CAAP Quarterly Report

Date of Report: April 1, 2020

Prepared for: U.S. DOT Pipeline and Hazardous Materials Safety Administration

Contract Number: 693JK31950001CAAP

Project Title: Improved NDT Detection and Probabilistic Failure Prediction for Interacting Pipeline Anomalies

Prepared by: Sijun Niu, Alexander Le, and Dr. Vikas Srivastava

Contact Information:
Vikas Srivastava
Assistant Professor of Engineering
Brown University
184 Hope Street, Box D,
Providence, RI 02912
Email: vikas_srivastava@brown.edu
Phone: 401-863-2863

For quarterly period ending: March 31st, 2020

Business and Activity Section

(a) Contract Activity

No modifications were made to the contract.

(b) Status Update of Past Quarter Activities

We generated a reasonably large dataset using finite element numerical simulations of ultrasonic testing of a steel plate with various crack geometries for our neural network study. One of the main research objective is to accurately predict size, location and orientation as three key features for an embedded elliptical crack through our combined computational simulations and machine learning research. We performed feature extraction for the *ultrasonic* signals by using wavelet packet transform and used these extracted features to train the neural network. Our early neural network analysis tool was built in MATLAB. In the last quarter, we developed a crack size prediction methodology using neural network. The results from our early research show that we are able to achieve very good accuracy in crack size prediction using our newly developed methodology combining finite numerical simulation data with a neural network algorithm.

(c) Cost share activity

Partial support for 1 graduate student tuition were provided by Brown University School of Engineering as per the cost share agreement.

(d) Task X: Task Title

1. Background and Objectives in the 2nd Quarter

1.1 Background

Application of popular ultrasonic non-destructive testing (NDT) technique remains challenging for crack characterization in pipelines as data interpretation if performed by people, leading to significant uncertainty in accurate crack feature predictions.[1][2] Using an automated solution to detect cracks is gaining more attention and has the potential to provide significantly more accurate results.[3][4][5] In order to use the full potential of automation, or machine learning, there are several underlying problems that need to be solved. The leading limitation for ultrasonic testing (UT) is the lack of dataset. Current machine learning algorithms often need a substantially large dataset to be trained in order to reach reliable accuracy. For example, the benchmark problem in machine learning aided pattern recognition uses the Modified National Institute of Standards and Technology database (MNIST) which contains 70000 images of hand-written digits in total for training, validation and testing. On top of that, each image is normalized to the same pixel value for standardization and consistency. Unfortunately a large UT dataset from field does not exist due to impracticality and high expense of such an endeavor. Another challenge is physically motivated correct feature selection for our neural network (NN). It is known that features (or inputs) have significant impact on the accuracy of a NN. Due to the high sampling rate of UT, it is not optimum to use the full signal, which is over thousands of pairs of time/amplitude data, as inputs to NN. Hence reliable feature extraction technique is necessary and crucial in our case.

1.2 Objectives in the 2nd Quarter

We developed a finite element numerical simulation platform for UT during the first quarter. In the recent (second) quarter, our aim was to study and address some of the leading limitations that are discussed in the previous background section. First, we aimed to generate a standard computational methodology for large number of physically correct simulations. Second, we aimed to study different crack sizes in a systematic way to build a numerical simulation based large UT dataset to train our NN. Our third aim was to find the best feature selection criterion for crack size detection to train our NN. Lastly, we aimed to build and train an early NN in MATLAB to demonstrate accuracy and feasibility of our newly developed method.

2. Experimental and Computational Program in the 2nd Quarter

2.1 Experimental design

No experimental results to report for the 2nd quarter.

2.2 Computational setup

All computations were conducted on an existing workstation desktop (early computations are relatively smaller sized).

We studied finite element based numerical simulation requirements for sound wave propagation in steel pipelines. All of our numerical study used an ultrasound wave of 5 MHz frequency and wavelength of ~1.2 mm, the numerically stability requirement we obtained that 10-15 meshes per wavelength provides a stable practical element size.

A steel plate geometry (which we will refer to as 'plate' later on) with width 60 mm and thickness 20 mm was used in our simulations. 50 elements in total on the bottom surfaces are assumed to be both the ultrasound signal exciter and receiver by monitoring the longitudinal wave (common in industry practice) for embedded cracks inside the plate. A short 5 mm long ultrasound signal exciter with 5 MHz raised-cosine type waveform was applied as boundary condition to one edge of the plate thickness. Profile for this waveform is shown in **Figure 1**. Step size is fixed at $2 \times 10^{-9}\ s$ which corresponds to a 500 MHz sampling rate. Artificial anomalies in the form of elliptical cracks are placed in the plate. We conducted dynamic numerical simulations in Abaqus/Explicit and analyzed the displacement history profile at the selected point receiver locations.

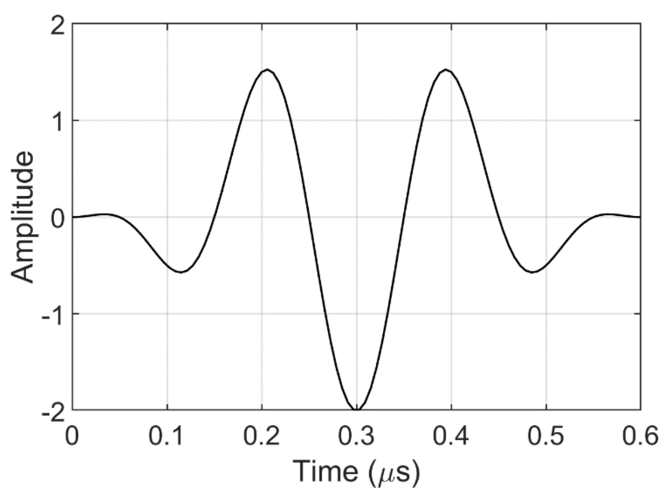

Figure 1. 5MHz, 3 period raised-cosine type pulse signal used in the simulations.

3. Results and discussion

As described in the object section, we aim to address three major problems in this quarter, namely

- To generate a large dataset using UT numerical simulations where crack size was systematically varied

- To search and perform state-of-the-art feature extraction technique
- To build a small NN for an early study

We will discuss all three problems in the following subsections.

3.1 Technical approach and result

One key ingredient to reliable use of a machine learning algorithm is the application of adequate training data. We performed a detailed literature search to study the past efforts towards creating a dataset for machine learning training. Selected work is summarized in **Table 1**. The feature extraction techniques are also listed here and will be discussed later on.

Table 1. Some previous work summarized

Reference	Feature extraction	Sample number	Size	Location	Orientation	Type
Sampath et al.[6]	DWT	240	No	No	No	Yes
Margrave et al.[7]	No	90	No	No	No	Yes
Veiga et al.[8]	No	50	No	No	No	Yes
Martin et al.[9]	Manual selection	483	No	No	No	Yes
Liu et al.[10]	WPT	600	No	No	No	Yes

Note that all the listed works above are experiment based, hence the sample numbers are generally small in comparison to the MNIST example. Moreover, it is obvious that all listed work focused on classification of crack types rather than geometric properties. It is mostly due to the fact that crack type posed a classification problem which is simpler than geometric properties such as crack size which requires a regression for a continuous output variable and creates additional complexity. Fabrication of embedded cracks can be very arduous and expensive which prohibits obtaining large dataset through controlled experimentation.

To address this problem, we conducted a systematic numerical implementation of UT simulations. Our current focus is elliptical embedded cracks (most prominent type of cracks), but our methodology can easily be generalized to other crack/flaw shape types. Five parameters were identified for an elliptical crack and illustrated in **Figure 2**.

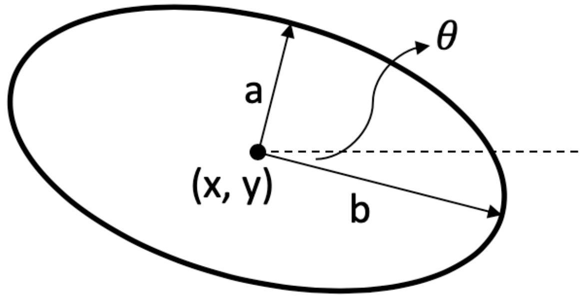

Figure 2. Five geometric parameters identified for an elliptical crack

The location of the crack is characterized by a vector with two parameters x and y, indicating the center of the ellipse. a is the short axis of the ellipse, b is the long axis of the ellipse and they both define the size of the crack. θ is the angle that long axis made with the horizontal direction, used to characterize the orientation. Through the implementation of a python script, we were able to submit a parametrized job in Abaqus with varying geometric properties.

For our first study of varying crack size, we fixed four of the parameters and allowed only for the long axis b to vary. We assigned, $x = 0, y = 17\ mm,\ y = 1\ mm, \theta = 0$ and increased b monotonically from 1 mm to 4 mm with 10 microns increment. We generated results from a total of 287 simulations. Two plate geometries are shown here in **Figure 3** which represent two limiting cases $b = 1\ mm$ and $b = 4\ mm$.

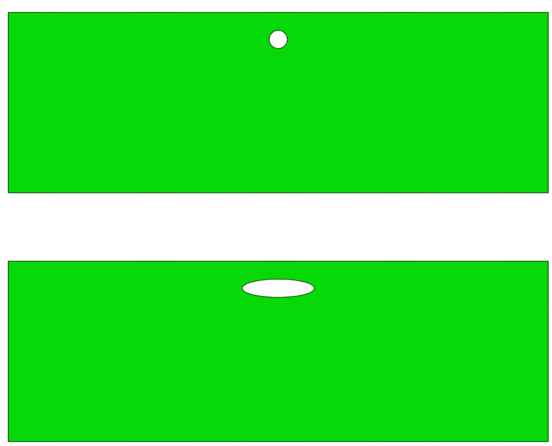

Figure 3. Simulated plates with elliptical cracks with long axis b = 1 mm (top) and b = 4 mm (bottom).

In addition to varying only the long axis of the ellipse, *we also created two additional datasets that will be used for future NN training and predictive studies of location and orientation*. Location is a variable with y ranging from 7 mm to 18 mm besides the varying long axis b, and the other one has the angle θ varying between 0 and π with all other parameters fixed. The parameters of three datasets are summarized in **Table 2**. Pulse generation and signal collection follows the same procedure as discussed in computation setup in section 2.2.

Table 2. Summarize of datasets with different parameters, bold numbers indicate a range in which the corresponding parameter varies

	x (mm)	y (mm)	a (mm)	b (mm)	θ	Number of simulations
Dataset 1 (for size)	0	17	1	**[1, 4]**	0	287
Dataset 2 (for location and size)	0	**[7, 18]**	1	**[1, 4]**	0	958
Dataset 3 (for orientation)	0	15	0.5	1.5	$\boldsymbol{[0, \pi]}$	446

A reliable machine learning algorithm will depend heavily on the selection of physically important input features. In signal processing, there are many existing techniques that can serve for feature extraction. One of the most traditional technique is the fast Fourier transform (FFT) and the FFT coefficients are used as the features. However, due to the transient nature and the limited time window of an ultrasonic signal, FFT based features are mostly outperformed by wavelet based techniques. Wavelets are different than normal waves in the sense that it may be irregular in shape, and normally lasts only for a limited period of time. Also, a wavelet can serve as both a deterministic and nondeterministic template for analyzing time-varying or nonstationary signals by decomposing the signal into a 2D, time-frequency domain.[11] There are many varieties of wavelet based transform such as continuous wavelet transform (CWT), discrete wavelet transform (DWT) and wavelet packet transform (WPT). We find WPT to be most appropriate state-of-the-art technique for our problem. Unlike DWT which loses some resolution on the middle and high frequency, WPT is its extension that simultaneously breaks up detail and approximate versions and have the same frequency bandwidths in each resolution.[3] Previous studies have found that WPT based feature extraction performs better than DWT.[12] Hence we selected WPT as our UT signal feature extraction technique. Signal decomposition performed by WPT is demonstrated in **Figure 4**.

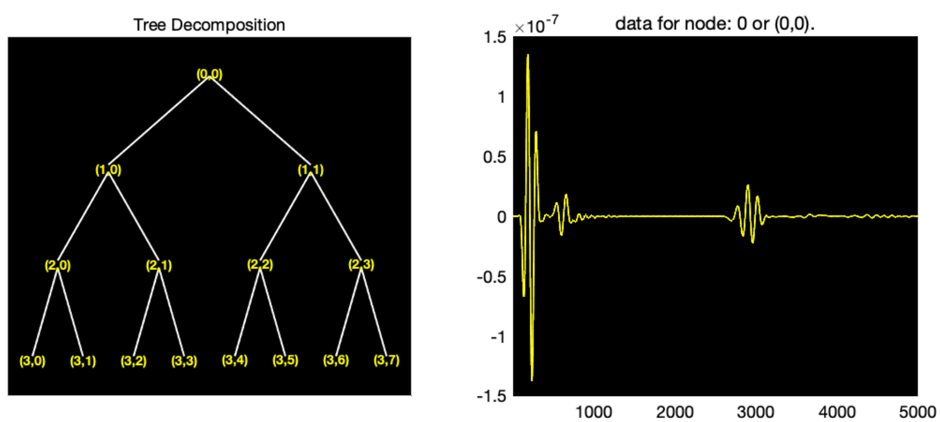

Figure 4. WPT decomposition. Left panel: level 3 tree decomposition schematic; right panel: original signal, denoted by tree node (0,0)

As illustrated in **Figure 4**, a level 3 decomposition breaks down the original signal into 8 pieces. More generally, a level n of decomposition breaks down the signal into 2^n sub-signals or resolutions. In each resolution, we have approximately $m/2^n$ sampling points where m is the original sampling number. Here we define the sampling number to be

$$\text{sampling number } m = \frac{time\ of\ duration\ t}{step\ size\ \Delta t}$$

All sampling points in each resolution have corresponding WPT coefficients $d_{j,i}^k$, where k is the number of the point, j is the current level and i is the node number. The coefficients are determined by the set of orthonormal wavelets we use as the base wavelets. The number of WPT coefficient is roughly the same as sampling number which is over 5000 in our case. Hence we define two quantities only for the nodes in the bottom level:

$$Energy: E_i = \sum_k \left(d_{b,i}^k\right)^2$$

$$Entropy: S_i = \sum_k \left(d_{b,i}^k\right)^2 log[\left(d_{b,i}^k\right)^2]$$

where b stands for the bottom level. This reduced our number of features to 2^n, where n is the level of decomposition. These two quantities will be the features that go into our NN.

We used MALTAB's deep learning toolbox to build our neural network. For a faster training process, we started with a simple three-layer structure: one input layer, one hidden layer and one output layer. We selected 'Levenberg-Marquardt backpropagation' function as our training function. Dataset 1 was selected and divided into training, validation and testing data. We performed NN regression for our *two selected features which are energy and entropy*. The performances are shown in **Figure 5**. Note that the performance will be different each time we run our NN since the split of training, validation and testing data is random. We have selected two typical performances to be shown here.

Two features have similar outputs, with energy feature having slightly less error comparing to entropy. It can be seen that for dataset 1 in which the crack size varies, the mean squared error converges very quickly and validation stops at only 15 epochs using energy feature and 20 epochs for entropy, which is a very good NN performance. The error histograms on the right plot the absolute error which has a unit of millimeter. It also reveals that the prediction peaks very near to the middle and represents very small error. **Figures 6 and 7** show the regression plots for the two features.

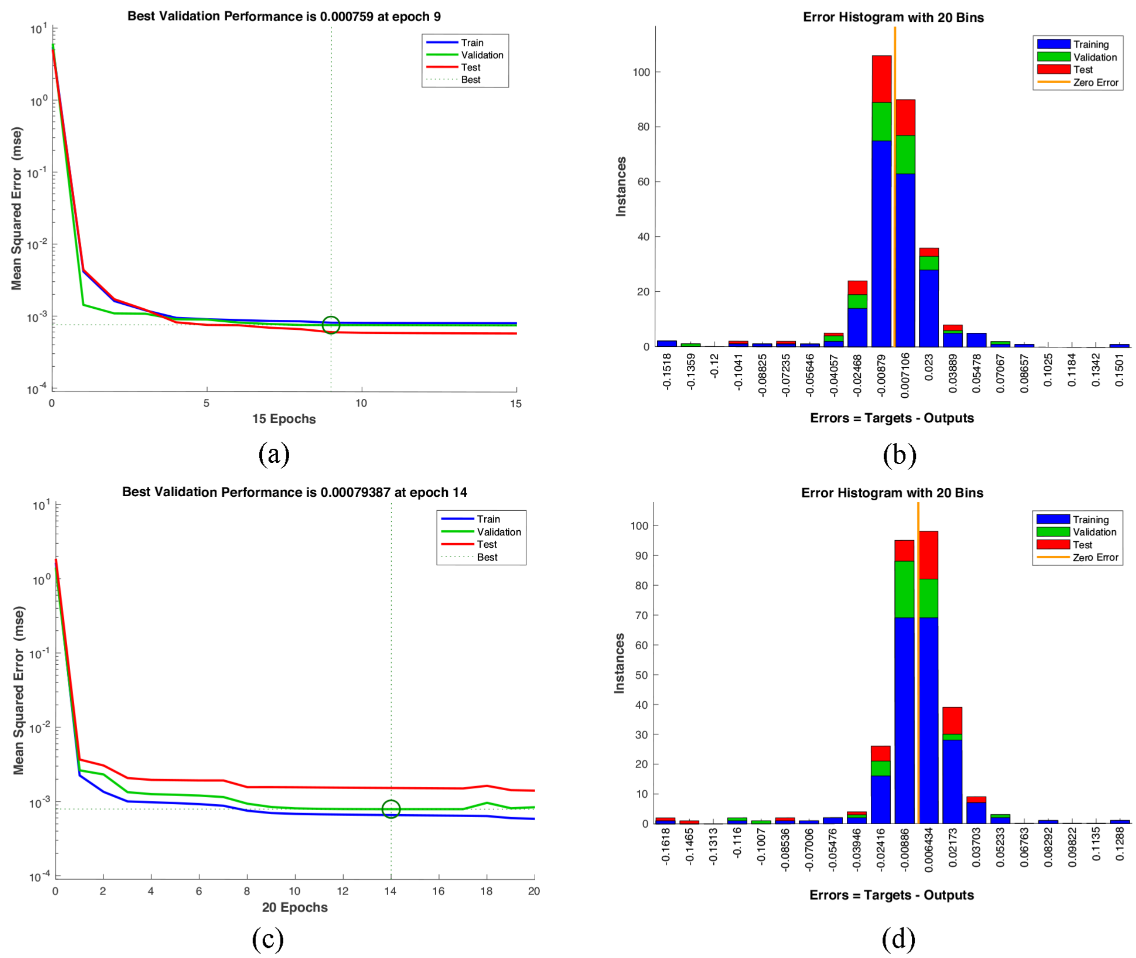

Figure 5. MATLAB NN performance using 'energy' feature (top) and 'entropy' feature (bottom). (a) and (c): MSE vs. epoch, (b) and (d): Error histogram

Again the results are very similar for the two types of features. For all three types of data we reached R values surpassing 0.999, and these results are very promising. The individual circles in the plot represents real data (long axis value *b*) and the straight line represents the output fit by our NN. It is seen that when the long axis *b* is small, there exists some error in the prediction. However, when *b* is roughly above 1.5 mm, the prediction from our NN matches extremely well with the real crack size data.

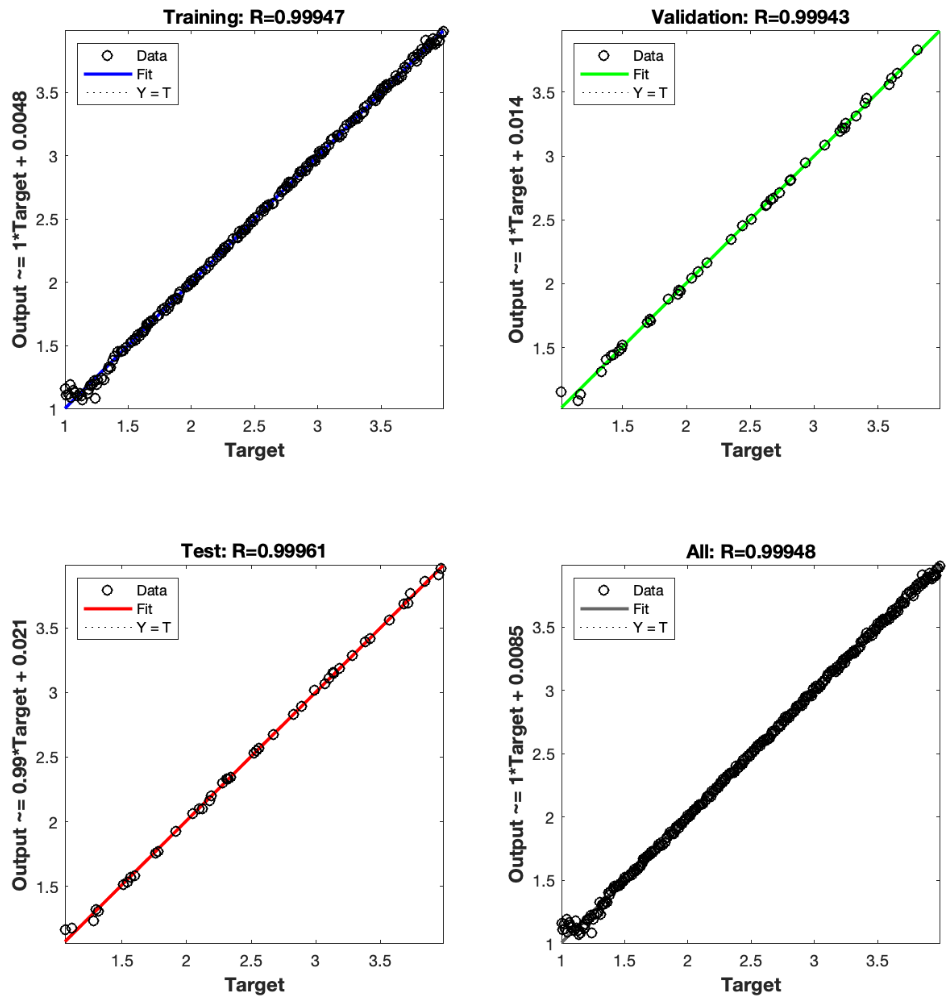

Figure 6. MATLAB NN regression (prediction vs. data) plot using energy feature

3.2 Discussion

In the recent quarter, we successfully completed three important studies including building datasets for various crack geometries, feature extraction and performing an early NN training and its prediction ability studies. We have independently generated three datasets regarding three important geometric properties of an embedded elliptical crack, i.e. the size, location and orientation. They contain over a thousand different crack cases which would have been impractical to create from an experimental approach point of view. Out of these three crack features, we studied crack size in the current quarter. We identified the most suitable feature extraction technique WPT for our algorithm and implemented it in MATLAB. We selected 'energy' and 'entropy' as the input features. Lastly we used MATLAB deep learning toolbox and use a three layer NN to characterize crack size variation using dataset 1. This preliminary result proves that our NN is capable of characterizing crack size, independent of any other variables. We compared the performance of energy input and entropy input, and both gave promising results regarding crack size prediction.

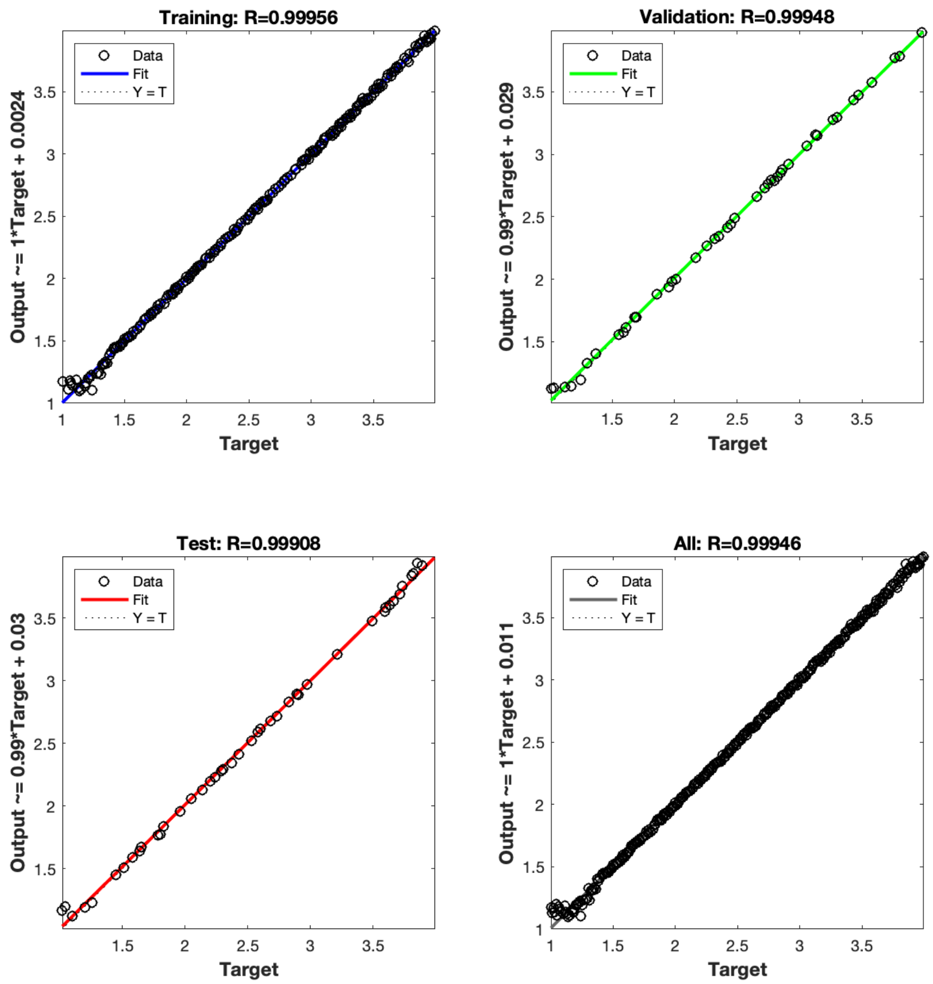

Figure 7. MATLAB NN regression (prediction vs. data) plot using entropy feature

4. Future work

We have demonstrated a NN capable of predicting crack size when it is perfectly horizontal. In the future, we will investigate and study cracks with combined geometric properties (including location and orientation of the crack) and several more comprehensive datasets will be built. Also, for our early studies, we have used a one hidden layer "shallow" NN. As the geometries of the cracks become more complicated, we study multiple layer NN and adapt the neurons in each layer and other NN parameters to achieve most accurate anomaly and interacting anomaly predictions.

References

[1]. Y. Yao, S.-T. E. Tung, and B. Glistic, “Crack detection and characterization techniques—An overview,” *Struct. Control Heal. Monit.*, 21, 1387–1413, 2011.

[2]. A. A. Carvalho, J. M. A. Rebello, M. P. V. Souza, L. V. S. Sagrilo, and S. D. Soares, “Reliability of non-destructive test techniques in the inspection of pipelines used in the oil industry,” *International Journal of Pressure Vessels and Piping*, 85, 745–751, 2008.

[3]. Jiaxing Ye et al., “Computerized Ultrasonic Imaging Inspection: From Shallow to Deep Learning”, *Sensors,* 18, 3820, 2018

[4]. Daniel Bacioiu et al., “Automated defect classification of Aluminium 5083 TIG welding using HDR camera and neural networks”, *Journal of Manufacturing Processes,* 45, 603–613, 2019

[5]. Min Meng et al., “Ultrasonic signal classification and imaging system for composite materials via deep convolutional neural networks”, *Neurocomputing,* 257, 128–135, 2017

[6]. S. Sambath et al., “Automatic Defect Classification in Ultrasonic NDT Using Artificial Intelligence”, *J Nondestruct Eval*, 30, 20–28, 2011

[7]. F.W. Margrave et al., “The use of neural networks in ultrasonic flaw detection”, *Measurement*, 25, 143–154, 1999

[8]. J. L. B. C. Veiga et al., “The Use of Artificial Neural Network in the Classification of Pulse-Echo and TOFD Ultra-Sonic Signals”, *J. of the Braz. Soc. of Mech. Sci. & Eng.*, Vol. XXVII, No. 4, 2005

[9]. Oscar Martin et al., “Artificial neural networks for quality control by ultrasonic testing in resistance spot welding”, *Journal of Materials Processing Technology,* 183, 226–233, 2007

[10]. Jing Liu et al., “Defect intelligent identification in resistance spot welding ultrasonic detection based on wavelet packet and neural network”, *Int J Adv Manufacturing Technol*, 90:2581–2588, 2017

[11]. Robert X. Gao and Ruqiang Yan, “Wavelets: Theory and Applications for Manufacturing”, *Springer Science+Business Media, LLC*, 2011

[12]. Peng Yang and Qiufeng Li, “Wavelet transform-based feature extraction for ultrasonic flaw signal classification”, *Neural Comput & Applic*, 24:817–826, 2014