Adaptive neuro fuzzy inference system in modelling/detecting cracks and porosity using liquid penetrant test

Bharat Mehta* and Raman Bedi

Department of Mechanical Engineering,
National Institute of Technology,
Jalandhar-144011, India
Email: bharatmehta106@gmail.com
Email: bediraman74@gmail.com
*Corresponding author

Abstract: This research paper is concerned with a successfully developed adaptive neuro-fuzzy inference system (ANFIS) for detection of indication size (of the penetrant) for liquid penetrant test. The evaluation of the neuro-fuzzy model has been comprehensive; it has been performed using a database of indication size containing 252 valid data points in a structural steel sample and 468 valid data points in high-alloy steel sample. The surface discontinuities are artificially drilled holes varying in diameter and depth from 0.5 to 1.75 mm, respectively, with an increment of 0.25 mm between each hole. The test is conducted at varying time intervals from 2 to 30 min for structural steel samples and 2 to 60 min for high-alloy steel samples, respectively, and the results are obtained till 1/100th of an mm. The ANFIS modelling accuracy is very high, with R^2 values reaching about 0.9924 for the test set when 83:17 ratios for train set:test set are taken. Use of this technique can be useful as this would help with the correct detection of discontinuities and in deciding whether to reject or accept a sample.

Keywords: adaptive neuro fuzzy systems; dye penetrant test; non-destructive evaluation; pin holes; subtractive clustering.

Reference to this paper should be made as follows: Mehta, B. and Bedi, R. (2016) 'Adaptive neuro fuzzy inference system in modelling/detecting cracks and porosity using liquid penetrant test', *Int. J. Experimental Design and Process Optimisation*, Vol. 5, Nos. 1/2, pp.117–132.

Biographical notes: Bharat Mehta is an Assistant Manager, Quality Assurance, at Maruti Suzuki India Ltd., Gurgaon. He received his bachelor's degree in Mechanical Engineering from National Institute of Technology, Jalandhar, India. He is interested in product development, manufacturing and soft computing techniques.

Raman Bedi is an Associate Professor at Department of Mechanical Engineering at National Institute of Technology, Jalandhar. His research interests include mechanical behaviour of composite materials and soft computing techniques.

1 Introduction

As per American Society of Mechanical Engineering (2010), liquid penetrant test (LPT) is a very effective technique to detect discontinuities which are open to the surface of nonporous metals and other materials. Typical discontinuities detectable by this method are cracks, seams, laps, cold shuts, laminations and porosity.

This technique has been use for more than a century and is still prevalent because of the low cost and applicability across different types of materials. It can be used with basically any material given that the surface roughness is low and the material is not porous. The application of this technique is widespread. It is used during testing of aircraft wings, forged/rolled sheets of steel, blades of turbine, porosity in welded joints etc. The disadvantages of this technique are as follows:

1 It can only detect discontinuities, which are open to the surface. Sub-surface crack detection is not possible using LPT (Mehta, 2015).

2 The detection of discontinuities and identification of defects are subjective to the technician performing the test, which leaves a chance of failure in the article due to error in human judgment.

The former disadvantage cannot be solved effectively because of the manufacturing flaws in material, which will always cause sub-surface cracks. To some extent, analysis of the design to remove manufacturing flaws helps reduce the chances of error. The use of novel computational methods in the past few decades has made analysis a crucial part of quality enhancement. Use of finite element methods and computer aided design has aided in making manufacturing processes all the more automated and the flaws in those processes less prevalent.

The latter can be solved by moderating human effort and reducing the chance of error by having an automated system. So, the author suggests the use of modelling techniques to develop a well-informed system for identifying the discontinuities in LPT. The use of artificial neural networks, genetic algorithms and fuzzy logic, being the novel techniques in optimisation and modelling, are used. Use of adaptive neuro-fuzzy inference system (ANFIS) in Baoguang Xu (2013) for the development of an intelligent system using ANFIS in eddy current testing to provide a user with a decision on whether a defect is present or not, and certain properties of unknown crack, such as depth, width and orientation, are anticipated. This paper considers eddy current testing, which is a non-destructive testing method for conductive material detection and uses ANFIS to train a neuro-fuzzy system using signals from known cracks and then utilises this system for predicting the crack information output from given signals of unknown crack.

There is another non-destructive technique, ultrasonic pulse velocity test (UPV), which is used to determine the compressive strength of concrete (Bilgehan and Turgut, 2010). This paper discusses the usage of UPV and density data to predict the compressive strength of concrete. Such a technique can be helpful for health monitoring of structures.

The paper deals with the ingenious use of artificial neural networks in detection of the dimensions of defect (two outputs: width and depth) using six inputs for pipes with the help of ultrasonic guided waves in pipe walls (Cannas, 2005). The utilisation of signal processing techniques in longitudinal and torsional wave modes has been done. The pipes were modelled using finite element method technique, signal processing was applied using discreet wavelet transform and blind separation techniques and, finally, multilayer perceptron networks were used as pattern classifiers.

Also, a paper working on similar lines as this was published, which discussed the use of signal processing for defect characterisation in the ultrasonic testing technique to help in decision-making for human operators (Meksen, 2009). The paper is based on a Kohonen self-organising algorithm, which clusters the signals in the similarity space and uses the results to distinguish between signals corresponding to non-defect, flat defects, and volumetric defects. The results obtained gave a 99% accuracy for the train set and 70% accuracy in the test set, which is average, considering a 75–25 train-test set.

In the present work, ANFIS is used to model the indication size of the penetrant in LPT. A valid database of 252 and 468 data points are used respectively for the application of the ANFIS models. A large part of the data (83%) are taken for training the ANFIS system and the rest (17%) are used to test the data, and the result is quite satisfactory for the predicted vs. actual values, which proves the validity of the proposed methodology. The work done in this paper is unique in its approach because it takes a novel approach towards automation of dye penetrant tests. To be able to identify the dimensions of a discontinuity and categorise it as a defect or otherwise is generally plagued by the possibility of human error. Applying a mathematical approach towards problem-solving rather than experience of a technician would be more reliable and accurate.

2 Indication size data

2.1 For artificial discontinuities (drill holes) on carbon structural steel

A valid database created by Janusz Czuchryj (2012) is used in this work. It refers to the LPT conducted on structural steel plates with simulated discontinuities made on them. The discontinuities varied in depth and diameter from 0.5 to 1.75 mm, respectively, with an increment of 0.25 mm. The LPT was conducted on these discontinuities after 2, 5, 10, 15, 20, 25, and 30 min. In total, 252 valid indication sizes were collected. The data points for 5, 10, 15, 20, 25, and 30 min are considered for developing an ANFIS model (216 data points). Similarly, for the second dataset, out of 468 data points, 432 are considered. The '2 min' data point is not considered to create a spatial distribution among data points and get a better learning curve.

2.2 For artificial discontinuities (pores) on welded joints made of high-alloy steel

A valid database created by Janusz Czuchryj (2015) is used in this work. It refers to the LPT conducted on welded joints made of X5CrNi18-10 Steel (high-alloy steel) with artificial discontinuities made on them. The discontinuities varied in depth and diameter from 0.5 to 1.75 mm, respectively, with an increment of 0.25 mm. The LPT was conducted on these discontinuities after 2, 5, 10, 15 … 60 min. In total, 468 valid indication sizes were collected. The data points for 5, 10, 15 … 60 min are considered for developing an ANFIS model (432 data points).

3 ANFIS modelling application

This application has been adopted as per Bedi (2008) and Tomohiro Takagi (1985):

3.1 Fuzzy logic methods

Fuzzy logic methods have been used to model various highly complex and nonlinear systems based on a set of sample data and fuzzy ‘if–then rules’. A fuzzy inference system can model the qualitative aspects of human knowledge without using any quantitative analyses. The following notation is common in fuzzy logic modelling and is adapted to serve the needs of the present study.

3.1.1 Linguistic variables

Form the basic concept underneath fuzzy logic, i.e., a variable whose values are words rather than numbers. The input linguistic variables specified herein for the indication size in LPT modelling are the following: diameter of discontinuity (*d*), depth of discontinuity (*h*) and development time (*t*). The diameter/size of indication (*D*) is used as the only output variable.

3.1.2 Fuzzy sets

In contrast to a classical set, a fuzzy set does not have a crisp boundary, i.e., the transition from the case ‘belong to a set’ to the case that does ‘not belong to a set’ is gradual. Normally, this smooth transition is characterised by a membership function (MF) which gives flexibility to the fuzzy sets in commonly used modelling linguistic expressions. For the case studied herein, a linguistic expression could be: ‘diameter of discontinuity is close to zero’ or ‘development time (*t*)’ is high and so on.

3.1.3 Membership function

It is the curve which defines the way each point in the input space is mapped to a membership value (or degree of membership) between 0 and 1. The MF type can be any appropriate parameterised MF like triangle, Gaussian or bell-shaped.

3.1.4 Linguistic rules

A set of linguistic ‘if–then’ rules applied on the defined linguistic variables. A single fuzzy ‘if–then’ rule assumes the form ‘If *x* is *A* then *y* is *B*’, where *A* and *B* are linguistic values defined by fuzzy sets on the ranges *X* and *Y*, respectively. The ‘if’ part of the rule ‘*x* is *A*’ is called the antecedent or premise, while the ‘then’ part of the rule ‘*y* is *B*’ is called the consequent or conclusion. Fuzzy ‘if–then’ rules with multiple antecedents like the following are often used.

Rule: If diameter of discontinuity is low, depth of discontinuity is low and the development time is low then the indication size is small.

The resulting output after the described fuzzy logic method has to be defuzzified or else converted to a crisp value by using any of the available defuzzification methods, like the centre of gravity method, etc. The MFs used to represent linguistic variables will have important effect on modelling performance as the type of the MF being used determines when a given rule is to be put into effect or not (in fuzzy logic 'the rule is fired'). Three types of MFs, namely the triangular type, the Gaussian type and the bell-shaped type, have been used in this study to examine the influence of each one of them on the produced data.

4 Artificial neural networks

ANNs are very efficient in adapting and learning, and for this reason they are used as modelling tools in a number of applications. An ANN is made of three types of layers: an input layer which accepts the input variables, herein *d*, *h*, *t*; a set of hidden layers (one or more); and an output layer made of a single neuron which, in the case examined herein, gives the indication size (*D*). Hidden and output layers are in general composed of a number of neurons that perform a specific nonlinear function such as a sigmoid. The neurons of one layer are interconnected to the neurons of the previous layer and after the next layer through weighted links. Each neuron of the hidden and output layers is offset by a threshold value. The back-propagation training algorithm is commonly used to iteratively minimise a cost function by updating the interconnection weights and neuron thresholds. The training process is terminated either when the mean square error (MSE) between the measured data points and the predicted ANN values for all elements in the training set reach a pre-specified threshold or after the completion of a preselected number of iterative learning processes, called learning epochs.

5 Adaptive neuro-fuzzy inference system

Although the fuzzy inference system has a structured knowledge representation in the form of fuzzy 'if-then' rules, it lacks the adaptability to deal with a changing external environment. Therefore, neural network learning concepts have been incorporated in fuzzy inference systems, resulting in adaptive neuro-fuzzy modelling. The adaptive inference system is a network, which consists of a number of interconnected nodes. Each node is characterised by a node function with fixed or adjustable parameters. The network is 'learning' the behaviour of the available data during the training phase by adjusting the parameters of the node functions to fit that data. The basic learning algorithm, the back propagation, aims to minimise a set measure or a defined error, usually the sum of squared differences between the desired and the actual model outputs. The fuzzy modelling was first explored by Tomohiro Takagi (1985). The ANFIS architecture that is used in the present study is based on the first-order Takagi-Sugeno model and it is schematically presented in Figure 1. It is assumed that the indication size of the penetrant (*D*) is a function of the depth of discontinuity (*d*) of the samples, the diameter of discontinuity (*d*) of the samples and the development time (*t*) for LPT. Thus, *d*, *h*, *t* are the input parameters, while the indication size which corresponds to each combination of the three input parameters is the unique output of the ANFIS model. In this model, the *i*th rule for the prediction of indication size of the penetrant can be expressed as follows:

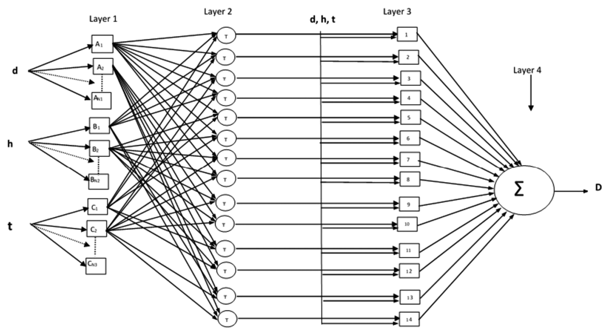
Rule 1:

If d is A_j, h is B_k, t is C_l,

$$\text{Then } f_i = n_i h + o_i d + p_i t \tag{1}$$

where $j = 1, \ldots, N_1$, $k = 1, \ldots, N_2$, $l = 1, \ldots, N_3$ and $i = 1, \ldots, N_1 N_2 N_3$.

Figure 1 Developed ANFIS topology based on the first order Takagi-Sugeno model



A, B and C are the fuzzy sets defined for d, h and t, respectively. N_1, N_2 and N_3 indicate the number of MFs defined by the indicated fuzzy input variables, f is a linear consequent function defined in terms of the input variables and n, o, p, q and r are the consequent parameters of the Takagi-Sugeno fuzzy model (Tomohiro Takagi, 1985). In this model, nodes of the same layer have similar functions, as described below. The output of the *i*th node in layer l is denoted as $O_{l,i}$.

The fuzzy inference system shown in Figure 1 is composed of three layers. Each layer involves several nodes. The output signals from the nodes of the previous layer will be accepted as the input signals in the current layer. After manipulation by the node function in the current layer, the output will be served as input signals for the subsequent layer.

Layer 1:

The first layer of this architecture is the fuzzy layer. Each node of this layer makes the membership grade of a fuzzy set. The membership relationship between the output and input functions of this layer can be expressed as:

$$\begin{aligned} &O_{l,j} = \mu_{A_j} \cdot (d), j = 1, \ldots, N_1 \\ &k = \mu_{B_k}(h), k = 1, \ldots, N_2 \\ &O_{l,j} = \mu_{C_l} \cdot (t), l = 1, \ldots, N_3 \end{aligned} \tag{2}$$

In this layer, the MF can be any appropriately parameterised MF like triangular, Gaussian or bell-shaped.

Layer 2:

Every node in layer 2 is a fixed node, marked by a circle, whose output is the product of all the incoming signals, i.e., T-norm operation:

$$O_{2,i} = \mu_{A_i}(d) \cdot \mu_{B_k}(h) \cdot \mu_{C_l}(t) = \omega_i \tag{3}$$

The output signal ω_i denotes the firing strength of the associated rule. The firing strength is also called 'degree of fulfilment' of the fuzzy rule and represents the degree to which the antecedent part of the rule is satisfied.

Layer 3:

Every node in layer 3 is an adaptive node marked by a square node with a node function like

$$O_{3,i} = \bar{\omega}_l f_i \tag{4}$$

where $\bar{\omega}_i = \frac{\omega_i}{N_i \cdot N_2 \cdot N_3 \cdot \Sigma_{L=1} \omega_L}$ is known as the normalised firing strength. The consequent parameters of f_i in this layer are to be adapted in order to minimise the error between the ANFIS outputs and their experimental results.

Layer 4:

Every node in layer 4 is a fixed node marked by a circle node. The node function computes the overall output by summing the incoming signals, i.e.,

$$O_{4,i} = \sum \bar{\omega}_l f_i = N_f \tag{5}$$

This ANFIS structure represents a four-dimensional space partitioned into $N_1 \times N_2 \times N_3$ regions, each one governed by a fuzzy 'if–then' rule. In other words, the premise part of a rule defines the fuzzy region, while the consequent part specifies the output within the region.

A hybrid learning algorithm is used to adapt the parameters of the first layer, called premise or antecedent parameters, and the parameters of the third layer, referred to as consequent parameters, are adapted to optimise the network. The network uses a combination of back-propagation and least squares method to estimate MF parameters. More specifically, in the forward pass of the hybrid learning algorithm, node outputs go forward till layer 3 and the consequent parameters are identified by the least squares method. In the backward pass, error signals propagate backward and the premise parameters are updated by a gradient descent method.

6 Application design

Accurate identification of indication size is quintessential in deciding the appropriate size of the discontinuity and to decide whether to accept or reject an article on that basis. The field of LPT still uses visual inspection for identification of defects and does not use a lot of technology with respect to the detection of whether discontinuity has been done.

This paper provides with a possible alternative solution to the problem of avoiding human error. The use of neuro-fuzzy inference systems is a first in the field of LPT analysis. The indication size (output) can be defined accurately by feeding in the input variables, namely diameter of discontinuity, depth of discontinuity and development time. The assumptions made in this paper are that the discontinuity is cylindrical in nature and the test is being conducted on a flat surface with the indication being produced parallel to the surface.

There was no need for pre-processing the data and the given data points were directly fed into the ANFIS GUI to obtain the necessary dataset. From the resulting dataset, a training set was constructed by selecting a portion of the data in a random way. The remaining portion was used for the construction of the test set. As one of the major objectives of this work was to investigate the effect of the size of the training set to the modelling ability of the generated ANFIS model, the portion of data used for training and testing was varying. It was decided to start with allocating a portion of 87% of the data for training and the remaining 13% for testing. This case is identified in the sequel as 87–13. During the analysis stages that followed, the portion of data used as a training set were gradually decreased while that used as a testing set were increased, reaching the extreme case of having 13% of the data used for training and 87% used for testing (case 13–87). The performance outcome of ANFIS in all these cases was evaluated. The scatter of the input values is critical for the quality of the evolved model. Less scatter leads to higher modelling efficiency. On the other hand, the speculative nature of the method characterises the input-output process, since difference in general output data should be expected, even for identical input values processed by the same ANFIS model. However, the output values that are produced that way follow the same statistical distribution.

For all the types of datasets described above, ANFIS was constructed using the 'here' types of MF, i.e., the triangular type, Gaussian type and the bell type MF. The number of MFs was chosen to be 5-5-5 corresponding to the inputs d, h, t respectively. In order to enhance the efficiency of the models, the available experimental data were clustered by the subtractive clustering algorithm.

7 Clustering of data

As per Bedi (2008), clustering of numerical data is the basis of many classifications and system modelling algorithms. The purpose of clustering is to identify natural groupings of data from a large dataset to produce a concise representation of system behaviour. A clustering technique can be used to generate a Takagi-Sugeno type fuzzy inference system which best models data behaviour using a minimum number of fuzzy rules, thus preventing the explosion of rules. The rules themselves can be partitioned according to the fuzzy qualities associated with each one of the data clusters. Various methods of clustering have been described in the literature (Yager and Filev, 1994; Chiu, 1994). The subtractive clustering method (Chiu, 1994) presents an efficient method for estimating cluster centres of numerical data. This method can be used to determine the number of clusters and their initial values for initialising iterative optimisation-based clustering algorithms such as fuzzy C-means.

As per subtractive clustering, one-pass algorithm is used for estimating the number of clusters and the cluster centres in a dataset. The cluster estimates, which are obtained from the *subclust* function, can be used to initialise iterative optimisation-based clustering methods (*fcm*) and model identification methods (like *anfis*). The *subclust* function finds the clusters by using the subtractive clustering method. The *genfis2* function builds on the *subclust* function to provide a fast, one-pass method to take input-output training data and generate a Sugeno-type fuzzy inference system that models the data behaviour.

So, data clustering primarily enhances the modelling accuracy, as would be clear from the results. This clustering would be involving the definition of 23 MFs for input variables as compared to 5 MFs for the triangular, Gaussian and bell-shaped MFs. The subtractive clustering is performed based on the default parameters, such as range of influence = 0.5, squash factor = 1.25, Accept ratio = 0.5, Reject ratio = 0.15. Range of influence is basically the cluster radius when considering the data space as unit hypercube. Squash factor is multiplication factor to the radii values to find neighbourhood of clusters involved. The accept ratio sets the potential, as a fraction of the potential of first cluster centre, above which another data point is accepted as a cluster centre. Reject ratio sets the potential, as a fraction of the potential of the first cluster centre, below which another data point is rejected as a cluster centre.

8 Results and discussion

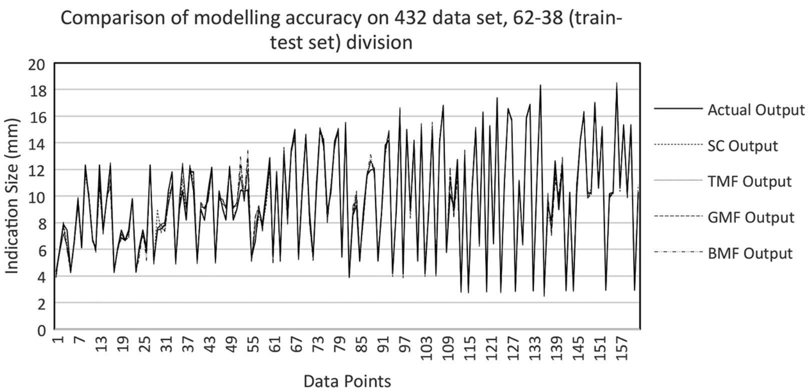
The basic purpose of this paper is to explore the applicability of analytics in NDT techniques. The results obtained to detect indication size predicted by ANFIS model produce a R^2 value as high as 0.9924, which is exemplary.

Two different datasets of 216 and 432 data points were taken to produce a total of eight separate ANFIS architectures (four for each dataset). The architectures were based on triangular MF, Gaussian MF, bell-shaped MF and subtractive clustering. Out of these, four architectures comply with the indication sizes for the assessment of surface discontinuities in carbon structural steel and four architectures comply with the size of pores in welded joints made of high-alloy steel.

Since ANFIS operates on the applicability of the given dataset, which is used in training the system to suit the test set which is in turn used for accurate prediction of the output, we saw that the dataset with 432 data points gives more accuracy because it avoids the problem of overfitting of the model, wherein the modelling software performs very well with the training data but not with the test data. Also, the results deal with two different types of metals: structural steel and high-alloy steels, which makes the applicability of these results all the more widespread. The applications can involve the development of a chart which can let the technicians know which sample to be rejected by looking at the size of the indication, thus reducing the human error of judgment and making the process of LPT all the more accurate.

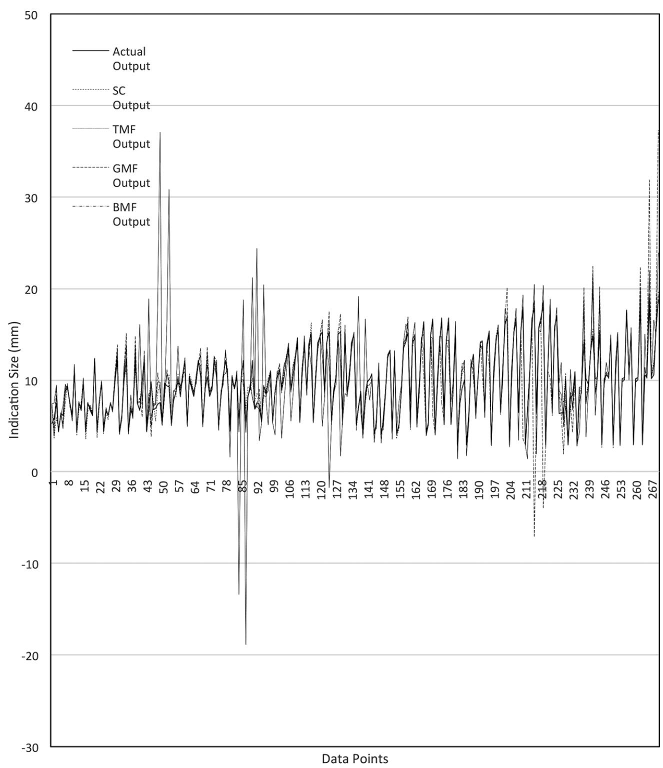
Comparison of modelling accuracy 432 training 62–38 compared the different sets of MFs for 432 data points divided in the ratio of 270/162 (train test) and showed the accuracy of each of them vs. the actual output set (Figure 2). We see that most of the MFs that provided a more or less equal value with the correlation coefficient are located between 0.9858 and 0.9963.

Figure 2 Comparison of modelling accuracy 432 training 62–38



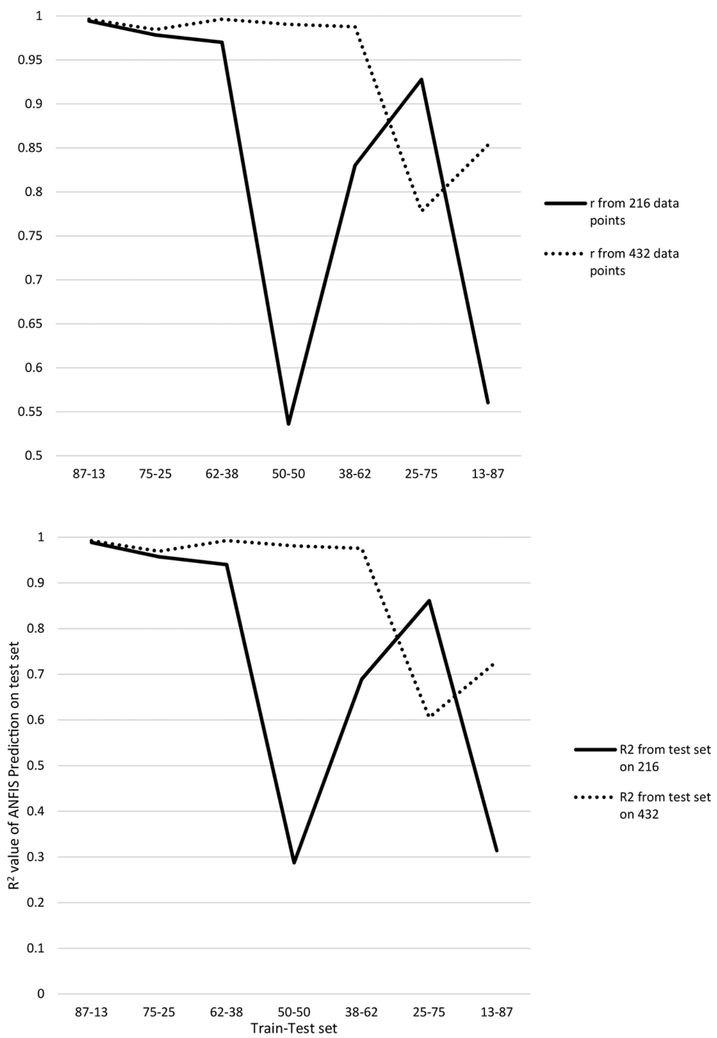
Comparison of modelling accuracy 432 training 38–62 compared the different sets of MFs for 432 data points divided in the ratio of 162/270 (train test) and showed the accuracy of each of them vs. the actual output set (Figure 3). We see that the correlation coefficient drops to a value as low as 0.7139 for bell-shaped MF, whereas the MF involving subtractive clustering technique still retained a high correlation coefficient of 0.9877. The chart of the MFs being compared demonstrated the same graphically.

Figure 3 Comparison of modelling accuracy 432 training 38–62



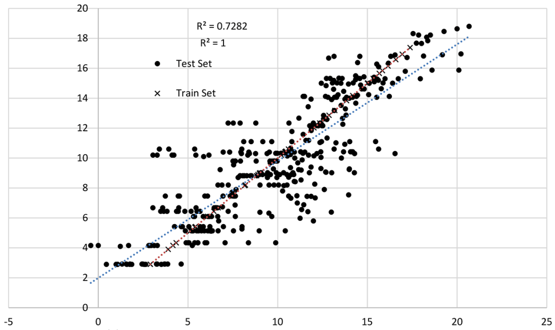
Comparison between 216 and 432, *r* and R^2 described the comparison of correlation coefficient and R^2 for both the datasets for a subtractive clustering dataset (Figure 4). It is used as a method to know how the variance in the above parameters can be compared with respect to a change in the train-test set. We observed that the dataset with 432 data points clearly performs more consistently than the dataset with 216 data points. This verified the fact that more the number of data points, more would be the accuracy of the neuro-fuzzy system.

Figure 4 Comparison between 216 vs. 432, *r* and R^2

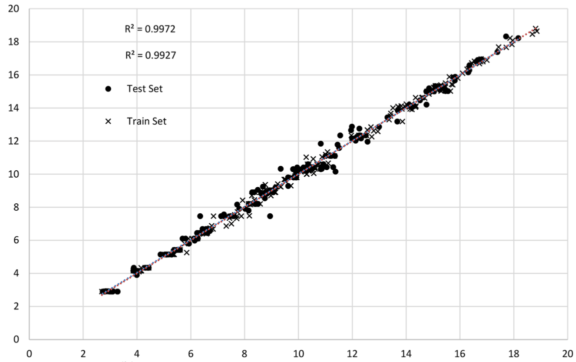


Comparison of modelling ability in ANFIS with SC has shown the comparison of R^2 values between the test set and train set for the datasets conditioned to subtractive clustering MF (Figure 5). These comparisons made it clear as to how R^2 for a test set decreases from 0.9924 to 0.6054, whereas for the train set the values are always as high as 0.99.

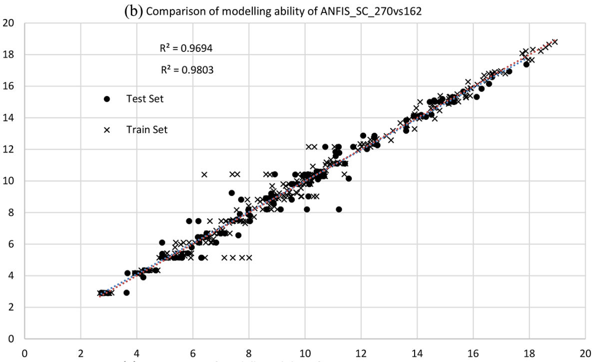
Figure 5 Comparison of modelling ability in ANFIS with SC



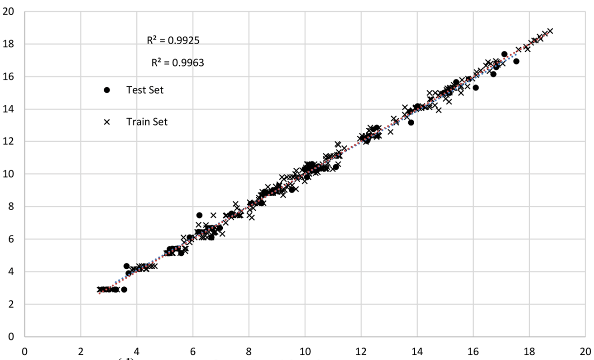
(a) Comparison of modelling ability of ANFIS_SC_54vs378



(b) Comparison of modelling ability of ANFIS_SC_270vs162



(c) Comparison of modelling ability of ANFIS_SC_324vs108



(d) Comparison of modelling ability of ANFIS_SC_378vs54

Adaptive neuro-fuzzy inference system for different ANFIS structures 50 vs. 50, 87 vs. 13 compared two test sets with a ratio of train-test set as 50–50 and 87–13 on the parameters of correlation coefficient, MSE, R^2 and maximum/minimum absolute error (Table 1). We saw that the errors increase as we decrease the number of training data. This is in direct relation to the accuracy of the values being predicted by the ANFIS architecture for those data points.

Table 1 ANFIS for different ANFIS structures 50 vs. 50, 87 vs. 13

	Dataset with 50% training - 50% testing set				*Dataset with 87% training - 13% testing set*			
	Triangular MF	*Gaussian MF*	*Bell MF*	*Subtractive clustering*	*Triangular MF*	*Gaussian MF*	*Bell MF*	*Subtractive clustering*
Correlation coefficient	0.9857	0.9782	0.9726	0.9904	0.9932	0.9909	0.9896	0.9962
MSE	0.4514	0.686	0.8732	0.2992	0.2183	0.2981	0.3394	0.1269
R^2	0.9716	0.957	0.946	0.981	0.9866	0.982	0.9794	0.9924
Minimum absolute error	0.0008	0.0001	0.0009	0.0014	0.0084	0.0407	0.0105	0.0018
Maximum absolute error	4.6407	4.6252	4.9497	4.8363	1.2572	1.4757	1.5512	1.2282

Adaptive neuro-fuzzy inference system for different datasets (75–25, 13–87) compared the two different datasets (216 and 432 data points) for different train-test sets and then comprehensively compared the two systems (Table 2). We could observe better accuracy of the system with 432 data points under parameters like correlation coefficient, R^2, MSE and minimum/maximum absolute error.

Table 2 ANFIS for different datasets (75–25, 13–87)

	Dataset with 216 data points				*Dataset with 432 data points*			
	Triangular MF	*Gaussian MF*	*Bell MF*	*Subtractive Clustering*	*Triangular MF*	*Gaussian MF*	*Bell MF*	*Subtractive clustering*
(a) ANFIS for different datasets (75–25)								
Correlation coefficient	0.927	0.865	0.6771	0.9784	0.9845	0.9836	0.98	0.9845
MSE	2.2159	5.282	26.0872	0.473	0.4226	0.4504	0.5479	0.4264
R^2	0.8594	0.7482	0.4585	0.9572	0.9694	0.9675	0.9605	0.9693
Minimum absolute error	0.0098	0.0006	0.0053	0.0092	0.002	0.0128	0.0004	0
Maximum absolute error	7.2845	11.5555	25.9519	2.2686	2.3865	2.4916	3.0745	3.0091

Table 2 ANFIS for different datasets (75–25, 13–87)

	Dataset with 216 data points				*Dataset with 432 data points*			
	Triangular MF	*Gaussian MF*	*Bell MF*	*Subtractive Clustering*	*Triangular MF*	*Gaussian MF*	*Bell MF*	*Subtractive clustering*
(b) ANFIS for different datasets (13–87)								
Correlation coefficient	0.4014	0.4268	0.4226	0.5605	0.46	0.506	0.5079	0.8533
MSE	43.1609	41.1833	41.8091	9.9659	34.419	28.7415	28.6827	5.1812
R^2	0.1611	0.1822	0.1786	0.3142	0.2116	0.2561	0.258	0.7282
Minimum absolute error	0.713	0.2047	0.143	0.0044	0.0118	0.0084	0.0019	0.0106
Maximum absolute error	13.47	13.1297	12.8515	14.3795	18.8	18.45	17.9388	7.1973

Error on training/testing set SC 432 compared the errors in terms of correlation coefficient, R^2, MSE and minimum/maximum absolute error and compared it for the datasets with seven ratios of train-test set, i.e., 87–13, 75–25, 62–38, 50–50, 38–62, 25–75, 13–87 (Table 3). These tables show how the errors increase with an increase in the number of test set. This meant that the predicted value using ANFIS was further away from the actual value of the output.

Table 3 Error on training/testing set SC 432

	87–13	*75–25*	*63–37*	*50–50*	*37–63*	*25–75*	*13–87*
(a) Error on testing set (subtractive clustering, 432 data points)							
Correlation coefficient	0.9962	0.9845	0.9963	0.9904	0.9877	0.778	0.8533
MSE	0.1269	0.4264	0.1142	0.2992	0.3982	8.1087	5.1812
R^2	0.9924	0.9693	0.9927	0.981	0.9757	0.6054	0.7282
Minimum absolute error	0.0018	0	0.0058	0.0014	0.0011	0.0067	0.0106
Maximum absolute error	1.2282	3.0091	1.4846	4.8363	2.1493	17.1798	7.1973
(b) Error on training set (subtractive clustering, 432 data points)							
Correlation coefficient	0.9981	0.9901	0.9985	0.9914	0.9982	0.9999	1
MSE	0.058	0.3218	0.0447	0.0265	0.055	0.0021	0
R^2	0.9962	0.9803	0.9971	0.9982	0.9964	0.9998	1
Minimum absolute error	0.0002	0.0027	0.0003	0.0001	0.0008	0.0001	0
Maximum absolute error	0.826	3.9934	0.7329	0.5969	0.993	0.2108	0.0001

9 Conclusion

This paper is a first of its kind in terms of its approach towards LPT. This paper not only quantifies the approximation made by NDT technicians all around the world while thinking of passing or failing a sample but also involves the use of a novel modelling technique in doing so. And, at the same time, the modelling technique was used comprehensively by involving the use of different types of steels and different MFs for evaluation. It was seen that using subtractive clustering technique enhances the accuracy of the model to a great extent. Such a system could be beneficial as there is no scientific study present in this non-destructive testing. To enhance the applicability of this technique, more testing using different types of penetrants, more comprehensive discontinuity sizes, different types of materials etc. should be done to develop a full-fledged system.

References

American Society of Mechanical Engineering. (2010) ASME section V: boiler and pressure vessel code, US.

Baoguang Xu, W.F. (2013). *Intelligent eddy current crack detection system design based on neuro-fuzzy logic*. April 24, 2015, *www.ndt.net*; *http://www.ndt.net/article/ndt-canada2013/content/papers/46_Xie.pdf*

Bedi, R. (2008) 'Adaptive neuro fuzzy inference system in modelling fatigue life of multidirectional composite laminates', *Computation Material Sciences*, pp.1086–1093.

Bilgehan, M. and Turgut, P. (2010) 'The use of neural networks in concrete compressive strength estimation', *Computers and Concrete*, Vol. 7, No. 3, pp.271–283.

Cannas, B. (2005) Neural NDT by means of reflected longitudinal and Torsional waves, *5th WSEAS/IASME Int. Conf. on Systems Theory and Scientific Computation*, Malta, pp.94–102.

Chiu, S. (1994) 'Fuzzy model identification based on cluster estimation', *Journal of Intelligent and Fuzzy Systems*, Vol. 2, No. 3, pp.267–268.

Fuzzy Logic Toolbox-User Guide. (n.d.) February 5, 2015, *http://in.mathworks.com*; *http://in.mathworks.com/help/fuzzy/anfis.html*

Janusz Czuchryj, K.H. (2012) *Dye-Penetrant Method Assessment of Size of Surface Discontinuities in Products made of Carbon Structural Steel*, Biuletyn Instytutu Spawalnictwa, Gliwice.

Janusz Czuchryj, P.I. (2015) *Dye-Penetrant Method Assessment of the Size of Pores in Welded Joints made of High-Alloy Steel*, Biuletyn Instytutu Spawalnictwa, Gliwice.

Mehta, B. (2015) 'Surface and sub surface crack analysis in steel using MPT and LPT', *Materials Evaluation Journal*, pp.464–469.

Meksen, T.M. (2009) 'Defects clustering using kohonen networks during ultrasonic inspection', *IAENG International Journal of Computer Science*.

Tomohiro Takagi, M.S. (1985) 'Fuzzy identification of systems and its application to modelling and control', *IEEE Transactions on Systems, Man and Cybernetics*, pp.116–132.

Yager, R.R. and Filev, D.P. (1994) 'Generation of fuzzy rules by mountain clustering', *Journal of Intelligent & Fuzzy Systems*, Vol. 2, No. 3, pp.209–219.

Nomenclature

d	Diameter of discontinuity (mm)
h	Depth of discontinuity (mm)
t	Time taken for development (seconds)
A	Fuzzy set defined for d
B	Fuzzy set defined for h
C	Fuzzy set defined for t
N_1	Number of membership functions for d, herein, $N_1 = 5$
N_2	Number of membership functions for h, herein, $N_2 = 5$
N_3	Number of membership functions for t, herein, $N_3 = 5$
f	Linear consequent function of Takagi-Sugeno fuzzy model
n,o,p,q,r	Consequent parameters of the Takagi-Sugeno fuzzy model
$O_{i,j}$	Output of the ith node of the jth layer of ANFIS
μ_{Aj}	Membership function for d, $j = 1 \ldots N_1$
μ_{Bk}	Membership function for h, $k = 1 \ldots N_2$
μ_{Cl}	Membership function for t, $l = 1 \ldots N_3$
ω_i	Firing strength of the ith rule