Fatigue Damage Detection and Risk Assessment via Wavelet Transform and Neural Network Analysis of Ultrasonic Signals

Hassan Alqahtani[1,2], Asok Ray[3,4]
[1] Department of Mechanical Engineering, Taibah University, Medina, KSA, 42353.
[2] School of Electrical Engineering and Computer Science, Pennsylvania State University, University Park, PA 16802
[3] Department of Mechanical Engineering, Pennsylvania State University, University Park, PA 16802
[4]Department of Mathematics, Pennsylvania State University, University Park, PA 16802.

Abstract

This paper develops a data-driven autonomous method for detection of fatigue damage and classification of the associated damage risk in mechanical structures, based on ultrasonic signal energy. The underlying concept is built upon attenuation of the signal and stability of the attenuation process. The attenuation provides pertinent information for damage quantification, whereas the stability represents resistance towards the fatigue damage growth. The proposed neural network (NN) model has been trained using the scaled conjugate-gradient back-propagation method. The NN model is capable of damage detection and damage classification into five classes of increasing risk. The Daubechies wavelet transform has been used to reduce the noisy pattern of the ultrasonic signal energy by using the associated approximation coefficients. The results show that the proposed method of approximation signal energy can detect and classify the damage with an accuracy of up to $\sim 98.5\%$.

Keywords: Fatigue damage detection, Risk classification, Signal energy, Daubechies wavelet transform, Neural networks

1. Introduction

Structural integrity of large-scale engineering systems (e.g, power plants, chemical processing plants, and transportation systems) deteriorates over time due to various types of degradation, which may cause irreversible damage in critical structures. According to Farrar and Wonden [1], the damage is the cumulative effect of changes that, upon initiation in a system, potentially degrade the system performance; this damage may lead to a failure unless appropriate timely actions are taken. Such damage in structural materials can evolve either at a micro-scale level from inherent local and global defects (e.g., voids and inclusions), or at a macro-scale level (e.g., corrosion and cracks). Of all types of structural damage, perhaps the most critical [2] is the fatigue damage that is caused by fluctuating stresses, which can be well below the respective yield points. It is well known [2, 3] that $\sim 90\%$ of the structural failures occur due to fatigue; in general, fatigue failures evolve through three stages: defect, damage, and fracture.

Two essential factors for mitigation of fatigue damage are damage-tolerant design and nondestructive inspection. An example of damage tolerance is the ability to resist the onset of fracture in material from pre-existent flaws. The methodology of the damage tolerance is vital especially for components whose failure could lead to a catastrophic loss of life or property. On the other hand, nondestructive inspections estimate the service life and/or inspection intervals under defined loading conditions and service schedule and type. Nondestructive inspection evaluates structural integrity by detecting damaged components that must be repaired or replaced. Because of the limitation of using inspection techniques and the difficulties in locating/detecting internal defects, the damage-tolerant design provides additional protection against premature fracture by incorporating the available information on structural design and material properties such that subcritical crack growth can be controlled and the final fracture can be avoided. Therefore, it is important to incorporate methods of nondestructive inspection and damage tolerance in the maintenance process to

ensure the operational reliability of machinery components.[4, 5]

The nondestructive evaluation (NDE) techniques that are widely used in practice are visual testing, ultrasonic testing, eddy current, magnetic particle, radiography, and dye penetrant [6]. In this paper, two NDE techniques have been used to study fatigue crack detection and growth. The first one is visual testing which is the most common technique for detecting surface damage, particularly when employed with lighting aids and magnification. The second one is ultrasonic testing which uses high-frequency sound waves to penetrate deep inside the structure to detect internal flaws and surface cracks.

Investigation of the reasons for damage resistance is very important for improving the damage-tolerant strategy. There is a large volume of open-source literature describing the role of the crack-tip plastic zone as one of the causes for material resistance against the crack driving force. During the crack propagation in polycrystalline alloys (e.g. aluminum), locally plastic deformation occurs around the area of a stressed crack tip. Thus, the fatigue damage of a cracked structure is significantly influenced by the size and shape of the plastically deformed zone. It is well known that the toughness of the material plays a major role in the plastic zone size, where the plastic zone size at the crack tip increases as the toughness of the material increases [7, 8].

The crack propagation is significantly influenced by the plasticity-induced crack closure mechanism. For the plane stress case, the volume elements in the plastic zone are extended and primarily balanced by an out-of-the-plane flow of the material such that, the thickness of the plastic zone is reduced. Generally, the plasticity-induced crack closure under plane stress case can be described as a result of an additional material layer behind the crack tip, which can be viewed as a wedge that is inserted in the crack path, where it decreases the deformation of the cyclic plastic at the crack tip and consequently the fatigue crack growth rate [9, 10].

This paper introduces a novel methodology for the damage-tolerant design that largely relies on the detection of the fatigue damage and classification of the damage risk. This methodology consists of two different approaches. The

first approach relies on the responses of ultrasonic testing (UT) signals, which estimates the energy of each UT signal to provide vital information about the crack onset and the crack size. The signal attenuation leads to crack detection because part of the signal is reflected due to the creation of the new crack surfaces. On the other hand, the stability of the attenuated ultrasonic signal refers to the resistance of the crack growth. The size of fatigue damage is assumed to be small when the signal attenuation is small and the stability period of the signal attenuation is long. The reason for this long period is the resistance of the crack growth, where the attenuated signals that are received during this period are similar. Therefore, the phenomenon of crack propagation can be illustrated using the energy of UT signals. This paper proves that as fatigue damage progresses, the signal attenuation becomes higher and the stability period of the attenuation decreases.The second approach makes use of a neural network to build an automated system that can detect the onset of fatigue damage and classify the risk based on the amount of the signal energy and the transition of this amount (the period of the signal stability).

In recent years, considerable developments in the field of computational power and the development of powerful statistical learning algorithms have led to a renewed interest in machine learning (ML) techniques in a variety of engineering applications. One potential advantage of using ML techniques is to predict the interdependencies and relationships of complex problems which are very difficult to apply using conventional modeling techniques. Many independent studies have shown the importance of ML techniques to identify and classify structural crack severity.Ljubov and Helle [11] used a deep learning neural network and the random forest method to detect the depth or location of cracks. Fang et al. [12] found that the neural network using frequency response functions (FRFs) as input data can evaluate damage conditions with very good accuracy in the cantilever. Alqahtani and Ray [13] employed convolutional neural networks (CNN) for crack damage detection in polycrystalline alloys, their CNN model successfully detects and classifies the severity of the damage into three classes. Rageh [14] developed a framework of automated damage detection

to identify stiffness degradation that results from the fatigue cracks propagation in stringer-to-floor beam connection of the riveted steel railway bridges. Pandey and Barai [3] applied different ANN architectures to identifying damage in a 21-bar bridge truss. Wu et al. [15] investigated the use of a neural network (NN) to identify member damage in a 3-story building frame. Zapico [16] studied the damage assessment of steel structures using the ANN method. Most of the researches has shown that NNs provide a good assessment for damage identification, generally when the generated data are simulated numerically and are error-free.

The major contributions of this paper are delineated as follows:-

1. *Fatigue damage detection & classification*: This research topic has been investigated by using NDE methods such as those based on ultrasonic testing (UT) and optical metrology (OP), where the ultrasonic signal energy has been used to classify the risk of the damage into five classes, based on the signal attenuation and the stability of the signal attenuation.
2. *Performance enhancement by NN-based pattern classification*: The performance of the signal energy-based model has been improved by applying the wavelet transform technique of Daubechies [17], where the approximation coefficients of the signal energy are used as an input data for the model instead of the original signal energy data.

The paper is organized into five sections. Section 2 describes the laboratory apparatus to validate the NDT in real time. Section 3 illustrates the methodology adopted in this paper, including an overview of discrete wavelet transform and neural networks, which consists of the following four subsections: the concept of the signal energy, an overview of discrete wavelet transform, a description of the proposed classification methodology for damage-tolerant design, and an overview of neural networks. Section 4 discusses the outputs of the proposed damage-tolerant model. Section 5 summarizes and concludes the paper with recommendations for future research.

2. Description of the Experimental Apparatus

This section describes the laboratory apparatus, as depicted in Figure 1, which is built upon a computer-instrumented and computer-controlled fatigue testing machine[1], equipped with ultrasonic testing (UT) probes[2], and a confocal microscope[3], The objective of this investigation is to infer conclusions from a synergistic combination of the heterogeneous measurement data, generated from ultrasonic sensors and optical images (of the Alicona confocal microscope) to improve the performance of the damage-tolerant design of mechanical structures. It is well known that a large part of the service life of ductile alloy structures under medium to high-cycle fatigue is consumed in crack initiation stage [18]. Therefore, detecting the crack initiation helps to maintain the performance of the structure and avoid the economic consequences of the structure failure, where increasing the reliability of the structure and decreasing the maintenance cost is the overriding goal.

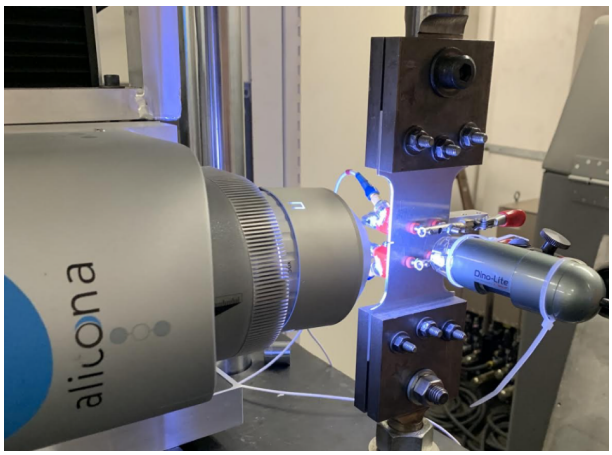


Figure 1: Picture of the experimental apparatus

[1]Manufacturer: MTS Systems Corporation, Berlin, NJ, USA
[2]Manufacturer: OLYMPUS, Shinjuku, Tokyo, Japan
[3]Manufacturer: Alicona Imaging GmbH, Dr.-Auner-Strasse 21a, 8074 Raaba/Graz, Austria

All tested experiments were conducted in the laboratory environment at room temperature to detect the initiation of small cracks, and to study the growth behavior of the crack (e.g., crack size). All tests were performed on specimens of 7075-T6 aluminum alloy.

In the reported work, 21 experiments have been conducted to build an automated monitoring system for the fatigue damage properties of polycrystalline materials. The dimensions of these specimens are 3 mm thick, 50 mm wide with (1 mm×3.5 mm) slot cut at the edge. Figure 1 presents an experimental apparatus for the NDT evaluation which contains two different sensors that are combined for information fusion to detect the damage initiation and to evaluate the risk of the damaged structure.

The computer-controlled and computer-instrumented fatigue-testing machine consists of the actuator, load cell, hydraulic system, and controller. Using this system, a specimen can be examined either under low-cycle or high-cycle fatigue tests at variable amplitude and frequency of random loads. All experiments were performed on tension-tension load cycles at 60 Hz. The target set-point load is 8000 N and the amplitude load is (+3500 N & -3500 N).

2.1. Ultrasonic testing:

High-frequency acoustic pulse (15 MHz ultrasonic waves) is injected into the specimen by a piezo-electric transducer called transmitter and received by another piezo-electric transducer called the receiver which is located on the other side of the transmitter. The strength of the signal is measured after it has propagated through the material. The strength of the signal is influenced by some material features (e.g., the grain boundaries, voids, and inclusions) that exist on the path of the propagated signals. However, the effect of the preexisting flaws such as voids, inclusions, or grain boundaries on the signal strength is stable. On the other hand, the strength of the signal is significantly affected (decreasing) by the fatigue crack creation because part of the signals are reflected and are not propagated to the receiver.

2.2. Optical metrology device:

The optical metrology device (Infinite-Focus, Alicona) provides 3D surface measurement. In the Focus-Variation system of Alicona, the topographical and color information are created from the variation of focus where the small depth of focus of an optical system is combined with vertical scanning. The vertical resolution of the Infinite-Focus system reaches 20nm. The size of the generated image using Alicona is 0.4 mm by 0.4 mm, as shown in figure 2, and each image has 4,161,600 pixels. Thus, Alicona can detect micro-cracks, that are considered to be the crack initiation regime.In this investigation, Alicona measurements were taken asynchronously with UT data to provide ground truth for UT signal weakness. In the reported work, Alicona was used to provide evidence for the fatigue damage initiation, as shown in figure 3. [19] [20]

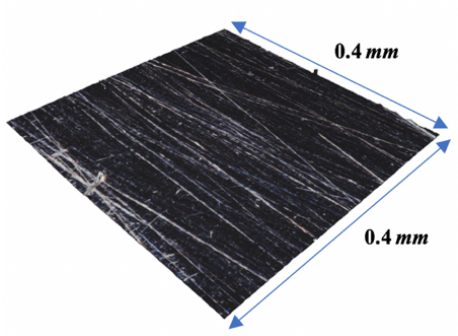


Figure 2: 3D surface generated using Infinite-Focus deceive.

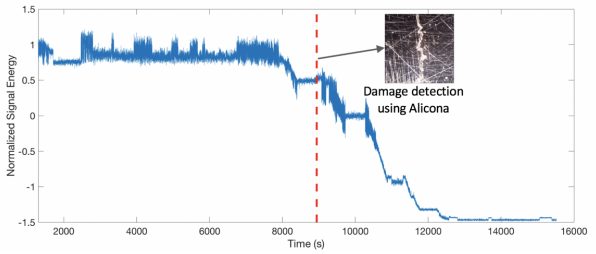


Figure 3: Damage detection using normalized signal energy data synchronized with evidence from Alicona.

3. Methodology

This section presents the methodology for fatigue damage detection and evaluation, improving the detection performance, classifying the damage risk, and building an automated model for damage detection and classification using a neural network.

3.1. Signal energy

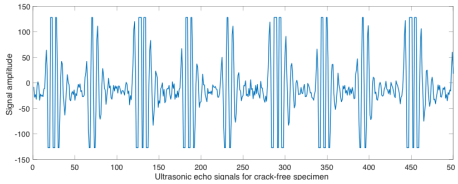
In industry, ultrasonic testing is one of the most extensively used non-destructive methods for determining the minor and internal damages and defects of mechanical structures. The response of UT signals usually leads to finding the location of the damage and the severity of that damage, where the attenuation of the measured signals provides evidence for the detection of damage (e.g., cracks). Figure 4 presents typical signal responses at the receiver end of an ultrasonic transducer for undamaged structure and damaged structure with the severity of each damage, respectively.

In signal processing applications, signal energy is widely used for quantifying signal strength. In this study, a decreasing trend in the signal strength serves to detect the onset of fatigue damage. It follows from Figure 4 that, by applying Eq. (1) on UT signals, the signal energy, $E(t)$, of a cracked specimen is seen to be less than that of a crack-free specimen.

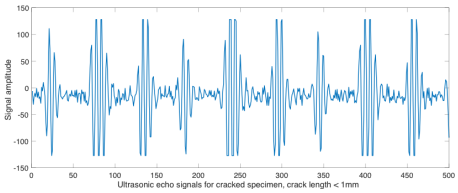
$$E(t) = \int_{-\infty}^{\infty} \mid x(t) \mid^2 dt \tag{1}$$

After estimating the signal energy of UT signals using Eq. (1), the outputs should be standardized because the response the signal energy for all tested specimens is similar, but the values of the energy may differ from one to another as shown in Figure 5. The standardization is called a z-score, and it transforms UT signal energy to have a mean of zero and a standard deviation of one. UT signal energy can be standardized by using the following equation:

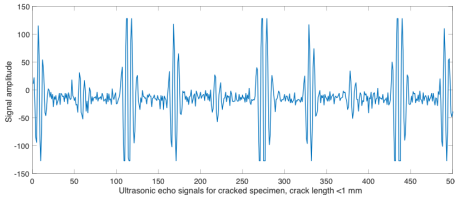
$$z_n = \frac{x_n - \bar{\boldsymbol{x}}}{\boldsymbol{S}} \tag{2}$$



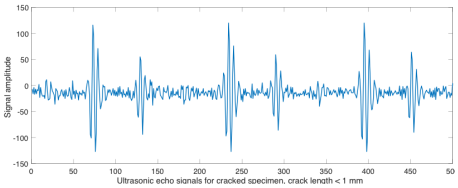
(a) Typical measured UT signal, healthy status of UT signal, all measured signals are detected .



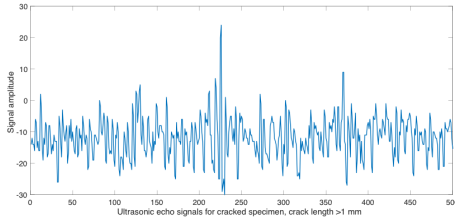
(b) Typical measured UT signal, Damage class-1 status of UT signal, most of measured signals are detected, and few measured signals are reflected .



(c) Typical measured UT signal, Damage class-2 status of UT signal, some of measured signals are detected, and some of the measured signals are reflected.



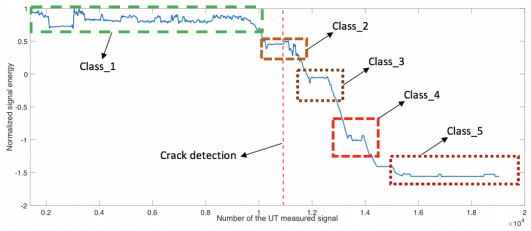
(d) Typical measured UT signal, Damage class-3 status of UT signal, few of measured signals are detected, and most of the measured signals are reflected.



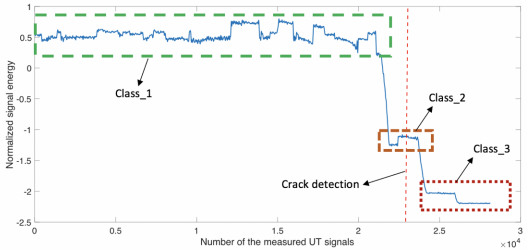
(e) Typical measured UT signal, Damage class-4 status of UT signal, all measured signals are reflected .

Figure 4: Comparison between different statuses of typical UT signals.

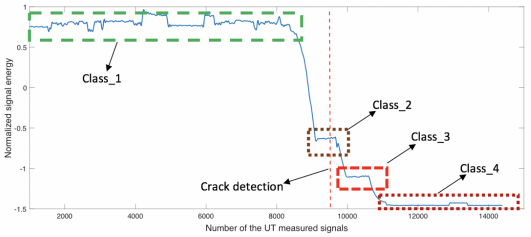
where, $\bar{\boldsymbol{x}}$ is the mean and $\boldsymbol{S}$ is the standard deviation.



(a) Smooth signal energy of experiment A.



(b) Smooth signal energy of experiment B.



(c) Smooth signal energy of experiment C.

Figure 5: The classification methodology of three experiments data,experiment A presents five classes, experiment B shows three classes, and experiment C contains four classes.

3.2. Discrete Wavelet Transform:

As shown in Figure 3,

the measurements of signals energy present in nonstationary forms make the estimation of the damage onset more difficult. However, the Bulk of the pertinent information in measurements exists in nonstationary forms and also has irregular wave structures. Therefore, it is preferable to decompose the signals

into building blocks that are well localized at different levels of scales and time-translations. In this task, the energy signals have been decomposed by digital wavelet transform (DWT) [21, 22] that is suitable for analyzing nonstationary measurements.

Multiresolution analysis (MRA) [21, 22], which belongs to the class of DWT, is suitable for decomposing UT signals into the *approximation* and the *detail* parts of the signal. The approximation part characterizes the low-frequency components of the signal, while the detail part represents the high-frequency components of the signal. Each approximation of the signal is further orthogonally decomposed into a hierarchical set of details and approximations.

Figure 6 presents an example for the process of DWT, where the Haar low pass filter H_0 operation is represented by the following matrix

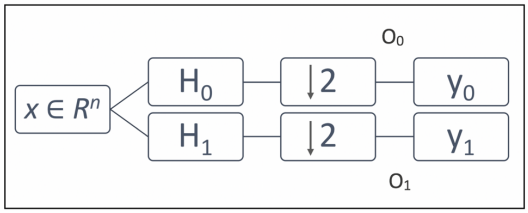


Figure 6: The block diagram of the two-channel analysis filter banks

$$H_0^{(n)} = \frac{1}{2}\begin{bmatrix} 1 & 1 & & & & & \\ & 1 & 1 & & & & \\ & & & \ddots & & & \\ & & & & \ddots & & \\ & & & & & 1 & 1 \\ & & & & & & 1 & 1 \end{bmatrix} \in R^{n\times n} \tag{3}$$

and the Haar high pass filter H_1 operation can be represented as

$$H_1^{(n)} = \frac{1}{2}\begin{bmatrix} 1 & -1 & & & & & & \\ & 1 & -1 & & & & & \\ & & & .. & & & & \\ & & & & & 1 & -1 & \\ & & & & & & 1 & -1 \end{bmatrix} \in R^{n\times n} \tag{4}$$

215

The down-sampling operation is represented as the following matrix, this step picks out the odd entries (i.e., first, third, etc.) of the matrices.

$$d^{(n)} = \begin{bmatrix} 1 & 0 & 0 & . & . & 0 & 0 \\ 0 & 0 & 1 & . & . & 0 & 0 \\ . & . & . & & & . & . \\ . & . & . & & & . & . \\ 0 & 0 & 0 & . & . & 1 & 0 \end{bmatrix} \in R^{\frac{n}{2}\times n} \tag{5}$$

The combined mathematical action of low pass filtering, followed by down-sampling, is represented as

$$O_0^{(n)} = d^{(n)}H_0^{(n)} = \begin{bmatrix} 1 & 1 & & & & & & \\ & & 1 & 1 & & & & \\ & & & & .. & & & \\ & & & & & .. & & \\ & & & & & 1 & 1 & \\ & & & & & & & 1 & 1 \end{bmatrix} \in R^{\frac{n}{2}\times n} \tag{6}$$

Similarly, the combined mathematical action of high pass filtering, followed

by down-sampling, is represented as

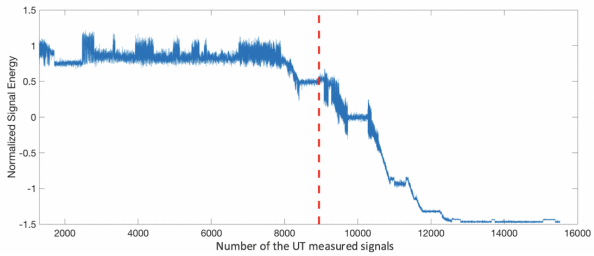
$$O_1^{(n)} = d^{(n)} H_1^{(n)} = \begin{bmatrix} 1 & 1 & & & & & & \\ & & 1 & -1 & & & & \\ & & & & \ddots & & & \\ & & & & & \ddots & & \\ & & & & & 1 & -1 & \\ & & & & & & 1 & -1 \end{bmatrix} \in R^{\frac{n}{2} \times n} \quad (7)$$

In summary, the filter bank is represented as

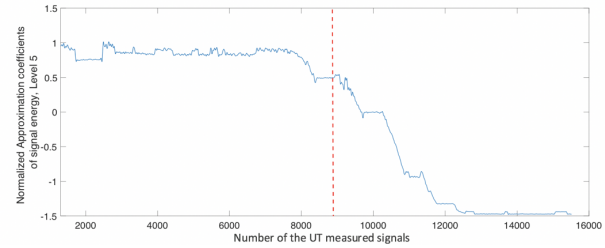
$$\text{Approximation part: } y_0^{(n)} = O_0^{(n)} x \in R^n \quad (8)$$

$$\text{Detail part: } y_1^{(n)} = O_1^{(n)} x \in R^n \quad (9)$$

In the reported work, the main objective of using DWT is smoothing the signal energy such that the nonstationary form is eliminated. The low-pass filtering serves the purpose of data smoothing by averaging adjacent entries of its input, and the difference between adjacent entries of its input is used in high-pass filtering of the signal. The process of DWT is first applied to the full-length vector of original data. Then, the outputs are smoothed and decimated by half such that the output vector length is half the input vector length. This process is continued until low-frequency smoothed data are reached as desired. Hence, each process of DWT provides a higher resolution of the data while smoothing the remaining data. Figure 7 illustrates how the smoothing data using DWT provides a better illustration for damage detection and risk classification.[23],[24],[25],[26],[27],[13].



(a) Typical measured UT signal energy of one experiment.



(b) Corresponding smooth signal energy data by DWT analysis.

Figure 7: Comparison between typical signal energy and corresponding level 5 approximation coefficients of UT signals, the red dashed line shows the moment of the crack onset

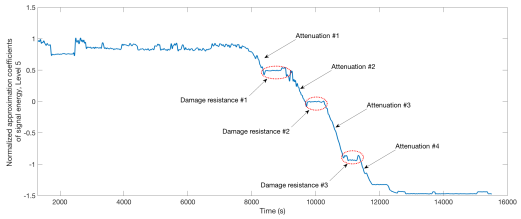
3.3. Classification

All mechanical structures contain defects at the nano/microscale level, which are usually hidden and may have been produced during the manufacturing process. Identification of a fatigue failure at an initial stage is a challenging task and hence prediction of onset of the next stage is uncertain and may often be ambiguous. These defects may differ from one component to another; hence the damage initiation and its growth vary from a component to another. Figures 5 (a),(b)& (c) present the signal energy response for three individual specimens, where the responses of signal energy during the fatigue damage growth for all specimens are not the same. For example, the first attenuation of signal energy response of Figure 5(b)& (c) is severe, which means the size of the first detected damage is large. On the other hand, the size of the first detected damage of Figure 5(a) is small, because the attenuation of the energy response is small as compared to those in Figure 5(b)& (c). Furthermore, the stability of signal

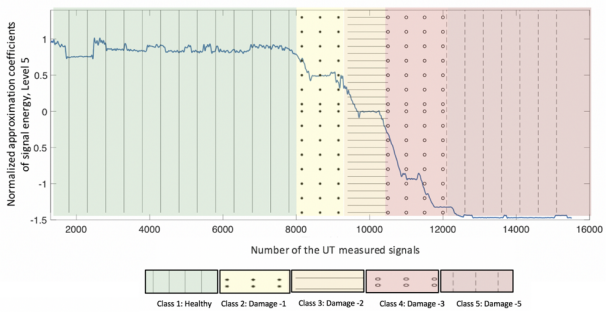
energy responses may also differ from one specimen to another. For example, figure 5(b) presents three stability regions and classes of signal energy responses, while Figures 5(c)&(a) show four and five classes, respectively.

The signal energy is estimated using equation (1). The best case of damage risk evaluation in this study is presented in Figure 8, where the highest signal energy characterizes the healthy condition class of the tested specimen. The healthy condition (i.e., crack-free component) is defined when the whole signal is received by the receiver transducer as shown in Figure 4(a). The next highest is the first damage class, where most of the signal is detected, but a small part of the signal is reflected due to the damage initiation, as illustrated in Figure 4(b). Below this comes the second damage class, where part of the signal is detected, while the other part is reflected because of the damage extension, as shown in Figure 4(c). The next class is the third damage class, where a substantial part of the signal is not transmitted, and just a few signals are detected, as shown in Figure 4(d). The last part represents the fourth class where all UT signals are practically reflected black. The length of each class depends on the stability of the attenuation which could refer to the constancy of the damage extension.

As shown in Figure 8, the classification strategy in this paper is defined based on the attenuation and the stability of the signal energy responses where more classes imply more resistance to the fatigue damage growth, and the higher the risk of damage, the lower the UT signal energy responses. The best damage classification case of this investigation is where there are five classes, as shown in Figure 5. Figure 9 presents the shape of the damage on the notch surface of the specimen for each damage class.



(a) Clarification of the damage growth and damage stability using smooth energy signal data



(b) Smooth energy signal data classification of experiment A, the first area from left side represents *no risk* (i.e., crack free), the second area represents *low risk*, the third area represents *medium risk*, the fourth area represents *high risk*, and the last area represents end-of-service-life (i.e., unusable or broken).

Figure 8: The classification strategy for the damage risk assessment of experiment A, the attenuation of the signal and its stability determine the classification strategy

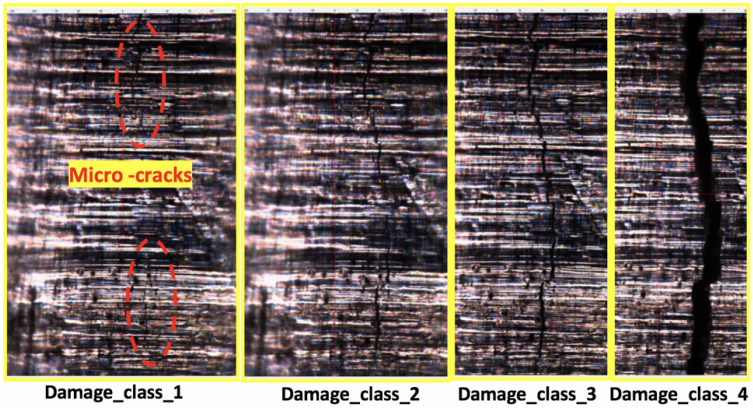


Figure 9: Illustration of the fatigue damage size on the notch of the specimen for each class using Alicona.

3.4. Artificial Neural Network

The artificial Neural network (ANN) is a computational method that attempts to mimic some of the functions of a human brain. The neural network is composed of a set of nodes, while the human brain is composed of connections of neurons. Hence, ANN imitates the most vital mechanism of the brain, which is the neural association. The structure of ANN is significantly affected by the connection process between nodes, and different types of connections can build different types of ANN. Feed-forward neural network is considered to be one of the common types of ANN [28].

3.4.1. Feed-forward Neural Network

The sequence of processes in this neural network is in one direction such that the ANN process starts from input nodes and ends at the output nodes. The neural network can be altered from a simple architecture to a very complex architecture. The single-layer neural network is the simplest neural network architecture. Figure 10 presents a single-layer neural network that consists of an input layer and a hidden layer. The second type of neural network architecture is known as a shallow or Vanilla multi-layer neural network, and it consists of an input layer, one hidden layer, and an output layer. However, when the hidden

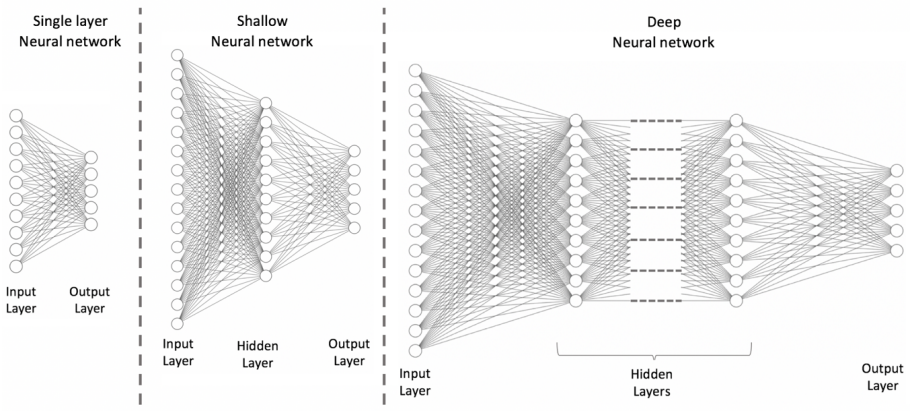


Figure 10: Three different types of neural network architecture.

layers become more than one layer, the neural network architecture is often defined as a deep neural network (DNN). As shown in the figure, the structure of the neural network consists of more than one hidden layer, input layer, hidden layers, and output layer. The reason for defining the hidden layers by that name is that they are not accessible from outside of the neural network [29].

Three major learning models are commonly used in neural network applications: supervised learning model, unsupervised learning model, and reinforcement learning model:

(i) *Supervised learning* is a learning rule that uses the training data to set the neural network parameters, where these parameters are determined based on already known output. The training data is represented usually in pairs of input and target (known output).

(ii) *Unsupervised learning* is an independent learning process, where the input data that share similar features are joined to form clusters. In ANN under unsupervised learning, the pattern of the input data determines the class to which input data belongs. Thus, the features and patterns of the input data must be discovered by the network itself to classify data in distinct clusters by their similarity.

(iii) *Reinforcement learning* is a learning rule that sets the parameters of ANN by interacting with the environment. This type of learning is essentially built upon a trial and error learning method, because the data are generally not provided. The reinforcement learning technique depends on the consequences of its actions and determines its actions that are largely dependent on its experience (exploitation) and the exploration of the latest choices.

.

Figure 11 presents the basic operation process of the neural networks, where the input data (x_1, x_2, x_3) are multiplied by weights (w_1, w_2, w_3) and added by a bias b before entering the node. The input data that has a significant effect has a higher weight value and vice versa. The results of each node are expressed as the following equations.

$$x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \; w = \begin{bmatrix} w_1 w_2 w_3 \end{bmatrix} \tag{10}$$

$$a = w * x + b = w_1 x_1 + w_2 x_2 + w_3 x_3 + b \tag{11}$$

$$y = \phi(a) = \phi(wx + b) \tag{12}$$

where ϕ is the activation function, and the output of the node is computed using an activation function [30].

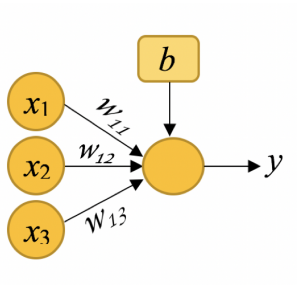


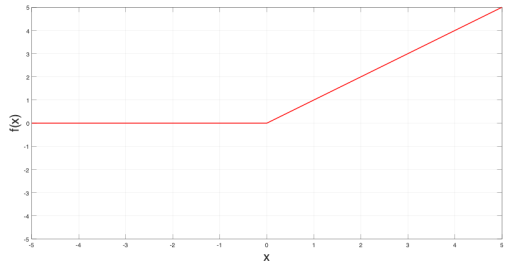
Figure 11: An artificial neuron.

3.4.2. Activation function

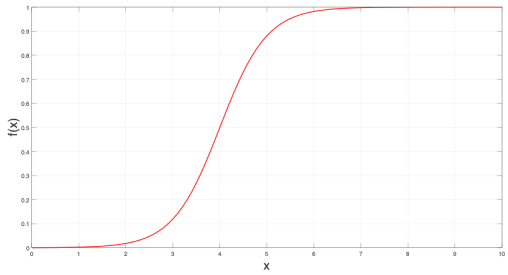
The activation function is used to estimate the output of neural network like yes or no. Based on the applied function, it maps the computed values in between 0 to 1 or -1 to 1. The Rectified Linear Unit (ReLU) function is perhaps the most used activation function in the artificial intelligent community, because it has been extensively used in deep learning and convolutional neural networks. The ReLU, $f(x) = max(x, 0)$, has no bounded output on the positive side such that $f(x)$ is zero when x is less than zero and $f(x)$ is equal to x when x is greater than or equal to zero. Also, the gradients of ReLU are always zeros and ones, as shown in Figure 12 a.

Another popular activation functions is the Sigmoid function that produces an analogue (nonlinear) output, bounded between 0 and 1, as seen in Figure 12 b. The Sigmoid function is described by:

$$y = \frac{1}{1 + e^{a}} \tag{13}$$



(a) The Rectified Linear Unit (ReLU) function.



(b) The Sigmoid function.

Figure 12: Popular activation functions used in NN

The information in the neural network system is stored in terms of weights. These weights are adjusted during the training of the neural network based on the value of the error, where the error is the difference between the output of the neural networks and the correct output.[31] [32] [33] [34].

3.4.3. Back propagation

During the training/learning phase of NN, the values of the weights (w_{ij}) are adjusted to enhance the performance of the ANN model. Each cycle of the training/learning phase refers to one epoch, where a set of input data is proceeded through the ANN architecture to create outputs of the NN. These outputs are compared with the desirable outputs. Based on the difference between the produced outputs and the desired outputs, the error is estimated. In the back-propagation process, the estimated error is passed in the reverse direction of the NN architecture, from the output layer to the input layer, to re-adjusted

weights by using the back-propagation algorithm. This technique is repeated continuously for the next epochs until the desired error is accomplished. The mean squared error of the ANN is expressed by:

$$J_{ANN} = \frac{1}{N} \sum_{t=1}^{N} \sum_{e=1}^{O} (T_{ANN}(t,e) - y_{ANN}(t,e))^2 \tag{14}$$

N: Number of trained data (input, output).

e: Index denotes the number of the output.

t: Index denotes the number of trained data.

O: Number of the ANN outputs.

T: The ANN target values.

3.4.4. Gradient descent (GD)

Gradient descent method (e.g., steepest descent) is used to adjust the weights in the direction of the performance function that declines most rapidly (e.g., the most negative of the gradient). Equation 15 presents the adjustment of One of the network weights by using the conjugate gradient (CG) algorithm.

$$w_{n+1} = w_n - 2\alpha J_n E_n \tag{15}$$

n: Number of iteration.

J: Jacobian matrix of J_{ANN}.

E: The computed error between ANN outputs and the target.

Although the gradient descent (GD) is an important optimization method, it has four main disadvantages: (i) the learning rate is low; (ii) the direction is not perfect-scaled. Thus, the iterations number mostly depends on the scale of the problem; (iii) determining the local minimum point could be missing, and (iv) highly influenced by noise [35][36].

3.4.5. Conjugate gradient (CG)

One of the methods that are established to improve the performance of the GD algorithm is the conjugate gradient algorithms. All the conjugate gradient

algorithms are performed by searching in the steepest descent direction on the first iteration. A line search is then implemented to find the optimal point to reach along the current search direction. Then the next search direction is established such that it is conjugate to prior search directions. Generally, determining the new search direction requires a combination between the new steepest descent direction and the previous search direction.

$$p_0 = -g_0 \quad , where \quad g = J^T E \tag{16}$$

$$w_{n+1} = w_n + \alpha_n p_n \tag{17}$$

$$p_n = -g_n + \beta_n p_{n-1} \tag{18}$$

Where the parameter(β_n) is a constant which is calculated to force the successive directions to be conjugate.

The types of CG algorithms are distinguished by the method in which the constant β_n is calculated. Most of the conjugate gradient algorithms involve a line search at each iteration. The line search technique is computationally expensive because the network response to all training data is processed various times for each search. In 1993, Moller established a method that overcomes the time-consuming line search. This method is known as the scaled conjugate gradient algorithm (SCG). The basic concept of this method is to combine the model-trust region approach, where the maximum distance is selected first, followed by the direction, with the CG approach.[37] [38] [39].

In this study, the SCG algorithm is used to set ANN weights. Equations 19, 20 present β_n parameter calculation and direction of the new search

$$\beta_n = \frac{|g_{n+1}|^2 - g_{n+1}^T g_n}{g_n^T g_n} \tag{19}$$

$$p_{n+1} = -g_{n+1} + \beta_n p_n \tag{20}$$

3.5. Neural network pattern recognition:

Pattern recognition is a process that is used to solve classification problems in the neural network. To solve a pattern recognition problem, the data must be prepared, where a set of D input vectors is arranged as columns in a matrix. Then, another set of D target vectors is arranged, such that target vectors indicate the corresponding classes of the input vectors.[40]

To ensure that the neural network model is going to be effective, the training dataset must satisfy two objectives; (i) Every group must be characterized so that each group patterns is represented in the data set; (ii) statistical deviation must be effectively represented within each class, where the neural network data must contain the total range of data including noise [41].

Figure 13 presents the necessary steps for building an NN model with a shallow architecture.

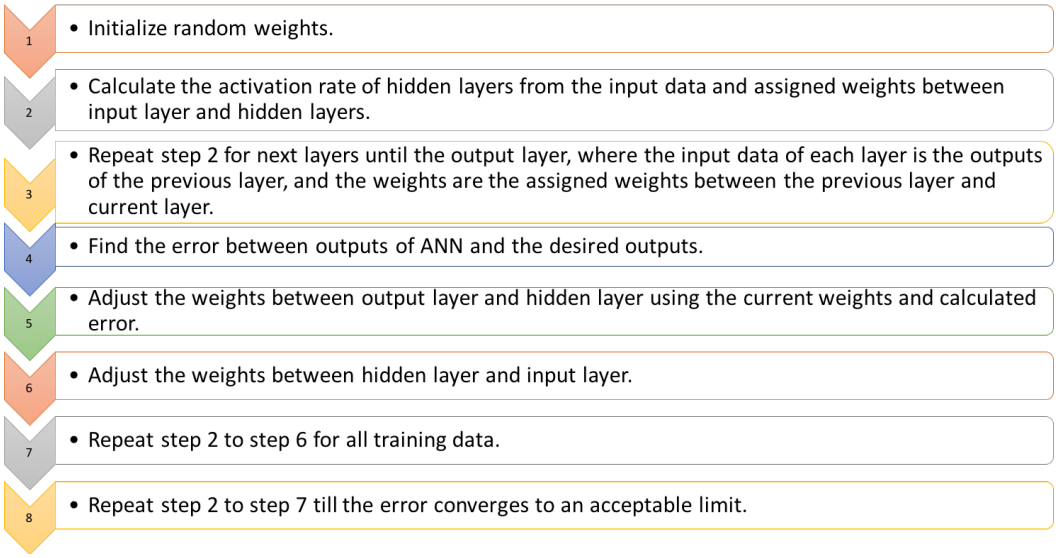


Figure 13: The process of the neural network analysis.

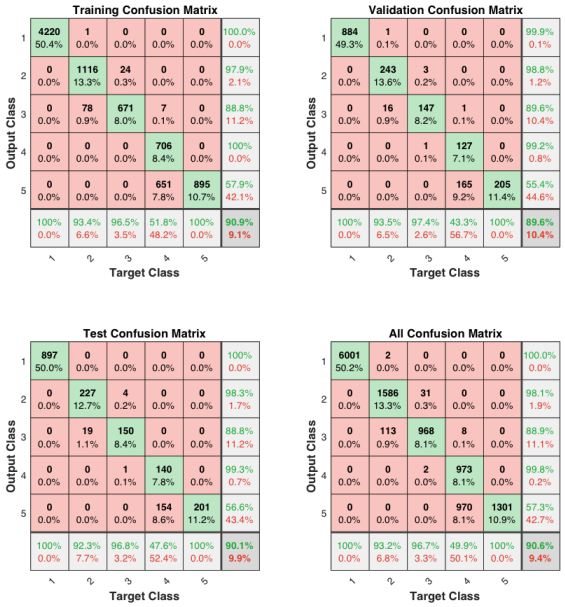
4. Results of Experimental Validation

This section presents the experimental results for validation of two NN models. The first model presents the results of the NN model for the signal energy data, and it is defined by the NNSE model. The second part shows the results of the NN model for approximation coefficients of the UT signal energy, NNAC model.

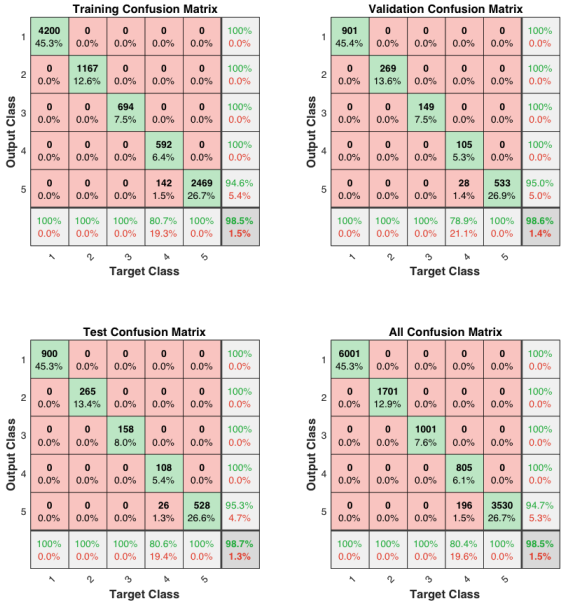
4.1. The model confusion matrices

Figures 14 (a) & (b) present respectively the confusion matrices of the original data (i.e., signal energy) and its approximation coefficients for training, testing, and validation. The outputs of the network models are around 90%, which is excellent; the high numbers of correct responses are observed as the green squares, while the low numbers of fault responses are indicated by the red squares. The lower right gray squares show the overall accuracy.

For the signal energy model, an accuracy of $\sim$ 90.9% is achieved for the training data set, $\sim$ 89.6% for the validation data set, and $\sim$ 90.1% for the testing data set; and the overall accuracy of the model is $\sim$ 90.6%. The model performance for classifying the first class is perfect (i.e., almost 100 %), while almost 50 % of the fourth damage data were classified correctly, which significantly affect the overall model performance. This problem has been resolved by smoothing the signal energy using DWT, which makes the data more separable. As shown in the figure 8, the approximation signal energy model yields an accuracy of $\sim$ 98.5% for the training data set, $\sim$ 98.6% for the validation data set, and $\sim$ 98.7% for the testing data set, and the overall accuracy of the model is $\sim$ 98.5%. The overall performance of this model increases by almost 8.7% from the previous model; the rationale for the enhanced performance is the improved classification performance for the last class.



(a) Confusion matrices of the signal energy model: $\sim 90.60\%$ performance.

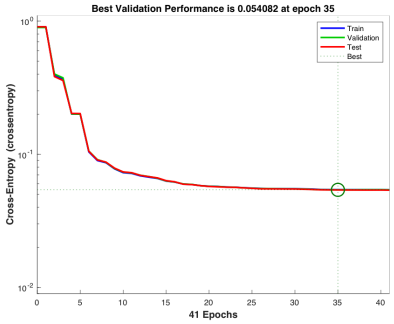


(b) Confusion matrices of the approximation signal energy model: $\sim 98.50\%$ performance.

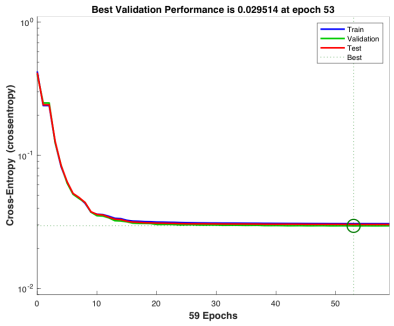
Figure 14: Comparison of the neural networks model performance between of the the signal energy and the approximation signal energy

4.2. The model performance

Figures 15 (a) & (b) show the NN model performance for training, validation, and testing phases for the NNSE model and the NNAC model, respectively. The performance is measured in terms of mean squared error. The best validation performance of the NNSE model fits the best value which is 0.054 after 35 epochs, while the best validation performance of the NNAC model reaches 0.029 after 53 epochs. Although the performance of both models are acceptable in many NN applications, the performance of NNAC model is significantly improved by reducing the mean squared error which is $\sim 46.3\%$ as compared to the NNSE model, which is more applicable for some sensitive NN applications that do not accept errors.



(a) Best validation performance of the NNSE model.

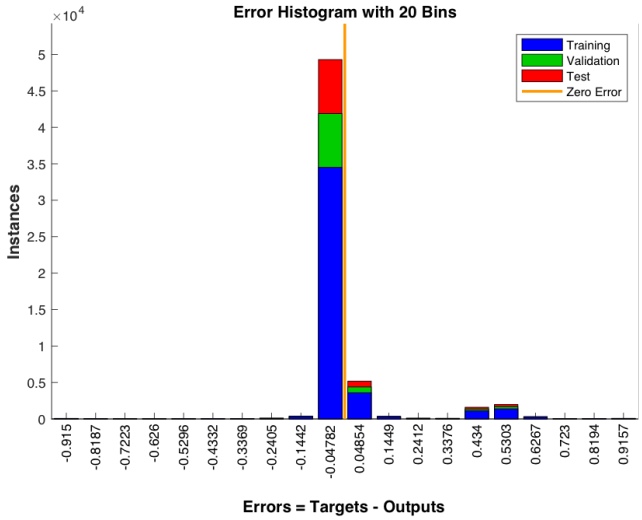


(b) Best validation performance of the NNAC model.

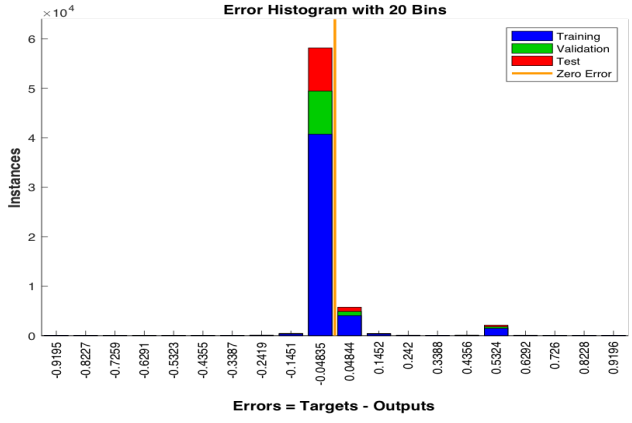
Figure 15: Best validation performance for all experiments.

4.3. Ther Error Histogram

Figures 16 (a) & (b) also display respectively the error histogram of the NNES model and NNAC model for the training, validation and testing steps. These figures show that the data fitting errors are located around zero, which is a reasonably good range.



(a) Error histogram of the first NN model.



(b) Error histogram of the second NN model.

Figure 16: Error histogram for all experiments.

5. Conclusion

This paper has developed a neural network (NN)-based model for a damage-tolerant design using the energy of the ultrasonic signals. Furthermore, this paper has improved the performance of the NN model by smoothing the input data by using the discrete wavelet transform (db10). The proposed damage-tolerant design makes use of the NN model to detect the fatigue damage and to classify the damage into five classes of increasing risk, which are: no-risk (i.e., healthy), low-risk, medium-risk, high-risk, and end-of-service-life (i.e., unusable or broken). The classification strategy, adopted in this paper, has been built upon the attenuation of the signal energy and the stability of this attenuation. The attenuation depends on the size of the damage, while the stability of the attenuation indicates the stationary characteristic of the damage size.

The overall performance of the NN model using the original data is around 90%. The difficult problem addressed by this model is classification of damage-4 data, where almost half of these data are mis-classified. This problem is resolved by smoothing the input data by using the discrete wavelet transform (db10). The overall performance of the NN model by using smoothed data increased by 8.7%, which reaches the level of 98.5% success. However, the generalization of these results is subject to certain limitations. For instance, the number of the classes is significantly affected by the damage growth behavior, which is influenced by the material properties, pre-existing defects, applied loads, and environmental conditions.

Future studies are required to better understand the effects of crack initiation and crack propagation on the signal energy of ultrasonic signals such that the NN model could classify the severity of the damage into many classes.

Acknowledgment

The first author gratefully acknowledges the financial support of the Saudi Arabian Cultural Mission (SACM). The work reported in this paper has been supported in part by U.S. Air Force Office of Scientific Research (AFOSR) under

Grant No. FA9550-15-1-0400 in the area of dynamic data-driven application systems (DDDAS).

References

[1] C. R. Farrar, K. Worden, Structural health monitoring: a machine learning perspective, John Wiley & Sons, 2012.

[2] S. Suresh, Fatigue of materials, Cambridge university press, 1998.

[3] F. C. Campbell, Elements of metallurgy and engineering alloys, ASM International, 2008.

[4] A. F. Grandt Jr, Fundamentals of structural integrity: damage tolerant design and nondestructive evaluation, John Wiley & Sons, 2003.

[5] F. Grandt Jr, Damage tolerant design and nondestructive inspection-keys to aircraft airworthiness, Procedia Engineering 17 (2011) 236–246.

[6] B. Raj, T. Jayakumar, M. Thavasimuthu, Practical non-destructive testing, Woodhead Publishing, 2002.

[7] A. Ray, R. Patankar, Fatigue crack growth under variable-amplitude loading: Part i – model formulation in state-space setting, Applied Mathematical Modelling 25 (11) (2001) 979–994.

[8] A. Ray, R. Patankar, Fatigue crack growth under variable-amplitude loading: Part ii - code development and model validation, Applied Mathematical Modelling 25 (11) (2001) 995–1013.

[9] D. S. Dugdale, Yielding of steel sheets containing slits, Journal of the Mechanics and Physics of Solids 8 (2) (1960) 100–104.

[10] R. Pippan, A. Hohenwarter, Fatigue crack closure: a review of the physical phenomena, Fatigue & fracture of engineering materials & structures 40 (4) (2017) 471–495.

[11] L. Jaanuska, H. Hein, Crack identification in beams using haar wavelets and machine learning methods, Int. J. Mech 10 (2016) 281–287.

[12] X. Fang, H. Luo, J. Tang, Structural damage detection using neural network with learning rate improvement, Computers & structures 83 (25-26) (2005) 2150–2161.

[13] H. Alqahtani, A. Ray, Forecasting and detection of fatigue cracks in polycrystalline alloys with ultrasonic testing via discrete wavelet transform, Journal of Nondestructive Evaluation, Diagnostics and Prognostics of Engineering Systems (2021) 1–15.

[14] A. Rageh, S. E. Azam, D. G. Linzell, Steel railway bridge fatigue damage detection using numerical models and machine learning: Mitigating influence of modeling uncertainty, International Journal of Fatigue 134 (2020) 105458.

[15] N. Bakhary, H. Hao, A. J. Deeks, Damage detection using artificial neural network with consideration of uncertainties, Engineering Structures 29 (11) (2007) 2806–2815.

[16] J. Zapico, M. Gonzalez, K. Worden, Damage assessment using neural networks, Mechanical Systems and Signal Processing 17 (1) (2003) 119–125.

[17] G. Luo, D. Zhang, D. Baleanu, Wavelet denoising, Advances in wavelet theory and their applications in engineering, physics and technology (2012) 634.

[18] E. Keller, A. Ray, Real-time health monitoring of mechanical structures, Structural Health Monitoring 2 (3) (2003) 191–203.

[19] R. Danzl, F. Helmli, S. Scherer, Comparison of roughness measurements between a contact stylus instrument and an optical measurement device based on a colour focus sensor, in: Proc. of the Nanotechnology Conference, 2006, pp. 284–287.

[20] R. Danzl, F. Helmli, S. Scherer, Focus variation–a robust technology for high resolution optical 3d surface metrology, Strojniški vestnik-Journal of mechanical engineering 57 (3) (2011) 245–256.

[21] G. Kaiser, A Friendly Guide to Wavelets, Birkhauser, Boston, MA, USA, 1994.

[22] S. Mallat, A Wavelet Tour of Signal Processing: The Sparse Way, 3rd ed., Academic Publishers, Amsterdam, The Netherlands, 2009.

[23] G. A. Alonso, J. M. Gutiérrez, J.-L. Marty, R. Muñoz, Implementation of the discrete wavelet transform used in the calibration of the enzymatic biosensors, Discrete Wavelet Transforms-Biomedical Applications.

[24] O. Rioul, M. Vetterli, Wavelets and signal processing, IEEE signal processing magazine 8 (4) (1991) 14–38.

[25] M. O. Oliveira, A. S. Bretas, Application of discrete wavelet transform for differential protection of power transformers, in: 2009 IEEE Bucharest PowerTech, IEEE, 2009, pp. 1–8.

[26] G. D'Angelo, S. Rampone, Feature extraction and soft computing methods for aerospace structure defect classification, Measurement 85 (2016) 192–209.

[27] A. R. Mane, P. Biradar, P. Shastri, Review paper on feature extraction methods for eeg signal analysis, in: Department Of Electronics and Telecommunication Engineering, VPCOE/Savitribi Phule University, IJEEBS, 2015, pp. 2349–6967.

[28] A. M. Zador, A critique of pure learning and what artificial neural networks can learn from animal brains, Nature communications 10 (1) (2019) 1–7.

[29] K. Gurney, An introduction to neural networks, CRC press, 2018.

[30] Y. Li, Deep reinforcement learning: An overview, arXiv preprint arXiv:1701.07274.

[31] A. C. Coolen, A beginner's guide to the mathematics of neural networks, in: Concepts for Neural Networks, Springer, 1998, pp. 13–70.

[32] M. Zakaria, M. Al-Shebany, S. Sarhan, Artificial neural network: a brief overview, Int J Eng Res Appl 4 (2014) 7–12.

[33] M. Islam, G. Chen, S. Jin, An overview of neural network, American Journal of Neural Networks and Applications 5 (1) (2019) 7–11.

[34] A. Zayegh, N. Al Bassam, Neural network principles and applications, in: Digital Systems, IntechOpen, 2018.

[35] X. Wang, Method of steepest descent and its applications, IEEE Microwave and Wireless Components Letters 12 (2008) 24–26.

[36] S. Ruder, An overview of gradient descent optimization algorithms, arXiv preprint arXiv:1609.04747.

[37] C. M. Bishop, Pattern recognition and machine learning, springer, 2006.

[38] M. F. Møller, A scaled conjugate gradient algorithm for fast supervised learning, Neural networks 6 (4) (1993) 525–533.

[39] F. Dkhichi, B. Oukarfi, Neural network training by gradient descent algorithms: Application on the solar cell.

[40] S. Ghorpade, J. Ghorpade, S. Mantri, Pattern recognition using neural networks, International Journal of Computer Science & Information Technology (IJCSIT) 2 (6) (2010) 92.

[41] R. Thirumalainambi, J. Bardina, Training data requirement for a neural network to predict aerodynamic coefficients, in: Independent Component Analyses, Wavelets, and Neural Networks, Vol. 5102, International Society for Optics and Photonics, 2003, pp. 92–103.