion as just a single class). In this case, it was possible to conclude that the hierarchical classifiers - which classify first the classes of easiest separation [12], enable better performances than those of nonhierarchical classifiers do. The general percent of accuracy of this type of classifier was 85%, when the inclusion class division was taking into account, and 97% without considering such division, which are significantly high percents when talking about linear classifiers.

In Silva [13,14], the results regarding the feature relevance evaluations were published. Two techniques were used: the evaluation by linear correlation between the features and the studied defect classes and the criterion of relevance that was previously described. However, in this previous case, the criterion of relevance was developed using also one linear classifier (implemented with only on neuron). The results showed that the features C (defect contrast [12]) and L/A did not have a significant relevance when compared to the others. For this reason, some combinations were also used among the relevant features to form input vectors for the linear classifiers, which allowed to prove that the reduction in the input data dimensionality is possible, without affecting the pattern classifier performance considerably.

Regarding the previous results, it is also interesting to point out that the criterion of reclassification [12] – according to which the class indicating neuron is the one with the highest internal excitation value, the criterion used for when all classifier outputs are negative or when there are two or more positive outputs-, allowed the best performance.

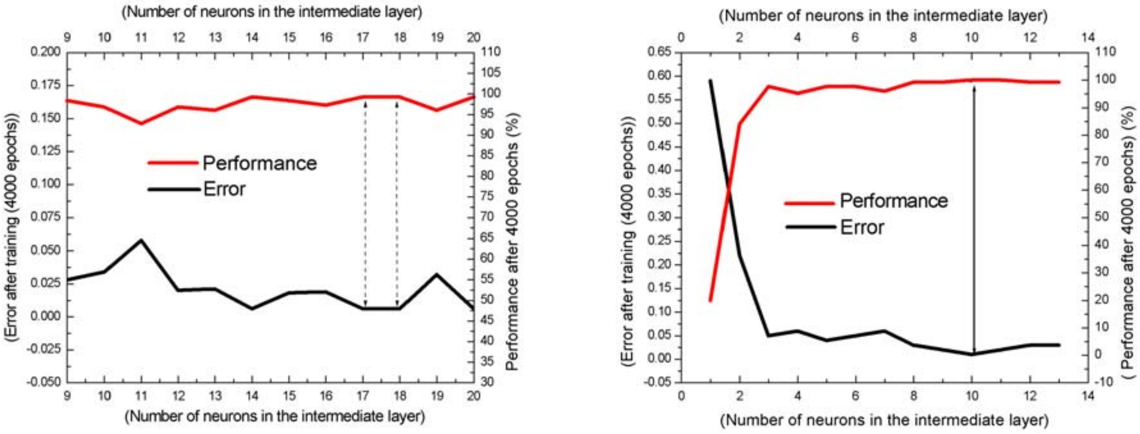
In terms of generalizing the classifiers, as the quantity of observations was small for some classes, the accuracy probability calculations were used within the criterion of reclassification. This methodology is described in details in Silva [13].

### 3.2 - *New Results*

#### 3.2.1 - *Nonlinear Classifiers*

Starting from the final point of the previous studies, the features C and L/A were discarded from the classifier input data, and, thus, the results here presented refer to input composed of those four remaining features (a-e/A-R-P).

In order to find a favorable number of neurons to be used in the intermediate layer of nonlinear classifier (developed as shown in Figure 1), the empirical criterion was used, which gradually increases the number of neurons in this layer and, consecutively, observes the classification error and performance. It was done considering the inclusion class divided into linear and nonlinear inclusion (total of five classes), Figure 4a, and inclusion considered as just a single class (total of four classes), Figure 4b. It was confirmed that, in the first case, 5 classes, the classifier reached the maximum performance percent (99.2%) and the minimum error percent, with the training data of 17/18 neurons in the intermediate layer (Figure 4a). In the four-class case, the maximum performance percent (100.0%) happened for 10 neurons. The confusion tables 1 and 2 show the numerical quantity and accuracy percent obtained for the following conditions: without reclassification criterion and with reclassification criterion for the five-class condition, respectively. It was confirmed that only one nonlinear slag inclusion observation (NLSI) was not classified at first, being afterwards confused with the lack of penetration (LP) class. For the four-class condition, the performance of 100.0% was obtained without the reclassification criterion, therefore, being unnecessary to present the confusion tables.




(a) (b)

***Figure 4***: (a) and (b) Graphs of neuron number optimization in the intermediate layer of the neural network to obtain the best performance with nonlinear classifier with the features: a, e/A, R and P as input. For divided and non-divided inclusion, respectively.

***Table 1***: Table of Confusion obtained with the features a-e/A-R-P.
***Without Reclassification***

| | UC | LP | PO | LSI | NLSI | More than one | None |
|---|---|---|---|---|---|---|---|
| **MO** | 25(100%) | 0 | 0 | 0 | 0 | 0 | 0 |
| **FP** | 0 | 25(100%) | 0 | 0 | 0 | 0 | 0 |
| **PO** | 0 | 0 | 25(100%) | 0 | 0 | 0 | 0 |
| **IEL** | 0 | 0 | 0 | 25(100%) | 0 | 0 | 0 |
| **IENL** | 0 | 0 | 0 | 0 | 24(96%) | 0 | 1(4%) |

***Table 2***: Table of Confusion obtained with the features a-e/A-R-P.
***With Reclassification***

| | UC | LP | PO | LSI | NLSI | Geral Performance |
|---|---|---|---|---|---|---|
| **UC** | 25(100%) | 0 | 0 | 0 | 0 | |
| **LP** | 0 | 25(100%) | 0 | 0 | 4 | |
| **PO** | 0 | 0 | 25(100%) | 4 | 0 | 124(99.2%) |
| **LSI** | 0 | 0 | 0 | 25(100%) | 0 | |
| **NLIS** | 0 | 1 | 0 | 0 | 24(96%) | |

As the quantity of samples was small, and, in order to divides the data set into a test and training group with the intention of evaluating the generalization [19] capacity of the classifier, once again the statistical inference tools were used. The outputs obtained in the classifier for this condition were evaluated according to the test of chi-square and Komolgorov-Smirnov [23] to check if they followed a normal distribution. Then, after being approved in the normal distribution tests, it was possible to calculate the accuracy probability with the reclassification criterion. In this case, as described in details in Silva [13], the probability of having the output of the class in question higher than the outputs of the other classes is calculated, remembering that the output of neurons is a measurement of the probability that a certain observation belongs to one class [13]. Considering that in only one observation, no class was indicated (all negative outputs – Table 1), the numerical difference between the indicated class outputs (positive) and the remaining classes (negative) resulted in one distribution that is distant from zero for all classes. Thus, the probabilities of accuracy for all classes are approximately 100.0%. One distribution example that was obtained for the UC class is shown in Figure 5 below. The other examples are not presented for being similar. This criterion is described in details in Silva [13].

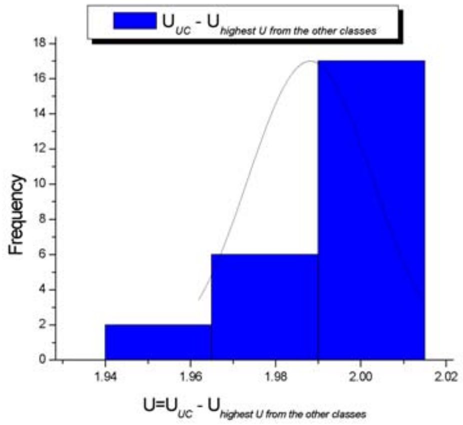



***Figure 5***: Distribution for the difference between the output obtained in the classifier for the UC class and the remaining classes.

**3.2.2 - *Criterion of Relevance***

In this study, a new evaluation of the feature relevance is done, implementing the criterion described with a nonlinear classifier (one double-layer network – Figure 1). In this criterion, it must be pointed out that there is no limit value for a feature to be considered as relevant, unlike the linear correlation criterion [13,14], this analysis is only comparative. Table 3 shows the current relevance values found for the six originally studied features. It can be seen that the feature P shows high relevance with almost all the classes, except for the NLSI, definitively proving to be a very important feature for the defect class classification considered in this study. The features R, a and e/A show a higher relevance in the classification of the PO class, a result that is compatible with those obtained before[13,14], and which can be justified by the fact that they are features of the geometric shape of the defect, and, for the PO class, it is known that the spherical aspect is very common. However, the features C and L/A have a low degree of relevance, if compared to the other features used in this study.

Based on these results, some combinations of inputs using these features were tested as input vectors for nonlinear classifier. Figure 6 shows the results obtained for such combinations of the features. The result obtained with the input vector a-e/A-P, discarding the feature R, was equivalent in the two situations (4 and 5 classes) to the result obtained with the four features, which indicates that it is possible to reduce the input vector dimension to 3, without affecting the classifier performance. The possibility of using only two features was also studied, as the figure shows, and, although the performance was lower if compared to the use of 4 or 3 features, the classification accuracy values are higher than 90% when we use one nonlinear classifier. This result is contrary to those found by Aoki [4]. In his study, in which, by using one input vector with ten features in one nonlinear classifier for the classification of the following classes: UC, LP, PO and SI, the performance was higher than when one or another feature was discarded [4]. Although Aoki [4] worked with features, which are different from those used in this study, we can certify that, in fact, the input vector dimension of the classifiers can be significantly reduced if the irrelevant features are discarded.

Kato et al [5] also used ten features to classify the following defects: crack, lack of fusion, lack of penetration, porosity and slag inclusion. The choice of relevant features to be used in the intelligent pattern classification system followed one criterion that is based on interviews done with radiography inspectors. In such interviews, Kato et al [5] concluded that it is difficult to choose features through this method, because the criterion is very subjective, and each inspector adopts one feature aspect of shape or defect geometry for the categorization.

***Table 3***: Results obtained for the criteria of relevance developed with nonlinear classifier.

| | UC | LP | PO | LSI | NLSI | SI |
|---|---|---|---|---|---|---|
| **C** | 0.23 | 0.16 | 0.11 | 0.14 | 0.14 | 0.21 |
| **a** | 0.10 | 0.24 | 0.58 | 0.07 | 0.15 | 0.36 |
| **L/A** | 0.12 | 0.07 | 0.14 | 0.11 | 0.05 | 0.15 |
| **e/A** | 0.18 | 0.17 | 0.38 | 0.10 | 0.11 | 0.25 |
| **R** | 0.11 | 0.33 | 0.58 | 0.11 | 0.19 | 0.28 |
| **P** | 0.62 | 0.65 | 0.34 | 0.54 | 0.25 | 0.71 |

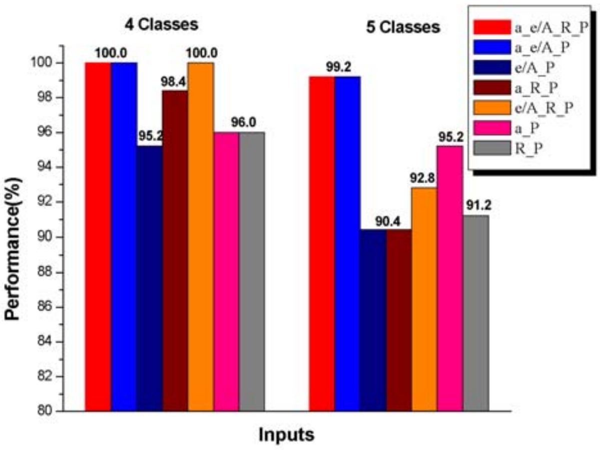
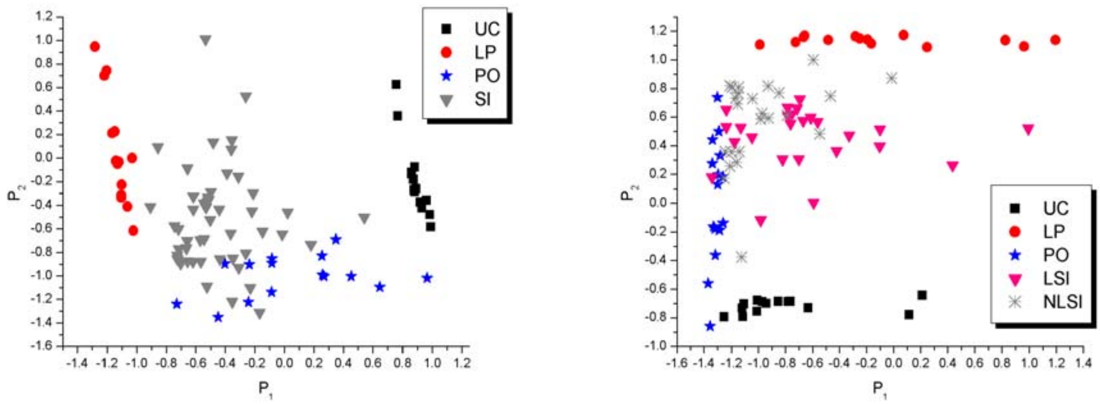



***Figure 6***: Graph with the performances obtained from the indicated combinations of features applied to nonlinear classifiers, for the two situations: 4 and 5 classes. See that the origin of the vertical axis is displaced.

### 3.2.3 - *Principal Components of Nonlinear Classification*

Figure 7 shows the two-bidimensional graphs built with the two principal components of nonlinear classification in independent performance. Figure 7a, which refers to the separation situation for four classes, shows that the classes UC and LP present well-defined regions and, thus, they are easy to be separated by nonlinear separators. Classes PO and SI present one confusion region in the graph, with some observations of SI in the region of PO. It had already been detected when linear classifiers were used, because the classification errors referred to those classes. As nonlinear classifiers were used this time, the possibility of separation of such classes became viable because the neural networks are able to develop very complex separation surfaces and allow a high accuracy performance, as it was proved. However, there is the problem of network generalization [19]. The graphs obtained for the distribution of nonlinear classifier outputs showed that the accuracy probability for not used data during the training is high. But, to make these results even more reliable, it is necessary to use test patterns, which surely will be done in future studies with the acquisition of new patterns. The same happens with Figure 7b, five classes, in which the confusion among NLSI, LSI and PO is even higher. It is easily explained because the distinction between the two classes of SI is complicated for having very similar features. In this case, the lack of generalization becomes more inclined to happen.

The utilization of these display graphs using the principal components is very useful for the systems of higher dimensions, such as the study conducted by Mery [24], that used 71 features to separate the following classes: real defect and structure in radioscopy of aluminum wheels.




(a) (b)

***Figure 7***: (a) and (b): Graphs formed by the two principal components of nonlinear classification of independent performance for the four classes (the inclusion class being considered as just a single class) and five classes (inclusion subdivided into linear and nonlinear inclusion).

The graph in Figure 8 represents the performance results obtained when the first component $p_1$ was used as the classifier input vector, as well as independent set ($p_1 + p_2$) (bidimensional vector) and cooperative ($p_1 + p_2$) (for the two types). The results showed that, using only the first component, the accuracy percent reaches 92.0% for the training data with four classes, and it is obviously lower for a more complex situation of five classes (66.4%). With two components, the percents are close to those results obtained when three or four features were used. Besides, it is also possible to say that there were no significant differences in performance among the types of components used in the study. These results confirmed the efficiency of the principal components to reduce the original data dimension, keeping the same classification accuracy capacity. For the studied case, reducing dimensions 3 or 4 to dimension 2 does not represent a justifiable dimension reduction. However, the results motivate similar studies to be done in a system of higher dimension, as the study conducted by Mery [24], because the quantity of calculations performed by the neural network can be significantly smaller. The same can also be done for when gray scale of weld joints are used, as in the study conducted by Liao [6,7], in which the classifier input vectors may have 500 or more components. These techniques to develop classification components are of recent use, and so far, it has not been found a similar use applied to the researches for the development of an automatic system of defect pattern recognition in radiographic images.

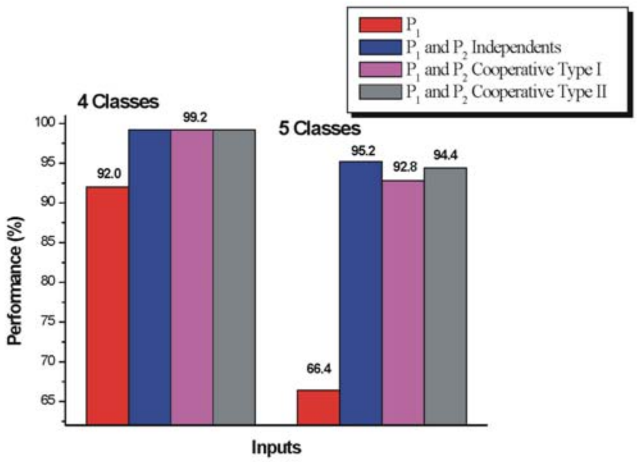



***Figure 8***: Graph showing the performances obtained when principal components of nonlinear classification with independent performance and cooperative performance are used as the classifier input. See that the origin of the vertical axis is displaced.

## 4 – CONCLUSIONS

Nonlinear classifiers provide higher performance results for all defect classes that were studied, when four features of weld defects were used.

The criterion of relevance, used with nonlinear classifiers, definitively confirmed that only four out of the six initial features were relevant to the classification of the defect classes that were studied. The inspection of the performance using only three features shows that the most important is the “quality” of the features used, and not the quantity.

The use of principal components of nonlinear classification to display the defect classes in two dimensions allowed a real dimension of the nonlinear classification problem for the data used in the study. When the components were used as input vectors of the classifiers, they allowed good accuracy results.

It is important to point out that the defect classes crack and lack of fusion, which are important in terms of welded joints, have not been evaluated by this technique yet, due to the insufficient quantity of reliable samples that are available. But surely it will be done in a future study, with the acquisition of new radiographic patterns, as well as the use of test data of pattern classifiers to confirm the generalization.


## 5 – ACKNOWLEDGEMENT

Authors wish to acknowledge CNPq, CAPES and FAPERJ for financial support and scholarships and to International Institute of Welding for the permission given to publish the present work using the radiographic standards.

## 6 – REFERENCES


1. Aoki K, Suga Y. “Intelligent Image Processing for Abstraction and Discrimination of Defect Image in Radiographic Film”. In: Proceeding of the Seventh International Offshore and Polar Engineering Conference. Honolulu. USA. May. 25-30, 1997. p. 527-531.
2. Kehoe A, Parker G A. “Image Processing for Industrial Radiographic Inspection: Image Enhancement”. British Journal of NDT 1990; 32(4):183-190.
3. Cherfa Y, Kabir Y, Drai R. “X-rays Image Segmentation for NDT of Welding Defects”. In: 7th European Conference on Non Destructive Testing. Copenhagen. 26-29 May, 1998. p.2782-2789.
4. Aoki K, Suga Y. “Application of Artificial Neural Network to Discrimination of Defect Type Automatic Radiographic Testing of Welds”. In: ISI International 1999; 39(10): 1081-1087.
5. Kato Y, Okumura T, Matsui S, et al. “Development of an Automatic Weld Defect Identification System for Radiographic Testing”. Welding in the Word 1992; 30 (7/8):182-188.
6. Liao T W, Ni J. “An Automated Radiographic NDT System for Weld Inspection: Part I – Weld Extraction”. NDT&E International 1996; 29(3):157-162.
7. Liao T W, Li Y. “An Automated Radiographic NDT System for Weld Inspection: Part II – Flaw Detection”. NDT&E International 1998. 31(3):183-192.
8. Liao T. W, Li D, Li Y. “Extraction of Welds from Radiographic Images using Fuzzy Classifiers”. Information Sciences 2000; 126: 21-40.
9. Liao T W, Li D, Li Y. “Detection of Welding Flaws from Radiographic Images with Fuzzy Clustering Methods”. Fuzzy sets and Systems 1999; 108:145-158.
10. Liao T W, Tang K. “Automated Extraction of Welds from Digitized Radiographic Images Based on MLP Neural Networks”. Applied Artificial Intelligence 1997; 11: 197-218.
11. Sankaran V, Chartrand B, Millard D, et al. “Automated Inspection of Solder Joints-A neural Network Approach. European Journal Mechanical Engineering; 43(3):129-153.
12. Silva R R, Siqueira M H. S, Calôba L P, et al. “Radiographics Pattern Recognition of Welding Defects using Linear Classifiers”. Insight 2001; 43(10): 669-674.
13. Silva R R, Siqueira M H. S, Calôba L P, et al. “Evaluation of the Revelant Characteristic Parameters of Welding Defects and Probability of Correct Classification using Linear Classifiers”. Insight 2002; 44(10): 616-622.
14. Silva R R, Siqueira M H. S, Calôba L P, et al. “Contribution to the Development of a Radiographic Inspection Automated System”. In: 8th Europen Conference on Non Destructive Testing. june 17-21, 2002.
15. ASME V. Article II. Mandatory Appendix III. Digital Image Aquisition. Display and Storage for Radiography and Radioscopy 1995. p. 36-37.
16. Castleman K R.. Digital Image Processing. New Jersey. Prentice-Hall. 1996.
17. Silva R R. “Radiographic Image Processing and Analysis of Defects in Weld Bead”. Master’s Degree. PEMM/COPPE/UFRJ. RJ, Brazil,1999. (In Portuguese).
18. Lashkia V. Defect detection in X-ray images using fuzzy reasoning. Image and Vision Computing 2001; 19: 261-269.
19. Haykin S. “Neural Networks – A Comprehensive Foundation”. Macmillian College Publishing. Inc. 1994.
20. Bishop C. M. “Neural Networks for Pattern Recognition”. New York. Oxford University Press. 1995.
21. Seixas J M, Calôba L P, Delpino I. “Relevance Criteria for Variance Selection in Classifier Designs. In: International Conference on Engineering Applications of Neural Networks”. 1996. p.451-454.
22. Dong D, McAvoy T J. “Nonlinear Principal Component Analysis-Based on Principal Curves and Neural Networks. Computers&Chemical Engineering”. 1996; 20(1): 65-78.
23. Ang A H S, Tang W H. “Probability Concepts in Engineering Planning and Design”. vol. 1. Basic Principles. John Wiley & Sons. Canada 1975.
24. Mery D. Filbert D. “Classification of Potential Defects in the Automatic Inspection of Aluminium Castings Using Statistical Pattern Recognition”. In: 8th European Conference on Non Destructive Testing. Barcelona 17-21 June. 2002. p. 1-13.