 doi:10.20944/preprints202204.0034.v1

Ultrasonic System to Automatically Measure Size of Cracks using the Snooker Algorithm, a Simulator and Machine Learning

Dr Channa Nageswaran CEng FIMechE

TWI Ltd, Granta Park, Great Abington, Cambridge, CB21 6AL, UK

channa.nageswaran@twi.co.uk

Abstract

Cracking in a wide array of industrial components and structures pose a significant threat to their integrity. Detecting cracks using ultrasonic inspection techniques is a widespread activity for economic reasons but there are limitations to the techniques due to the morphology of cracks, such as fatigue cracks. In addition to detection there is a need to measure the size of the cracks which are often within the volume of the material. Ultrasonic techniques are well-suited to look inside the volume of the material but achieving sufficient sensitivity to the tip of the cracks in particular is practically difficult. Without an accurate knowledge of where the tip of the crack lies there can be significant uncertainty in sizing measurements. Machine Learning (ML) techniques are being developed to aid in the inspection and monitoring tasks but presenting the ultrasonic data in a suitable way for ML is very important. Following on from recent work presenting the development of the snooker algorithm to create images termed parameter-spaces, this paper presents how these images can be input into neural network based ML systems to automatically size these critical cracks.

Keywords

Fatigue, cracks, ultrasonic, inspection, snooker algorithm, machine learning, imaging, techniques

Introduction

The snooker algorithm was developed by Nageswaran[2021] to aid the effective use of ML for ultrasonic inspection applications, specifically for the detection and through-wall sizing of fatigue cracks, which is a difficult task for human operators. This difficulty arises from the need to detect a diffracted signal from the tip of the crack which can be extremely weak or be non-existent in the highest resolution reference images available to human operators to interpret (Nageswaran, 2021). Use of deep neural network algorithms, such as Convolutional Neural Networks (CNN), is difficult because they cannot function when signals are absent in the image, the presence of noise in the reference images used by human operators can degrade performance and very large amounts of datasets are required to train a CNN to a sufficiently performant level. In the inspection industry the criteria for detection performance is very high, often greater than 95% is required in critical industries such as aerospace and nuclear power generation. Sizing the detected flaws is deemed critical to the operation of many critical industrial components but it is known that human ultrasonic operators struggle with sizing crack-like flaws [Moran et al, 2010]. Hence, this paper reports on the ongoing development to utilise artificial intelligence (AI) technology in ultrasonic inspection systems.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 6 April 2022 doi:10.20944/preprints202204.0034.v1

Components

Snooker Algorithm

The snooker algorithm, as presented by Nageswaran (2021), is copied below:

1. Select the first element of the array as the cue. For each element in the array, including the transmitting cue, set it as the pocket.
2. Calculate the TOF from the cue to the pocket by reflection off the cushion. This calculation is based on the geometry, where the travel paths are straight lines when the material is isotropic, i.e. when the velocity of the sound wave remains constant in all directions.
3. In the FMC data matrix, using the A-scan for the cue-pocket (transmit-receive) pair, set a 'gate' over the expected arrival time of the wave. A gate is a concept where a signal within a defined period is captured or recorded for additional processing.
4. Using the gated signal, identify the maximum amplitude within the gate above a defined threshold. Selection of this threshold will depend on the noise level in the signal, but typically for the carbon steel material used for illustration in this paper the noise levels are low. Hence the threshold was set to 10% of the maximum possible signal level, so that all signal levels above this threshold will be considered valid for processing.
5. Record the maximum amplitude level of the signal in the gate, above the threshold, and store it in an n x n 2D matrix for each cue-pocket pair.
6. Repeat the process over all cue-pocket pairs. Once done, visualise the n x n 2D matrix, which is the parameter-space image for training the ML system.

The algorithm is analogised with the game of snooker and is relevant in the context of ultrasonic wave propagation, leaving open many options for it to be used in future AI/ML work. The reader is referred to Nageswaran (2021) for definition of terms such as TOF (time-of-flight), transducer array and its elements, and the FMC data matrix. The signal collected in the gate can be processed to get its maximum amplitude, TOF and its frequency content using fast fourier transform (FFT) algorithms. In this paper we make use of the absolute maximum amplitude to yield the parameter-space images, whereas in Nageswaran (2021) the amplitude above the threshold was convolved with the TOF of the gated signal to illustrate differences between the reference images and the parameter-space images.

Parameter-Space Images

Figure 1 shows the parameter-space image for a fatigue crack in an approximately 12.5mm thick block and the corresponding high resolution total focusing method (TFM) image used by an operator.

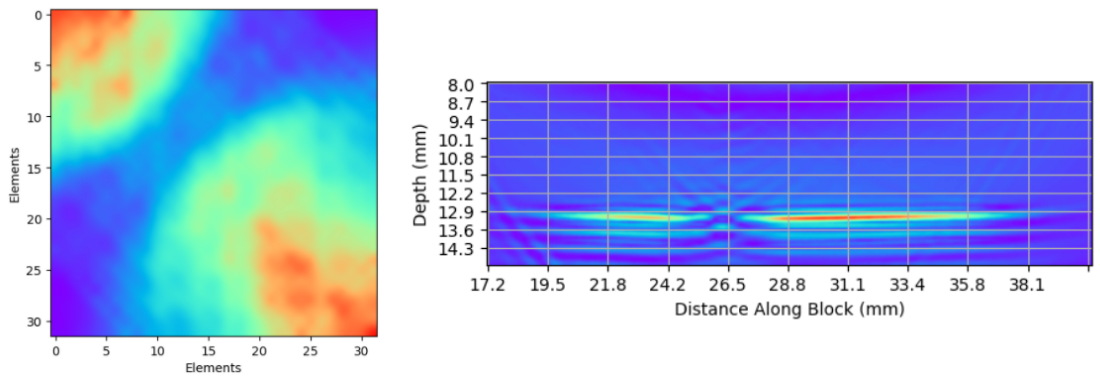

Figure 1 The absolute amplitude parameter-space image (left) and the corresponding TFM image, using data from a 5MHz 32 element transducer array, for a fatigue crack in a carbon thick steel block.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 6 April 2022 doi:10.20944/preprints202204.0034.v1

It is possible to see from the TFM image that a crack may exist by the signal at the root, but no signal is resolvable from its body or its tip, which often leads to operators underestimating sizes of cracking. The transducer frequency was chosen to be 5MHz with 32 elements as it has less sensitivity to diffracted signals of crack tips - ie making it difficult for an operator - but is low enough for practical use for inspection, as opposed to the higher 10MHz 64 element transducer array used by Nageswaran (2021), which is more sensitive to the tip diffracted echo but is also more sensitive to surface conditions of typical engineering components making its practical use in inspections difficult.

Both axes of the parameter-space image is n, where n is the number of elements (here 32). The information in the parameter space image is related to the presence (or not) of a crack. The work reported in this paper applied the technique to fatigue cracking, which is considered a smooth reflector, but other forms of rough cracking (e.g. stress corrosion cracking) yield similar images. In this work the aim was to size the through-wall extent of the cracking but other features of the cracking can also be established following a similar approach, which will be reported in future works.

On a parameter-space image, there are a number of parameters that can be used to develop the ML system for automatic sizing. Figure 2 illustrates these parameters, which can all be measured from the image using standard processing techniques and tabulated. This is essentially a feature engineering step for the ML system, where these parameters are synonymous with features that correlate with cracking damage, because the snooker algorithm is only sensitive to cracking. This is the key insight which enables the use of both well-established ML algorithms as well as neural network approaches, since some of these parameters can be placed on a standard scale: from 0 to 1.

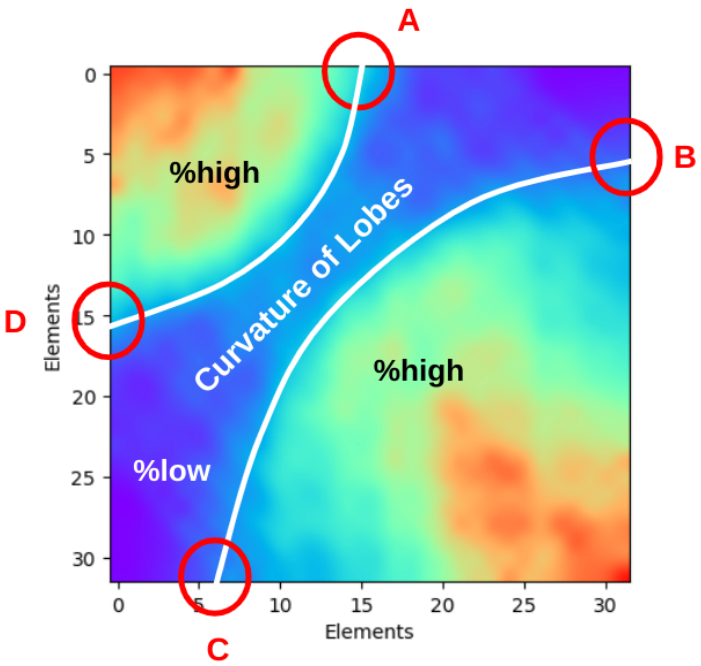

Figure 2 The parameters, or features, on a parameter-space image that are extracted from it using image processing techniques:

- A, B, C and D are the intercept of the characteristic lobes with the axes;
- % of area above a threshold and % of area below, termed %high and %low;
- curvature of the lobe fronts represented as coefficients of the polynomial of best fit.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 6 April 2022 doi:10.20944/preprints202204.0034.v1

The Simulator

Finding data from real cracks for training is difficult in the inspection industry for several reasons, therefore the use of approaches that rely on large amounts of data will have onerous requirements. Hence, the primary motivation for the approaches under development is to reduce this burden and so data augmentation strategies are recognised throughout the AI/ML community to be an important area of work, and viable approaches are being developed for ultrasonic testing (Virkkunen et al, 2019). The simulator is a numerical model developed in Python with the aim of playing the game of snooker in the component and has been conceived for use in a number of different approaches being investigated. Figure 3 shows a point in time during a simulation with the presence of a planar crack positioned under the array at a tilt, at a time when the propagating wave reflects from the crack. The code is designed to be parallel such that on an 8-core CPU a full simulation event leading to the generation of an FMC data matrix can be completed in less than 2 minutes for the cases in this paper. With these durations the simulator was integrated into the system architecture for use by the engine. Using the results of simulations, both TFM and parameter-space images can be generated for developing the model; Figure 4 shows example images for the simulation case presented in Figure 3.

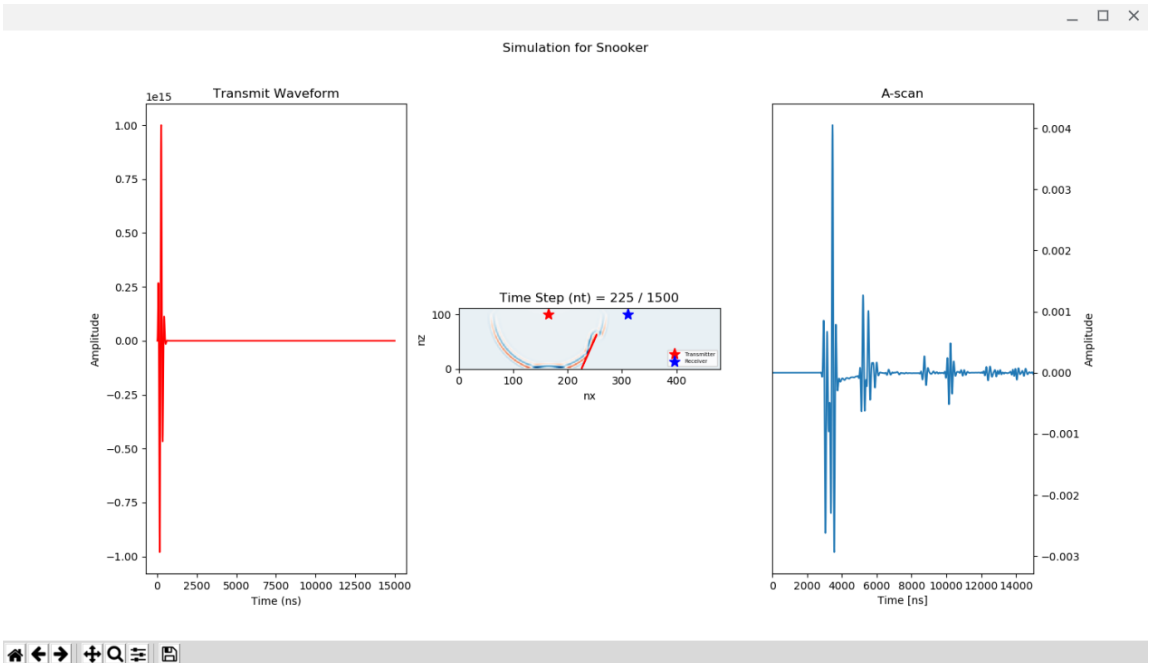

Figure 3 Image showing a crack tilted at 3° to the vertical and with a through-wall size of 7.5mm in the simulator with a propagating wavefront at a time just after it was incident onto the crack surface.

Simulation images are ‘cleaner’ than experimental images, as can be seen when comparing the parameter-space images of Figure 1 from an experimental data set to the parameter-space image of Figure 4. However, the images have strong similarities and by processing the experimental image with appropriate thresholds it is possible to ‘clean’ them to a sufficient level to clearly extract the parameters of interest, as shown in Figure 2. The application of tilt has the effect of changing the curvature of the lobes, the boundary between %low valley and the %high plateau defined in Figure 2. The empirical evidence in the simulation database shows the curvature to be a powerful indicator of the tilt of the flaw but this will not be discussed in this paper, as presenting the polynomial coefficients of the curvature on a standard scale to the neural network model at the same time as the other parameters (which are easy to express on a standard scale) leads to instability in the prediction. Therefore, tilt is handled separately by the engine and will be presented in a future paper.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 6 April 2022 doi:10.20944/preprints202204.0034.v1

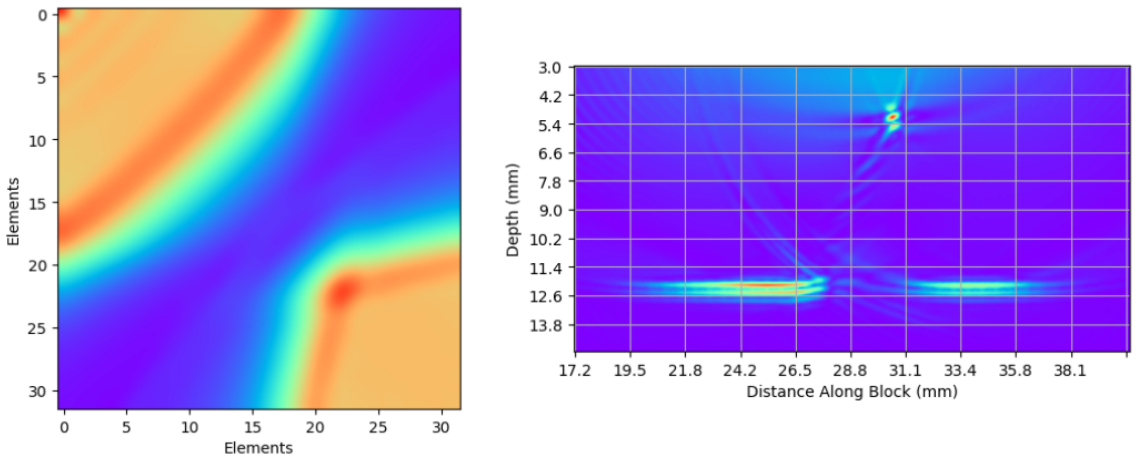

Figure 4 Parameter-space and corresponding TFM images generated from the simulation in Figure 3.

The Neural Network Model and the Engine

The scikit-learn framework (Pedregosa et al, 2011) was selected for development of the neural network architecture in this work. The reason for this selection is the uniformity of the Python libraries which allows for rapid investigation of standard ML algorithms as well as neural networks. As the goal of the work was to provide a measurement of size in mm the underlying concepts are related to regression. Figure 5 presents the system architecture with the key functional components.

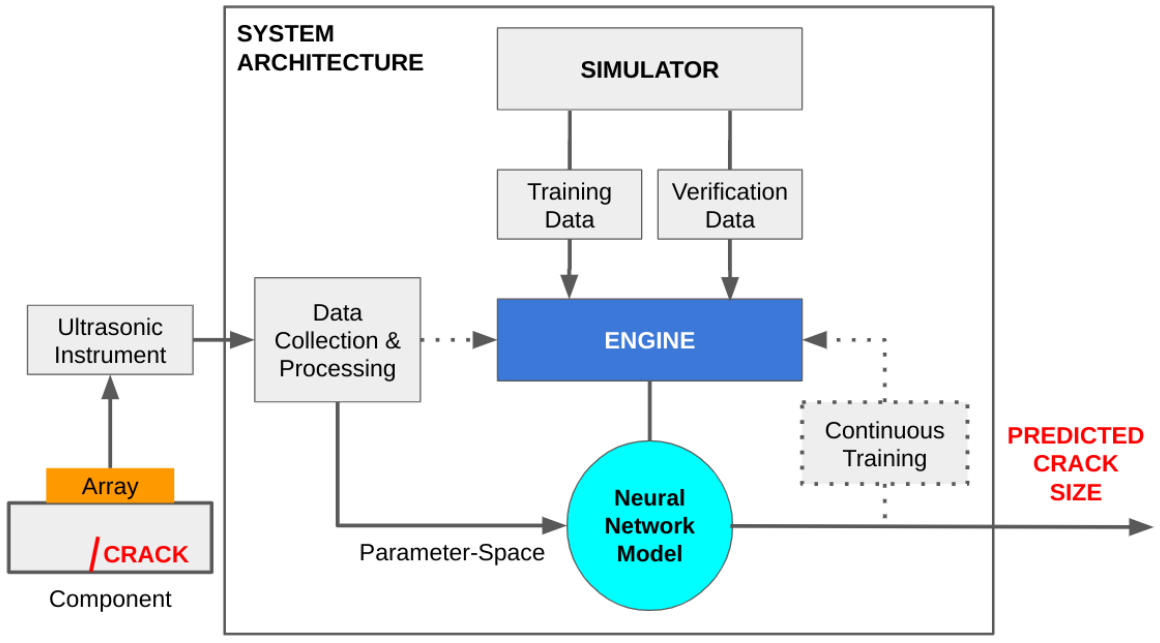

Figure 5 Architecture of the system, showing the key components which are all expressed in the Python language, except for components interfacing with ultrasonic hardware which are written in C.

The engine is a complex set of routines that aim to adapt the neural network model, which is based on MLPRegressor of scikit-learn's neural network module. The simulator is the primary source used by the engine to tune the model, but this is tempered by real data and by a continuous training feedback route. The engine routines use these datastreams to prevent overfitting and to aid in tuning the hyperparameters when the model is being configured for a particular application scenario. The functional aspects of the engine will be reported in a future publication once fully validated by trials.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 6 April 2022 doi:10.20944/preprints202204.0034.v1

Results and Discussion

For the case presented in this paper the simulator was used to generate 5852 cases for the engine to train the neural network model. The simulator was used to provide a random set of additional cases for verification of the model, ensuring that the verification models did not overlap with the set generated for training. The engine was able to randomly request a batch of cases from the simulator. A library of simulation data was created as a Python list data structure, defining the position, orientation and size of the flaw, its parameter-space and the TFM images as 2D NumPy arrays. This data structure was suitably small in its memory footprint to be held and rapidly processed. The raw simulation results, from which the training data structure was created, occupied 70 GB of disk space.

An image processing routine was used to extract from each parameter-space image the A, B, C, D, %high, %low and (maximum amplitude - minimum amplitude), yielding seven features for input to the neural network model. Each case had three labels: root position of the crack, its through-wall height (size) and its tilt. Note that tilt has a relationship with A, B, C and D, but a stronger relationship to the curvature of the lobes, and so the engine has another type of model to estimate tilt. Similarly the root position of the crack is not discussed in this paper as in a practical procedure for inspection the operator or the automated scanning system will move over the crack and so by sampling over several data sets an average position for the crack can be extracted accurately. Therefore, the neural network model discussed in this paper is configured for prediction of the through-wall size of cracks.

The engine configured the neural network model by tuning the hyperparameters of MLPRegressor and parameters related to the image processing routine to extract the features, in particular the amplitude threshold between the high and low plateaus separated by the curved lobes on the parameter-space image (see Figure 2). This parameter is particularly important for real data as the boundary between these regions can be difficult to distinguish - as can be observed on the parameter-space image of the real data shown in Figure 2 compared to the simulated example of Figure 4. A routine was developed as part of the continuous training feedback loop to the engine, as shown in Figure 5, to optimise the neural network model by tweaking this imaging threshold. These novel aspects will be reported on once the routine itself is optimised during the validation programme using real crack data (see below), which will allow for exploration of continuous training.

The criteria for optimal performance during model verification was set on the error in predicted through-wall size from the simulated size to be within 0.25mm. Training and verification were done using simulated data with the goal to tune the model. The range of the labels in training dataset was:

a. Root position of the cracks varied from 5 to 95% of the length under the array;
b. Through-wall size of the cracks varied from 1 to 11.5mm (block thickness was 12mm);
c. Tilt of the cracks varied from -3 to +3°.

These values were taken from observation of actual fatigue cracks from a starter notch (typically 1mm in size), as shown in Figure 2 of Nageswaran (2021), and have been selected with a view to the sample set being created for the upcoming validation programme. Five fatigue cracks were available whose through-wall sizes could be measured by a combination of visual assessment on the edges and by combining several ultrasonic imaging approaches to measure the size in the measurement plane of the transducer within the volume of the samples. Hence there was some confidence in their through-wall size. To enlarge the testing dataset three separate readings were taken from each of the cracks with them placed at different positions under the array (changing their root position). The most accurate measurement of the crack sizes can only be done by destructive examination. The validation programme will use different methods to measure crack size while generating the samples.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 6 April 2022 doi:10.20944/preprints202204.0034.v1

Once processed the parameter-space dataset could be explored within the simulation library and several key results are presented in this paper. Figure 6 shows the relationship between the (max - min) amplitude feature from the parameter-space image and the crack size for a crack growing at a root position below the centre of the array. The different tilt values are included and it can be seen that there is a greater spread in feature values at lower crack sizes. Figure 7 shows the same plot but for the crack growing under one edge of the array, which shows that significant spread appears as the crack size increases for different tilt values. This can be explained from an ultrasonic perspective.

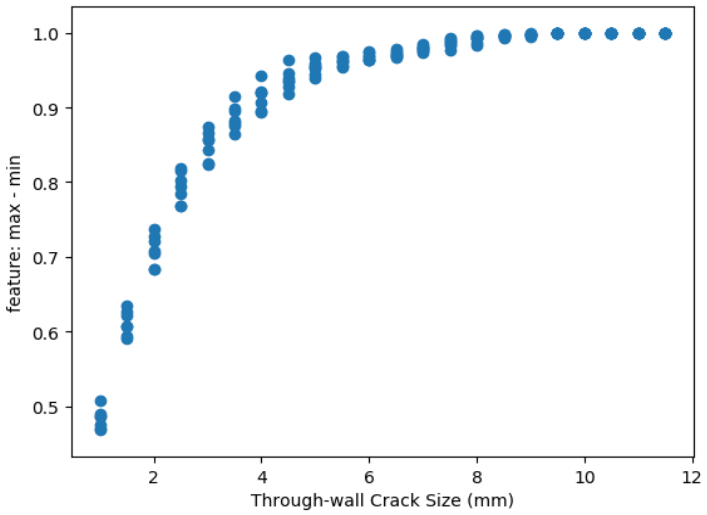

Figure 6 Plot showing relationship between the feature (max - min) amplitude of the parameter-space image and the through-wall crack size for a crack growing under the middle of the array, with the scatter at different through-wall size values due to different tilt values of the cracking.

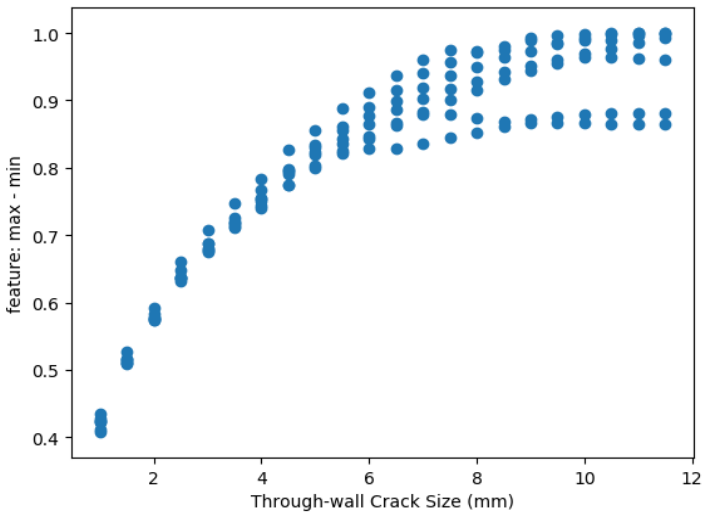

Figure 7 Plot showing relationship between the feature (max - min) amplitude of the parameter-space image and the through-wall crack size for a crack growing under the edge of the array, with the scatter at different through-wall size values due to different tilt values of the cracking.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 6 April 2022 doi:10.20944/preprints202204.0034.v1

Figure 8 shows the relationship between the percentage of the parameter-space image represented by the high amplitude plateau (%high) and the crack size for a crack growing under the middle of the array. There is again a trend and the scatter on values due to different tilts of the cracking, but an interesting aspect, which is later observed through the neural network model, is the plateau for cracks smaller than 2mm in through-wall size. In trials the ability of the system to make accurate predictions of the crack size for small cracking close to a mm in size was found to be poor. Analysis of the parameter-space library indicates that changes in values of many features are small for small cracks. Figure 9 shows the same scenario but for cracks growing under the edge of the array, illustrating large scatter induced by tilt. Hence these aspects necessitate a need for a neural network.

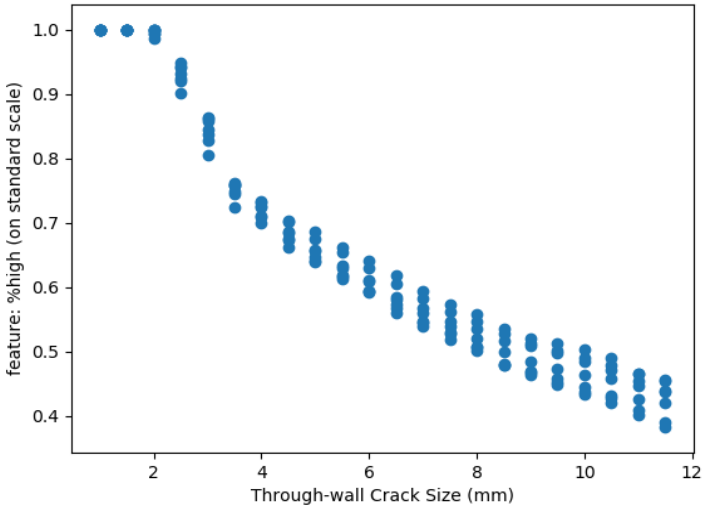

Figure 8 Plot showing relationship between the feature %high amplitude (on standard scale) of the parameter-space image and the through-wall crack size for a crack growing under the middle of the array, with the scatter at different through-wall size values due to different tilt values of the cracking.

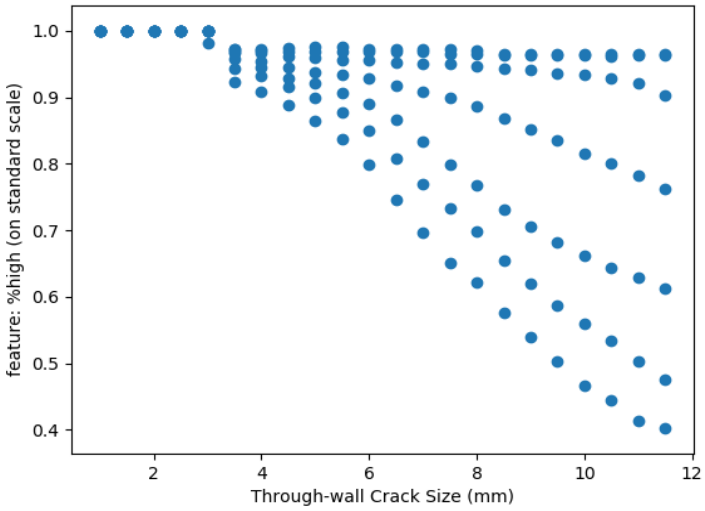

Figure 9 Plot showing relationship between the feature %high amplitude (on standard scale) of the parameter-space image and the through-wall crack size for a crack growing under the edge of the array, with the scatter at different through-wall size values due to different tilt values of the cracking.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 6 April 2022 doi:10.20944/preprints202204.0034.v1

Much larger variation for the different features can be observed when the tilt value is held at a specific value, for example -3.0°. Figure 10 shows the variation in feature D (an intercept of a lobe with a parameter-space axis) for varying crack sizes for all positions of the crack root under the array. Any relationship is not easy to establish visually in this case and once all cases are introduced a very rich space is revealed, including features with relationships to each other as shown in Figure 11, for cracks growing under the middle of the array, between the %low amplitude and intercept C features. These results pointed towards the use of a neural network model which could build the connections between all these feature-label and feature-feature relationships to create a predictive system for measurement of the crack size, which would necessarily be an approximate approach, but one that could be more accurate than the performance of human operators using procedures on TFM images.

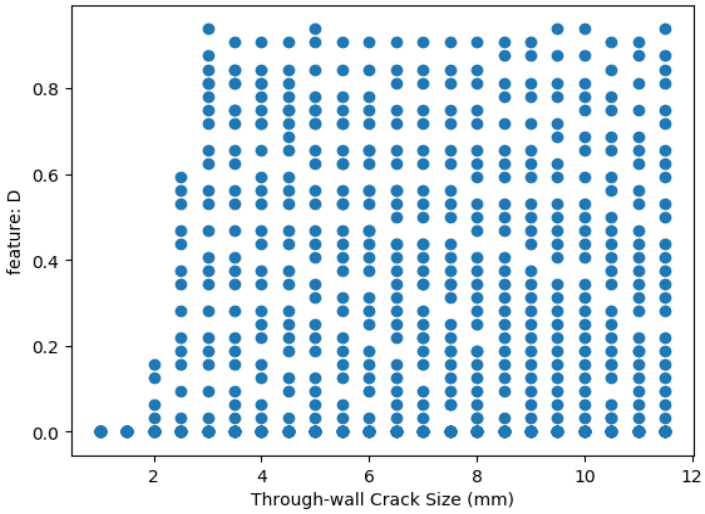

Figure 10 Plot showing relationship between feature D (axis intercept on the parameter-space image) and the through-wall crack size for cracks, growing in all root positions below the array with -3.0° tilt.

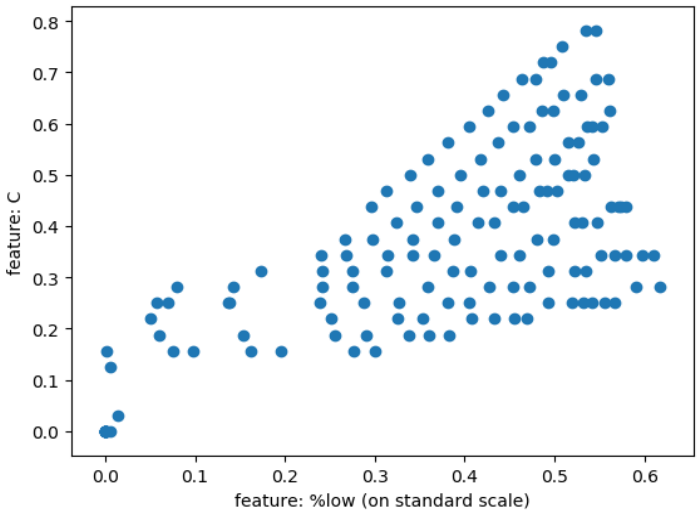

Figure 11 Plot showing relationship between two features, %low of amplitude in the valley of the parameter-space image (on standard scale) and the parameter-space image axis intercept feature C.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 6 April 2022 doi:10.20944/preprints202204.0034.v1

The model setup involves creation, at which time certain hyperparameters related to the neural network model have to be selected; in particular, the number of neurons in a layer, the number of layers, activation function type, regularisation, learning rate and the optimisation algorithm. There are several other hyperparameters related to the MLPRegressor implementation of a neural network that have an influence on the quality of the model. The work presented in this paper is restricted to exploration of these standard parameters which appear to provide good results but ongoing exploratory work indicates that fundamental changes to the neural network architecture itself can provide significant improvements in performance. These changes are inspired by the domain specific aspects of ultrasonic inspection as expressed through the snooker algorithm and the parameter-space images, as well as modern theoretical developments in the field of AI/ML. Future paper will present details of the modified neural network model and the engine under development.

The engine contains routines that iterate through selected hyperparameters while checking performance on the verification data set to confirm that it is within tolerance. Figure 12 shows a typical output from the script during the setup process. As noted earlier the tolerance is for the error between the predicted and actual values to be within 0.25mm for through-wall size (Flaw Heights in Figure 12) and it can be seen that some errors are significantly larger than this value. In particular note the two cases with a Flaw Height of 11.80mm and Root Position of 17.30, but differing tilts of 3.4° and -2.0°; the error in the prediction is significantly above tolerance at 5.44mm in one case and 1.36mm in the other. This illustrates the importance of tilt, especially when the through-wall size approaches the scanning surface - in this case the crack was 11.80mm in through-wall size, only 0.2mm from the top scanning surface of the block. It is well known that specular reflections from smooth targets are governed by very tight tolerances and this domain specific knowledge is also manifested in the neural network model. Hence, as noted earlier, tilt is handled in a separate system that then adapts the inputs of the neural network model using the curvature of the lobes in Figure 2.

Figure 12 also illustrates a well-known aspect of AI/ML systems in that it is very difficult to ‘explain’ how the output was arrived at - they are considered ‘black boxes’. Indeed it is complex to visualise the flow from input to the output, but there is causality based on the weights and the activation functions, even though the latter can be non-linear. Nevertheless, the need to explain is paramount in many scenarios and also in the inspection industry. Many sectors which are likely to benefit significantly from this emerging technology, such as the aerospace and nuclear industries, are highly regulated. Stringent qualification programmes are run to ensure that any inspection system deployed on critical components are fit for purpose and that their performance, to include that of the operators, is quantified to high standards. Independence in the assessments is a key component of these qualification programmes and these same principles will need to be applied for AI/ML systems.

The European Network for Inspection and Qualification (ENIQ) published a recommended practice (ENIQ, 2021) outlining the key principles in accordance with already well-established practice in the nuclear industry. In this framework, AI/ML is considered ‘software’ with a specific focus on ‘data analysis’ to be done in partnership with human operators or, in the extreme of four possible scenarios, to completely replace the human operator. The document highlights many of the unique aspects for the ML software, such as the need to prevent overfitting to the training data, the need to version the software to make it immutable and prevent fitting to ‘blind samples’ during qualification. Another important aspect is to prevent the operators from becoming overly reliant on the software because, as illustrated in this paper, error is an inherent aspect of AI/ML systems and there remains the danger that the output of a system - however well designed, conditioned and qualified - can be wrong, with the potential to lead to adverse outcomes for industrial components and system. However, the motivations for utilising AI/ML in industry are strong, and they will have a valuable role.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 6 April 2022 doi:10.20944/preprints202204.0034.v1

```
AI Demo Program
Fatigue Cracks in 12mm Plates
Version: 0.4

VERIFICATION
------------

Root Positions:

Prediction   5.34              Actual   7.80          Error   2.46
Prediction  16.88              Actual  17.30          Error   0.42
Prediction   9.92              Actual   9.30          Error   0.62
Prediction   3.56              Actual   3.70          Error   0.14
Prediction  11.43              Actual   9.40          Error   2.03
Prediction   6.17              Actual   8.20          Error   2.03
Prediction   8.92              Actual   7.20          Error   1.72
Prediction   5.89              Actual   7.80          Error   1.91
Prediction  17.19              Actual  17.30          Error   0.11
Prediction   9.93              Actual   9.30          Error   0.63
Prediction   2.53              Actual   3.70          Error   1.17
Prediction   6.92              Actual   9.40          Error   2.48
Prediction   6.29              Actual   8.20          Error   1.91
Prediction   5.79              Actual   7.20          Error   1.41

Flaw Heights:

Prediction   4.12              Actual   4.10          Error   0.02
Prediction   6.36              Actual  11.80          Error   5.44
Prediction   1.92              Actual   1.30          Error   0.62
Prediction   2.35              Actual   2.90          Error   0.55
Prediction   8.48              Actual   9.10          Error   0.62
Prediction   9.76              Actual  10.40          Error   0.64
Prediction   7.58              Actual   7.80          Error   0.22
Prediction   3.11              Actual   4.10          Error   0.99
Prediction  10.44              Actual  11.80          Error   1.36
Prediction   1.91              Actual   1.30          Error   0.61
Prediction   3.09              Actual   2.90          Error   0.19
Prediction   8.86              Actual   9.10          Error   0.24
Prediction   9.74              Actual  10.40          Error   0.66
Prediction   7.65              Actual   7.80          Error   0.15

Flaw Tilts:

Predition  -0.45               Actual  -2.30          Error   1.85
Predition   0.65               Actual   3.40          Error   2.75
Predition   0.15               Actual  -1.60          Error   1.75
Predition  -0.27               Actual   1.90          Error   2.17
Predition   0.25               Actual   2.70          Error   2.45
Predition  -0.01               Actual  -1.90          Error   1.89
Predition   0.22               Actual   2.60          Error   2.38
Predition  -0.22               Actual  -2.00          Error   1.78
Predition  -0.75               Actual  -2.00          Error   1.25
Predition   0.15               Actual  -2.00          Error   2.15
Predition  -0.48               Actual  -2.00          Error   1.52
```

Figure 12 Screenshot of a routine in the process of setting up using simulated verification data.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 6 April 2022 doi:10.20944/preprints202204.0034.v1

Figure 13 shows the root mean square error (RMSE) of the predicted through-wall size on the verification dataset for the total number of neurons in the model. Figure 14 shows the RMSE when separated into the number of neurons in each layer and number of layers. The results (for a model in this case with a hyperbolic tangent activation function and Adam as the optimisation algorithm (Kingma & Ba, 2015)) indicate that with increasing number of neurons the error reduces (Figure 13) and that the depth of the model is more important than the number of neurons per layer (Figure 14).

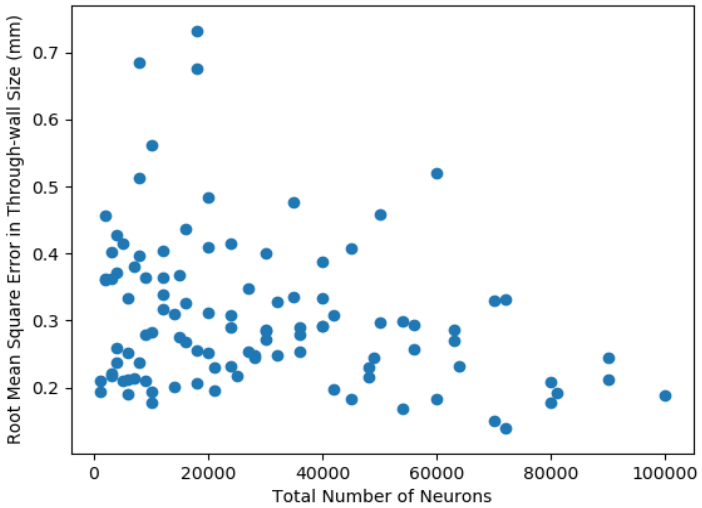

Figure 13 Plot showing the reduction in RMSE as the total number of neurons in the model increases.

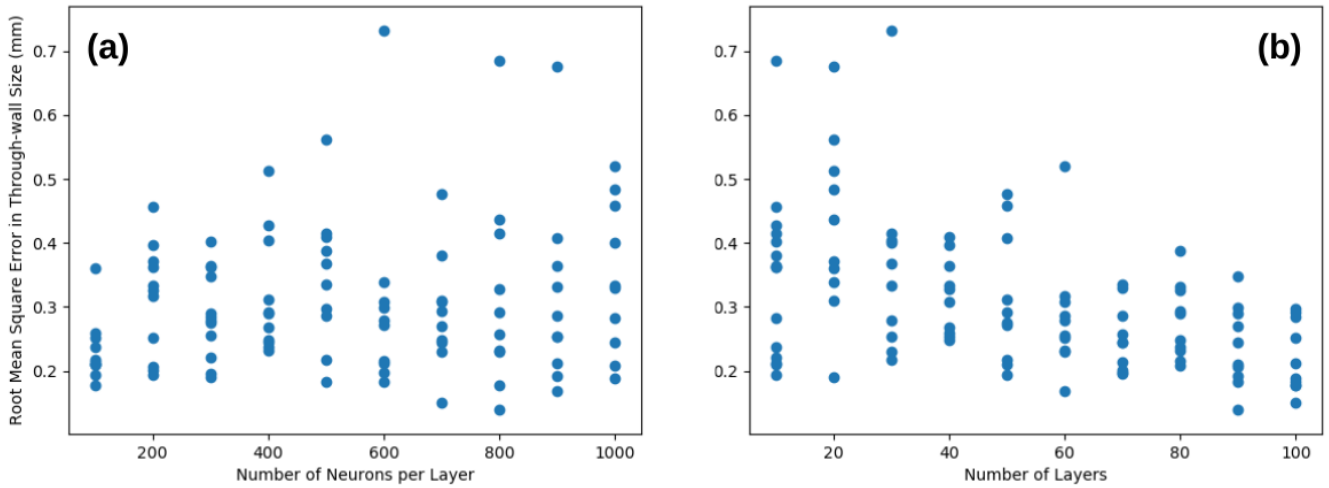

Figure 14 Plots showing influence on RMSE by (a) number of neurons/layer and (b) number of layers.

At this stage of the development, the model is well-conditioned at approximately 100,000 neurons with a consistent RMSE in through-wall sizing less than 0.25mm, based on a randomly generated set of verification data from the simulator which do not exist in the simulated training data set. A limited study on actual cracks indicates that RMSE errors in through-wall sizing is around 1.5mm but there is insufficient test data for the validation to be reliable. Furthermore, the evidence indicates that the largest influence on the outcome now is not the model but the processing of the parameter-space image to extract the features. The inherent noise in real data from real cracking will pose a challenge.

 doi:10.20944/preprints202204.0034.v1

Conclusions

1. The parameter-space images generated by the snooker algorithm were defined and shown to be viable as feature engineering for the development of a machine learning system capable of predicting the through-wall size of cracking. This is applicable to many cases in isotropic material.

2. The model was set up using a numerical simulator through a process of iterative training and verification to ensure the neural network model was well-conditioned. A set of routines were used to set hyperparameters of the model to achieve acceptable tolerances on the sizing output.

3. Trials on a limited number of actual cracks indicate that sizing accuracy can be within 1.5mm of the best endeavours measurement of a human operator - the actual size cannot be accurately established until these specimens are sectioned. **However, if the indicative 1.5mm error is achieved in validation, the system will offer significant advantages, as it can automatically analyse large amounts of data and report results in a fraction of the time required by a human.**

4. The next stage of work is to validate the system using a large independent (or blind set) of samples following robust principles as outlined in industry documents, such as in the ENIQ RP 13.

5. In further work, convolutional neural networks (CNNs) will be investigated using suitable ML frameworks such as Google's TensorFlow and appropriate hardware. However, the architecture presented in this paper is elegant and ongoing work indicates that it may be better suited for 'explainability', a key requirement for acceptance within the highly regulated inspection industry.

6. In addition to use for data analysis, which is a major theme for ongoing work in industry, the system presented in this paper can be used by monitoring systems to automatically track known defects and provide continuous information on their growth, for example stress corrosion cracks.

7. Robotic systems working in harsh and remote environments where humans cannot operate can use this system for inspection of critical equipment and make maintenance decisions confidently.

References

ENIQ. **Recommended Practice 13: Qualification of Non-Destructive Testing Systems that Make Use of Machine Learning**, Issue 1, European Network for Inspection and Qualification, Report No. 65, 2021, https://snetp.eu/wp-content/uploads/2021/06/ENIQ_RP13_Issue1.pdf

Kingma D P and Ba J L. **Adam: A Method for Stochastic Optimization**, arXiv, 2017, https://arxiv.org/pdf/1412.6980.pdf

Moran T L, Ramuhalli P, Pardini A F, Anderson M T and Doctor S R. **Replacement of Radiography with Ultrasonics for the Nondestructive Inspection of Welds - Evaluation of Technical Gaps - An Interim Report**, Pacific Northwest National Laboratory, Prepared for US Nuclear Regulatory Commission, 2010, https://www.nrc.gov/docs/ML1010/ML101031254.pdf

Nageswaran C. **The Snooker Algorithm for Ultrasonic Imaging of Fatigue Cracks in order to use Parameter-Spaces to Aid Machine Learning**, Preprints, 2021, doi:10.20944/preprints202107.0269.v1

Pedregosa et al. Scikit-learn: Machine Learning in Python, JMLR 12, pp. 2825-2830, 2011.

Virkkunen I, Koskinen T, Jessen-Juhler O and Rinta-aho J. **Augmented Ultrasonic Data for Machine Learning**, arXiv, 2019, https://doi.org/10.48550/arXiv.1903.11399