

Automatic classification of weld defects using simulated data and an MLP neural network

T Y Lim, M M Ratnam and M A Khalid

An effective weld defect classification algorithm has been developed using a large database of simulated defects. Twenty-five shape descriptors used for the classification were studied and an optimal set of nine descriptors with highest discriminative capability was selected using a statistical approach. A multi-layer perceptron (MLP) neural network was trained using shape parameters extracted from the simulated images of weld defects. By testing on 60 unknown simulated defects, the optimised set of nine shape descriptors gave the highest classification accuracy of 100%. Defect classification on 49 real defects from digitised radiographs produced maximum overall classification accuracy of 97.96%.

1. Introduction

The radiography technique (RT) used in non-destructive testing (NDT) of welds has evolved rapidly in recent decades and has become an established technology in the field of weld assessment. RT is widely used as an inspection tool for detecting flaws inside welded structures, pressure vessels, structural members and pipelines⁽¹⁾. Most radiographic exposures and film interpretations in RT are still carried out manually⁽²⁾. Human interpretation of weld defects, however, is tedious, subjective and is dependent upon the experience and knowledge of the inspector.

Much effort has been made recently to converge RT into an automated process of radiographic interpretation. With the advancement of digital image processing and computer architecture, the automated interpretation of weld radiographs is made possible with a system consisting of film digitisation, pre-processing, defect detection and classification. The automatic interpretation of radiographs using digital image processing reduces human involvement, thus making the inspection more reliable and faster.

Basically, automatic defect interpretation consists of five stages: image digitisation, pre-processing, weld extraction, defect segmentation and defect classification. A digitised radiographic image is often corrupted by non-uniform illumination, noise and poor contrast⁽³⁾. Due to the poor quality of a radiographic image, image pre-processing is normally carried out as an initial stage of defect detection. This includes noise removal and contrast enhancement. Poor quality radiographic images have led to the development of various automatic defect detection algorithms that focus on extracting defects using various image segmentation methods⁽⁴⁻¹²⁾. In addition, rapid growth of artificial intelligence methods such as artificial neural networks (ANNs) and fuzzy techniques has become an important component to assist and enhance the task of defect detection in poor quality images.

T Y Lim and M M Ratnam are with the School of Mechanical Engineering, Engineering Campus, Universiti Sains Malaysia, 14300 Nibong Tebal, Penang, Malaysia.

M A Khalid is with the Malaysian Institute for Nuclear Technology Research, Bangi, 43000 Kajang, Malaysia.

Corresponding author: M M Ratnam, E-mail: mmaran@eng.usm.my

Weld extraction is usually the first step in the development of automated inspection of weld radiographs. Several techniques of weld extraction are available in the literature, such as weld extraction based on the observation that the intensity plot of a weld profile looks more like Gaussian than the other objects in the image⁽⁴⁾, extraction of features from line image of welds for use in training a multi-layer perceptron (MLP) neural network⁽⁵⁾ and extraction of multiple curved welds from a radiographic image using fuzzy *K*-NN and *C*-means methods⁽⁶⁾.

Problems relating to poor quality images have been addressed by several authors. For instance, Kehoe *et al*⁽⁷⁾ introduced the sigma-norm contrast enhancement and mean gradient edge detector to improve the quality of the image. Bonser and Lawson⁽⁸⁾ developed filtering and 'window' based variance operator for segmentation of suspected defect areas inside the weld region of partially competed welds. Murakami⁽⁹⁾ proposed a local arithmetic operation to a limited region and was followed by thresholding methods. Various image processing techniques were used by other authors to process and segment the weld images⁽¹⁰⁻¹²⁾.

Research in weld defect detection has also been carried out actively in the past. For instance, Lashkia⁽³⁾ proposed a detection algorithm based on fuzzy reasoning to detect low-contrast defects using local image characteristics, such as spatial contrast, spatial variance and distance between two contrast regions. Liao and Li⁽¹³⁾ developed welding flaw detection based on the fitted line profiles of a weld image and successfully detected 93.33% of various defects from linear welds. Jacobsen *et al*⁽¹⁵⁾ extracted parameters from the intensity profile for the network with five techniques, namely morphological filter, derivatives of Gaussian filter, Gaussian Weighted Image Moment Vector-Operator (GWIMV) filter, Fast Fourier Transform (FFT) filter and wavelet transform.

Past research in automatic weld radiography focused mainly on enhancing poor quality weld images and detecting defects from the radiographs, whilst the development of automatic defect classification systems is limited. Some of the research studies on automatic weld defect classification are reviewed briefly.

Jacobsen *et al*⁽¹⁵⁾ used several features to differentiate between crack, undercut and absence of defect. Perner *et al*⁽¹⁶⁾ introduced a framework for distinguishing classification methods, namely neural nets and decision tree. Their system only classified crack, undercut and absence of defects. Feher⁽¹⁷⁾ presented a simple classification of three types of defects, namely porosity, undercut and incomplete penetration by simple rules. However, the system was unable to detect small defects that did not differ much from the background. Aoki and Suga⁽¹⁸⁾ used 10 parameters to characterise the defects and classified them into five possible classes using a MLP neural network. Their investigation on 27 defects showed that 25 defects were classified correctly with a success rate of 92.6%. In their work, 35 training samples and 27 test samples were used. These are, however, too small for practical applications. Wang and Liao⁽¹⁹⁾ used a set of parameters to classify six possible defects and obtained the highest accuracy of 92%. In their work, 108 data sets were used for training while 12 were used for testing. However, the 12 test samples used for classifying six types of defect are considered

small and the success rate for individual defect was not reported. Silva *et al.*⁽²⁰⁾ also used neural network to study the accuracy of defect classification, though only 50 datasets were used for training the network. In another work published by Silva *et al.*⁽²¹⁾, some of the defect classes with less data were duplicated to overcome the lack of sufficient samples for training the neural network. Such duplication of data can easily influence the prediction of the network.

The performance of a weld defect classification system often depends on the set of radiographic images used. Different authors used different sets of characterisation parameters and different sets of radiographic images. Moreover, final appearance of defects extracted depends on the image segmentation method used. Thus, comparison of the effectiveness of various classification techniques is not possible unless it is based on a standard set of 'ideal' weld defects. As stated by Wang and Liao⁽¹⁹⁾, there is a need to establish a benchmark image set in order to perform the comparison between the performances of various classification techniques. Silva *et al.*⁽²¹⁾ also concluded that the lack of high number of samples to increase the reliability of the classification is a common problem in the automatic interpretation of weld radiographs.

In this work, an efficient classification system using a large number of simulated images of weld defects that are considered as 'ideal' defects is developed. A set of shape descriptors was defined and the number of descriptors required in the classification was optimised using a statistical approach. The classification is divided into two parts. Firstly, a multi-layer perceptron (MLP) neural network was trained using features extracted from the simulated defects and the network was used to classify a set of simulated defects. Secondly, classification was carried out using features extracted from real defects while the training was done using the same simulated data set. The classification accuracy of individual defect type was also evaluated.

2. The simulation process

There are two main advantages of using a large database of simulated images to develop the classification task. Firstly, simulated images are created manually by imitating the appearance of real radiographs and, therefore, a large number of images could be created to simulate the variation of shapes in real defects. Secondly, different real radiographs have different intensity variations. Thus, for a given set of images a particular shape feature may give certain accuracy in defect classification while the same feature may give different accuracy in a different set of images due to the intensity variation. Simulated images overcome this shortcoming by providing a basis to test and select the appropriate features for the classification.

Before generating the simulated images, real radiographs were digitised using an X-ray film scanner (Cobrascan CX-612-T) in 12-bit resolution and saved in TIFF image format without compression. The original radiographs were a collection of reference radiographs previously evaluated by qualified inspectors. These radiographs were used as the basis for creating simulated weld defects. The simulation process is divided into four stages: (i) Generation of simulated defect images, (ii) shape definition, (iii) shape parameter optimisation and (iv) classification evaluation.

2.1 Generation of simulated defect images

Six types of defects were considered in creating the simulated defect images. These are longitudinal crack, incomplete penetration, porosity, cavity, slag inclusion and transverse crack. To create defects having high similarity to the real defects, the real image for each defect type was used and was subjected to edge detection process using the Sobel edge detector. The Sobel edge detector uses two convolution kernels, one to detect changes in vertical contrast and another to detect horizontal contrast. Due to noise in the real

image, the resulting edges are irregular and discontinuous. These edges were touched up manually by linking the broken edges and eliminating pixels due to noise. A total of 300 simulated defects, 50 images for each defect type, were generated for training the neural network. Additional 10 images for each defect type were generated for testing the accuracy of classification. Figures 1(a)-(d) show an example of a real image of crack, the edge of the defect, the edge after touching up and the simulated defect image. Figures 2(a)-(f) show samples of simulated defects generated for various types of defects.

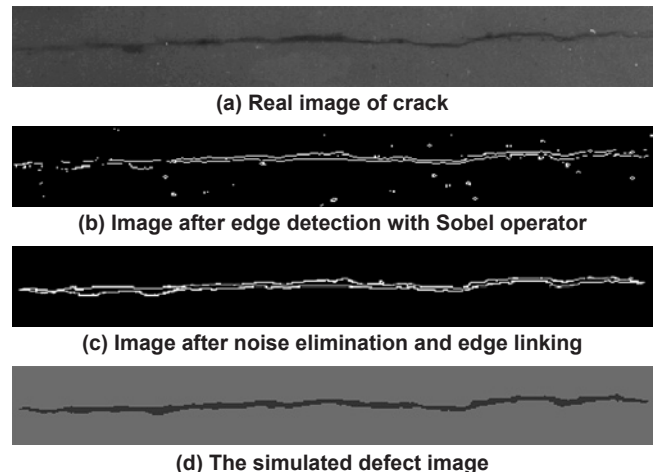


Figure 1. Generation of simulated defects

2.2 Shape definition

Shape recognition and classification are fundamental problems in many applications of machine vision⁽²²⁾, especially in vision-based inspection. In weld defect detection, a radiograph interpreter will inspect the defect shape, geometry and its orientation to recognise the defect. The shape or outline of a defect is particularly important because a certain defect type may have a different shape compared to other defects. Incomplete penetration, for instance, is defined as dark continuous or intermittent line in the middle of the weld. In contrast, porosity is defined as dark shadows of rounded contour.

The main purpose of shape descriptors is to measure geometric attributes of the shapes that can be used to classify and recognise an object. Shapes can be represented using various types of descriptors and techniques. Different types of methods that characterise a shape can be viewed from different context, such as description based on boundaries and region, local and global shape characters, statistical or syntactic object description, object reconstruction ability or incomplete shape recognition ability⁽²³⁾. A summary of common shape description techniques is given in the literature⁽²⁴⁾. A thorough discussion of shape analysis and recognition can be found in the literature⁽²³⁻²⁵⁾.

In this study, 25 common shape descriptors were defined in order to evaluate their capability in discriminating different types of weld defect shapes. Most of the descriptors use area, perimeter, diameter,

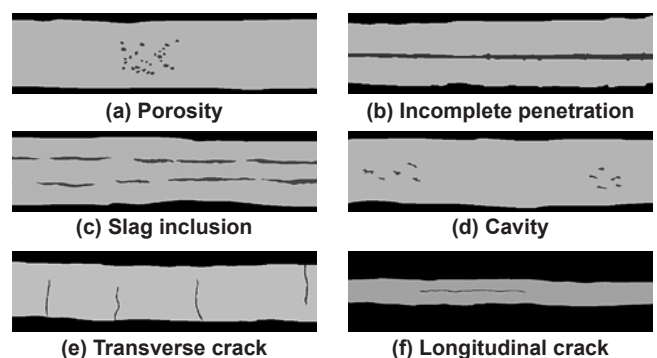


Figure 2. Simulated images for various defect types

width and length in describing shapes that are simple attributes of shape boundary. On the other hand, feature like curvature investigates the overall localised properties of shape boundary⁽²⁶⁾. Other features such as standard deviation, skewness and kurtosis of radius⁽²²⁾ tend to characterise the variability of radius data. A list of the shape descriptors used in this study is given in Table 1.

In order to choose a combination of optimal shape descriptors that provides the highest ability to classify the weld defects, a multiple comparison procedure was carried out and is discussed in the following section.

2.3 Shape parameter optimisation

A multiple comparison procedure was employed to compare the mean differences among each type of defect for every shape descriptor. The statistical test was conducted to determine if there are significant differences between each of the defect groups for a particular shape descriptor. A typical procedure to perform this test is as follows:

- Select the statistical hypothesis:
 $H_0 : \mu_i = \mu_j$ for all classes i and j where $i \neq j$. This hypothesis indicates that all pairs of groups have equal means.
 $H_1 : \mu_i \neq \mu_j$ if at least one pair of means is different.
- Select the significance level α .
- Determine the critical value from F -test. This is the one tailed 'Analysis of Variance' (ANOVA) F -test to find out whether data from several groups have a common mean. Rejection of H_0 leads to the multiple comparison test to determine specific differences among those groups.
- Determine the critical value from the multiple comparison procedure.
- Evaluate the results and draw conclusion.

The purpose of multiple comparison procedure is to determine where mean differences lie among a set of means in a one-way ANOVA because the rejection of H_0 in F -test does not indicate which of the means are different⁽²⁷⁾. Multiple comparison procedures compare two means to determine if they are statistically different. With every type of shape descriptor, six groups of defects were tested to determine if they are significantly different from others. The discriminative criteria of each descriptor is based on the concept that the smaller the number of joining groups in the test (which means most of the means are significantly different), the better the descriptor in classifying different types of defects.

Figure 3 shows an illustration of the multiple comparison procedure where \bar{y}_i is the mean of the individual defect type. The lines underneath indicate the mean values of defects that are insignificant among others, and thus combine the defects into several groups. Each group indicates that there is no significant difference in mean values among the group members. From Figure 3, four groups of defects have different means with an overlapping result for \bar{y}_5 . In the current work, two overlapping results are considered as one joining group. Therefore, the shape descriptor (*mean*) in Figure 3 is considered to have the ability to identify three main groups of defects.

By using this criterion, a comparison among all 25 shape descriptors was carried out and the optimal groups were identified.

Defect Descriptor	Defect					
	Crack	Incomplete Penetration	Slag	Transverse Crack	Porosity	Cavity
Mean	\bar{y}_1	\bar{y}_2	\bar{y}_5	\bar{y}_6	\bar{y}_3	\bar{y}_4
	Group 1 (joining sub-groups)			Group 2	Group 3	

Figure 3. Example of multiple comparison result

Table 1. Shape descriptors

Shape descriptors		
1) Form	8) Convexity	17) Curvature ^[24]
2) Area to perimeter ratio	9) Solidity	18) Diameter ratio
3) Roundness	10) Compactness	19) Curl
4) Aspect ratio 1 (maximum diameter/minimum diameter)	11) Extent	20) Orientation
5) Aspect ratio 2 (minimum diameter/mean diameter)	12) Heywound diameter	21) Standard deviation of radius ^[20]
6) Aspect ratio 3 (maximum diameter/mean diameter)	13) Major to minor axis length	22) Skewness of radius ^[20]
7) Elongation	14) Miu ^[18]	23) Kurtosis of radius ^[21]
	15) Sigma ^[18]	24) Horizontal length to area ratio ^[17]
	16) Circularity ^[18]	25) Vertical length to area ratio ^[17]

There are several multiple comparison procedures that can be adopted depending on how error rates are controlled. In this study five types of comparison procedures were used, namely Tukey HSD, Tukey LSD, Scheffe, Bonferonni and Dun-Sidak⁽²⁷⁻²⁸⁾ and the result of each test was compared among each other. The numerical values of data were normalised linearly before performing the mean comparison. The assumptions made in order to satisfy the multiple comparison procedure (as well as the F -test) are as follows: (a) the samples are randomly and independently selected, (b) the populations are normally distributed, and (c) the populations all have the same variance.

The defects from the simulated images were extracted in order to obtain the numerical data for every shape descriptor. The data were then used to perform the multiple comparison tests. The various comparison procedures were found to produce the same results. The result for one of the comparison procedures is shown in Table 2. Among the 25 shape descriptors, the highest discrimination capability produced by the tests was five groups using three shape descriptors followed by four groups using seven shape descriptors. Figures 4 and 5 show examples of the distribution of the shape descriptors *Solidity* and *Circularity* that have five and two joining groups respectively. From these Figures, it can be concluded that *Solidity* has better ability to classify defects compared to *Circularity*.

2.4 Classification of defects using MLP network

A multi-layer perceptron (MLP) model using back-propagation (BP) algorithm was used for learning and classifying the defects. MLP networks are gaining popularity in classification task due to their flexibility, robustness and high computational rates. The MLP model is shown in Figure 6.

The general model of MLP consists of a number of nodes arranged in multiple layers with connections between the nodes in the adjacent layers by weights. The model consists of an input layer

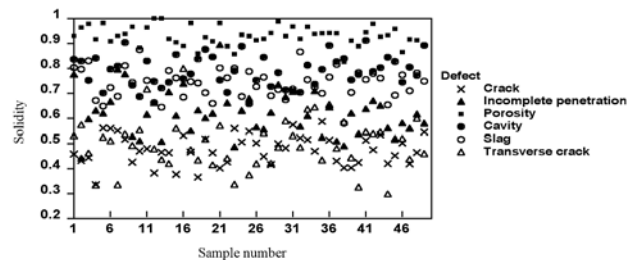


Figure 4. Distribution of *Solidity*

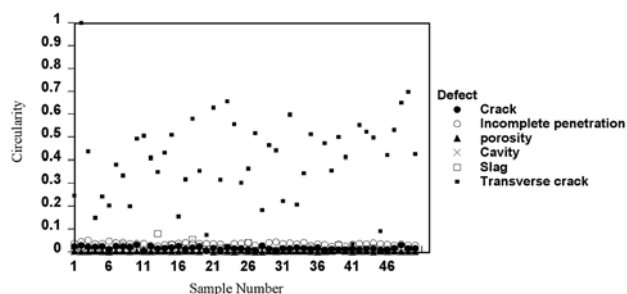


Figure 5. Distribution of *Circularity*

Table 2. Results of multiple comparison procedure

Number of groups	Shape descriptors
5	Solidity, extent, standard deviation of radius distribution.
4	Compactness, Heywood diameter, vertical length to area ratio, major to minor axis length, curvature, orientation.
3	Form, area to perimeter ratio, roundness, aspect ratio 2, aspect ratio 3, elongation, horizontal length to area ratio, eccentricity.
2	Aspect ratio 1, circularity, curl, kurtosis of radius distribution.
1	Convexity, miu, sigma, diameter ratio, skewness of radius distribution.

that accepts the input variables used in the classification, hidden layers and an output layer. A summation of each neuron j in the hidden layer by its input nodes x_i after multiplying the connection weights w_{ij} gives the output y_j as a function of the sum, that is:

$$y_i = f\left(\sum w_{ij}x_i\right) \dots\dots\dots(1)$$

where f is the sigmoidal or hyperbolic tangent transfer function. Using the BP training algorithm, the weights are minimised based on the squared differences between the actual and desired output values in the output neurons given by:

$$E = \frac{1}{2} \sum_j (d_j - y_j)^2 \dots\dots\dots(2)$$

where y_i is the actual output of the neuron and d_j is the desired output of neuron j . During classification, input data is fed into the network and the classification is performed by assigning a class number to a pixel or segment using the numerical values computed at the output layer. The weight w_{ij} is updated with an increment Δw_{ij} and the error E is reduced until an acceptable value of E is reached. The output node that gives the highest value is set to 1 whereas the others are set to 0 in order to obtain the output class vector.

3. Results and discussion

From the statistical analysis, the 25 shape descriptors were divided according to their classification ability (Table 2). To determine the suitable input features for training the MLP network, three sets of analyses were carried out using three features (five groups), nine features (five and four groups), and 17 features (five, four and three groups) as given in Table 2. In each set, 50 simulated images were used for each defect type, giving a total of 300 images for six defect types for training. For testing the classification accuracy 60 images were used (10 images for each defect type). The output layer consists of six nodes corresponding to the six types of defects. The learning rate and error rate were set, respectively, to 1.0 and 0.

The results of comparison of the classification accuracies of the three sets of descriptors (at training epoch 100) are shown in Figure 7. The three features set gave the lowest overall accuracy compared to the other two. For the set containing nine features, the accuracy is generally higher than that for the 17 features set. The comparison between the nine features set and 17 features set shows that increase in the number of features will not

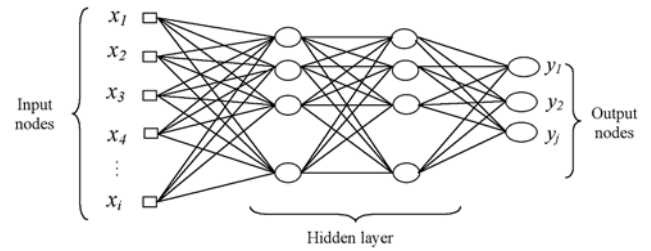


Figure 6. A schematic of a MLP model

necessarily increase the classification accuracy at the same number of hidden nodes. In addition, a larger features set will increase the processing time while resulting in similar results. Thus, in the current research nine features set were selected as the optimised set of input features to train the network and classify the defects. The classification accuracy using simulated defects for various numbers of hidden nodes and epoch numbers for the nine features set is shown in Table 3. The maximum accuracy (100%) was found to occur when 24 hidden nodes were used.

The classification accuracies attainable using the shape descriptors proposed by Aoki and Suga⁽¹⁸⁾ and Wang and Liao⁽¹⁹⁾ were also investigated in the simulation study. Only the shape parameter sets defined by both authors as shown in Table 4 were used, whereas the defect images were the same as those used in the current study.

Figure 8 shows a comparison of the classification accuracy between the three sets of shape descriptors. The optimised set

Table 3. Classification accuracy of simulated defects

Number of nodes	Number of epochs	Accuracy/%			Number of nodes	Number of epochs	Accuracy/%		
		3 features	9 features	17 features			3 features	9 features	17 features
6	20	88.3	98.3	96.7	20	20	90.0	98.3	96.7
	40	90.0	96.7	98.3		40	91.7	98.3	95.0
	60	88.3	96.7	98.3		60	93.3	98.3	96.7
	80	88.3	96.7	98.3		80	93.3	98.3	96.7
	100	88.3	96.7	98.3		100	93.3	98.3	96.7
8	20	88.3	98.3	93.3	22	20	88.3	100.0	96.7
	40	88.3	98.3	93.3		40	91.7	100.0	96.7
	60	88.3	98.3	93.3		60	93.3	98.3	95.0
	80	86.7	98.3	93.3		80	95.0	98.3	95.0
	100	88.3	98.3	93.3		100	93.3	98.3	95.0
10	20	90.0	100.0	96.7	24	20	90.0	100.0	98.3
	40	95.0	98.3	96.7		40	91.7	100.0	96.7
	60	93.3	98.3	96.7		60	93.3	100.0	96.7
	80	93.3	98.3	96.7		80	91.7	100.0	96.7
	100	93.3	98.3	96.7		100	90.0	100.0	95.0
12	20	88.3	100.0	96.7	26	20	91.7	98.3	98.3
	40	93.3	98.3	96.7		40	93.3	98.3	98.3
	60	99.3	98.3	96.7		60	93.3	98.3	96.7
	80	93.3	98.3	96.7		80	93.3	98.3	96.7
	100	93.3	98.3	96.7		100	91.7	98.3	96.7
14	20	88.3	100.0	96.7	28	20	90.0	98.3	96.7
	40	91.7	100.0	96.7		40	91.7	96.7	96.7
	60	93.3	96.7	96.7		60	91.7	96.7	96.7
	80	93.3	96.7	96.7		80	91.7	96.7	96.7
	100	93.3	96.7	96.7		100	91.7	96.7	96.7
16	20	88.3	96.7	93.3	30	20	91.7	100.0	96.7
	40	88.3	96.7	91.7		40	91.7	98.3	98.3
	60	88.3	96.7	91.7		60	91.7	98.3	100.0
	80	93.3	96.7	91.7		80	91.7	98.3	100.0
	100	93.3	96.7	91.7		100	90.0	96.7	100.0
18	20	86.7	95.0	98.3	32	20	86.7	100	98.3
	40	90.0	95.0	92.3		40	90.0	98.3	98.3
	60	91.7	95.0	98.3		60	91.7	98.3	98.3
	80	91.7	93.3	98.3		80	93.3	98.3	98.3
	100	93.3	93.3	98.3		100	93.3	98.3	98.3

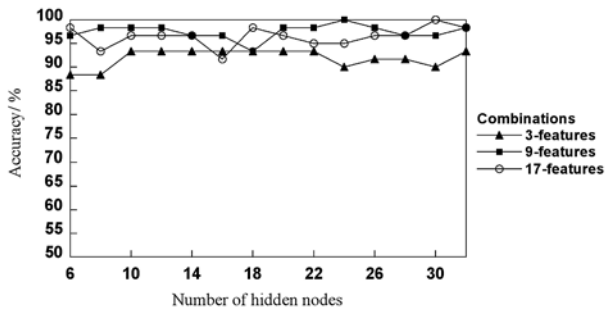


Figure 7. Classification accuracy of different shape descriptors set

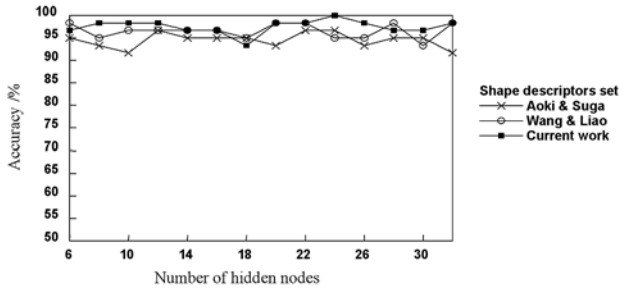


Figure 8. Comparison of classification accuracy with different shape descriptors from various authors

of shape descriptors used in the current work gave a maximum classification accuracy of 100% using 24 nodes in the hidden layer. This classification accuracy is slightly higher than that obtained using the descriptor sets proposed by the other authors. Although the use of the optimised set of shape descriptors in the current work did not produce a significant improvement in the results compared to those of the other authors, this part of the study demonstrates the possibility of comparing different classification techniques using the set of simulated images.

3.1 Application to real images

In order to investigate the usefulness of the simulation study, real images of defects were used for the classification while the training was conducted using the same data extracted from the simulated images. A series of pre-processing procedures using contrast enhancement and noise suppression were carried out to enhance the images. To extract the weld defects, the Background Subtraction Method (BSM) and automatic thresholding were applied⁽¹⁶⁻¹⁷⁾. The background was estimated using the second order polynomial given by⁽²³⁾:

$$B(x,y) = a_0 + a_1x + a_2y + a_3x^2 + a_4y^2 + a_5xy \dots\dots\dots(3)$$

where x and y are pixel coordinates and a_0, \dots, a_5 are fitted constant. The automatic thresholding method was employed where the images were subdivided into a grid of smaller rectangular sub-regions and the pixel having the highest grey value in each sub-region was located. A total of 81 pixels from 9×9 sub-regions were selected

Table 4. Different combination set of shape descriptors

Authors	Details of shape descriptors	Number of shape descriptors
Aoki and Suga ⁽¹⁷⁾	Ratio between horizontal and perpendicular length, vertical length to area ratio, horizontal length to area ratio, complexity, formal coefficient, Heywood diameter.	6
Wang and Liao ⁽¹⁸⁾	Mu, sigma, circularity, compactness, major axis, width and length, elongation, Heywood diameter.	9
Current work	Solidity, extent, standard deviation of radius distribution, compactness, Heywood diameter, vertical length to area ratio, major to minor axis length, curvature, orientation.	9

and the information obtained, *ie* grey value and coordinates, were used to determine the coefficients in equation (3). Subsequently, the background image was estimated by substituting each coordinates into $B(x,y)$.

The results of applying the BSM method for porosity defect are shown in Figure 9. The original image was preprocessed using histogram equalisation to enhance image contrast and noise was eliminated using median filter. The background image was estimated and subtracted from the enhanced image. After the binarisation, noise was removed and the shape descriptors were extracted from the resulting images. These descriptors were then used in the classification. A total of 49 defects from the real images were used for testing the classification accuracy. Some of the real images are shown in Table 5.

Table 5. Real images of radiographs

Defect	Image	Defect	Image
Crack		Slag inclusion	
Cavity		Transverse crack	
Incomplete penetration		Porosity	

The highest classification accuracy achieved by the neural network using features extracted from the real weld images is 97.96% (Figure 10). This value is slightly lower than the accuracy achieved for classifying simulated images (100%). This is expected because real images usually contain much more irregularities and noise than the simulated images of 'ideal' defects.

The classification accuracy of individual defect type is shown in Figure 11. For individual comparison, it was found that the accuracy

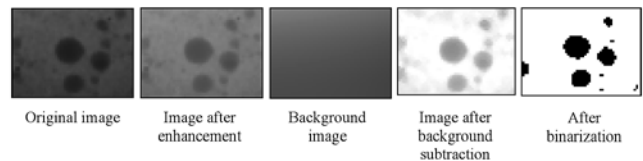


Figure 9. Result of background subtraction

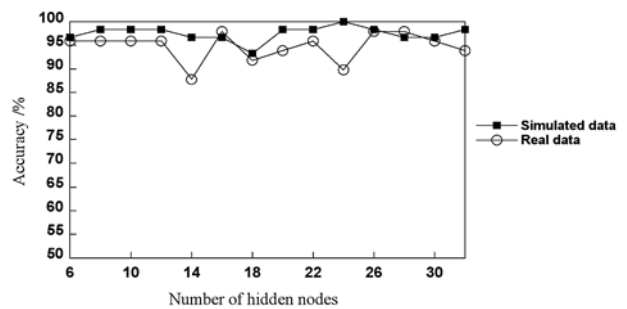


Figure 10. Classification accuracy of simulated and real defects

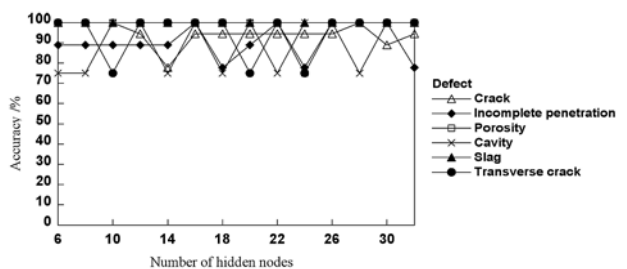


Figure 11. Classification accuracy of individual defects

varies with the number of nodes. The highest average accuracy is 100% for slag inclusion and the lowest is for incomplete penetration (89.29%). In some cases, incomplete penetration was misclassified as crack. As a matter of fact, only the feature *Extent* shows a significant difference from multiple comparison tests while the rest of the features show joining groups for incomplete penetration and crack. In other words, among the nine features selected only one feature, *ie Extent*, offers the highest ability to discriminate these two defects.

5. Conclusion

An effective weld defect classification method using shape descriptors extracted from simulated defects has been developed and demonstrated by classifying real welding defects. The general shape descriptors, which are used to characterise weld defects, were studied and an optimised set of shape descriptors than can effectively classify the weld defects have been identified using a multiple comparison procedure. In order to test the effectiveness of the optimised shape descriptors in classifying defects, shape descriptors extracted from the simulated defects were used to train the network. The simulation study produced a maximum classification accuracy of 100%.

Classification of real defects using the simulated training samples gave the highest overall accuracy of 97.96%. Thus, the proposed defect classification method overcomes the problem of limited real defect samples for classification using neural network. As demonstrated in this work, the simulated defects can be used as ideal or 'benchmark' images to compare the performance of various classification techniques proposed by other authors.

6. Acknowledgements

The authors would like to thank Universiti Sains Malaysia and the Ministry of Science, Technology & Environment (Malaysia) for the offer of the IRPA Grant (Project No: 03-02-05-2190 EA005) that enabled this research to be carried out.

7. References

1. T W Liao and Y M Li, 'An automated radiographic NDT system for weld inspection: Part II: Flaw detection', *NDT&E International*, Vol 31, No 3, pp 183-192, 1998.
2. F Fucsook and M Scharmach, 'Human factors: The NDE reliability of routine radiographic film evaluation', *Proceedings of 15th World Conference on Non-Destructive Testing*, Roma 2000, <http://www.ndt.net/article/wcndt00/papers/idn740/idn740.htm>, (Accessed: 26 July 2006).
3. V Lashkia, 'Defect detection in X-ray images using fuzzy reasoning', *Image and Vision Computing*, Vol 19, pp 261-269, 2001.
4. T W Liao and J W Ni, 'An automated radiographic NDT system for weld inspection: Part I – Flaw detection', *NDT&E International*, Vol 29, No 3, pp 157-162, 1996.
5. T W Liao and K Tang, 'Automated extraction of welds from digitized radiographic images based on MLP neural networks', *Applied Intelligence*, Vol 11, pp 197-218, 1997.
6. T W Liao, D M Li and Y M Li, 'Extraction of welds from radiographic images with fuzzy classifiers', *Information Science*, Vol 126, pp 21-42, 2000.
7. A Kehoe, G A Parker and A LeBlanc, 'Image processing for automatic defect detection in industrial radiographic images', *Image Processing and its Applications – IEEE Third International Conference*, pp 202-206, 1989.
8. K Murakami, 'Image processing for non-destructive testing', *Welding International*, Vol 4 No 2, pp 144-149, 1990.
9. G Bonser and S W Lawson, 'Defect detection in partially completed SAW and TIG welds using online radiography and

- image processing', *Proceedings of SPIE: Process Control and Sensors for Manufacturing*, Vol 3399, pp 231-239, 1998.
10. V Kaftandijan, 'Automatic detection and characteristics of aluminum weld defects: comparison between radiography, radioscopy and human interpretation', *Proceedings of the 7th European Conference on Non-Destructive Testing*, pp 1179-1186, Copenhagen, 26-29 May 1998.
11. R M Palenchika and A V Alekseichuk, 'Flaw detection in radiographic images by structure-adaptive binary segmentation', *Proceedings of International Symposium on Computerized Tomography for Industrial Applications and Image Processing in Radiology*, pp 221-232, Berlin, 15-17 March 1999.
12. H I Shafeeka, E S Gadelmawlab, A A Abdel-Shafiyb and I M Elewab, 'Assessment of welding defects for gas pipeline radiographs using computer vision', *NDT&E International*, Vol 37, pp 291-299, 2004.
13. T W Liao and Y M Li, 'An automated radiographic NDT system, for weld inspection: Part II – Flaw Detection', *NDT & E International*, Vol 31 No 3, pp 183-192, 1998.
14. T W Liao, D-M Li and Y-M Li, 'Detection of welding flaws from radiographic images with fuzzy clustering methods', *Fuzzy Sets and Systems* Vol 108, pp 145-158, 1999.
15. C Jacobsen and U Zscherpel, 'Automated evaluation of digitized radiograph with neuronal methods', *Proceedings of Computed Tomography for Industrial Application and Image Processing in Radiology*, pp 141-152, Berlin, 15-17 March 1999.
16. P Perner, U Zscherpel and C Jacobsen, 'A comparison between neural network and decision trees based on data from industrial radiographic testing', *Pattern Recognition Letters*, Vol 22, pp 47-54, 2001.
17. Z Feher, 'Computer aided processing of industrial radiographs', *Periodica Polytechnica Ser. El. Eng*, 44(3-4), 241-248, 2001.
18. K Aoki and Y Suga, 'Application of artificial neural network to discrimination of defect type in automatic radiographic testing of welds', *ISIJ International*, Vol 39, No 10, pp 1081-1087, 1999.
19. G Wang and T W Liao, 'Automatic identification of different types of welding defects in radiographic images', *NDT&E International*, Vol 35, pp 519-528, 2002.
20. R R da Silva, M H S Siqueira, M P V de Souza, J M A Rebello and L P Caloba, 'Estimated accuracy of classification of defects detected in welded joints by radiographic tests', *NDT&E International*, Vol 38, pp 335-343, 2005.
21. R R da Silva, L P Caloba, M H S Siqueira, J M A Rebello, 'Pattern recognition of weld defects detected by radiographic test', *NDT&E International*, Vol 37, pp 461-470, 2004.
22. K Xu, A R Luxmoore and F Deravi, 'Comparison of shape features for the classification of wear particles', *Engineering Application of Artificial Intelligence*, Vol 10 No 5, pp 485-493, 1997.
23. M Sonka, V Hilavac and R Boyle, *Image Processing, Analysis, and Machine Vision*, Second Edition. PWS Publishing (USA), 1998.
24. B M Metre, M S Kankahalli and W F Lee, 'Shape measures for content based retrieval: A comparison', *Information Processing & Management*, Vol 33, No 3, pp 319-337, 1997.
25. J C Russ, *The Image Processing Handbook*. 3rd edition, CRC Press (USA), 1999.
26. D J Williams and M Shah, 'A fast algorithm for active contours and curvature estimation', *CVGIP: Image Understanding*, Vol 55, pp 14-26, 1992.
27. D C Weber and J H Skillings. *A First Course in the Design of Experiments: A linear Model Approach*. CRC Press (USA), 2000.
28. D C Montgomery, *Design and Analysis of Experiments*, John Wiley & Sons (USA), 1996.