ORIGINAL ARTICLE

Feature extraction and neural network-based fatigue damage detection and classification

Hassan Alqahtani[1] · Asok Ray[2,3]

Abstract

This paper proposes a methodology for detection and classification of fatigue damage in mechanical structures in the framework of neural networks (NN). The proposed methodology has been tested and validated with polycrystalline-alloy (AL7075-T6) specimens on a laboratory-scale experimental apparatus. Signal processing tools (e.g., discrete wavelet transform and Hilbert transform) have been applied on time series of ultrasonic test signals to extract features that are derived from: (i) Signal envelope, (ii) Low-frequency and high-frequency signal spectra, and (iii) Signal energy. The performance of the neural network, combined with each one of these features, is compared with the ground truth, generated from the original ultrasonic test signals and microscope images. The results show that the NN model, combined with the signal-energy feature, yields the best performance and that it is capable of detecting and classifying the fatigue damage with (up to) 98.5% accuracy.

1 Introduction

Fatigue damage may occur in mechanical structures that are subjected to cyclic stresses even if the peak values of these stresses are much lower than the yield stress of the material. Fatigue damage is considered to be a major source of failures in mechanical structures, which may take place in notches, rough surfaces, and due to various internal defects (e.g., cracked second-phase particles and voids [1]). The size of a fatigue crack may keep on increasing with load cycles until the remaining cross section cannot support the applied load [2].

✉ Hassan Alqahtani
hhqahtani@taibahu.edu.sa

Asok Ray
axr2@psu.edu

1   Department of Mechanical Engineering, Taibah University, Medina, KSA 42353, Saudi Arabia

2   Department of Mechanical Engineering, Pennsylvania State University, University Park, PA 16802, USA

3   Department of Mathematics, Pennsylvania State University, University Park, PA 16802, USA

All mechanical structures contain defects at the level of nano-/micro scales, which might have been produced during the manufacturing process, and these defects usually stay hidden. Such defects may often evolve as fatigue damage, leading to structural failures. Therefore, early detection of evolving fatigue damage is very crucial for operation & maintenance of machinery. In this context, real-time sensing of fatigue damage is considered to be one of the successful tools for averting unexpected failures in mechanical structures. In many engineering applications, visual inspection (VI) is widely used for both damage assessment of various mechanical components (e.g., storage tanks, pressure vessels, and pipelines); however, there are several limitations to using VI. Usually, the errors introduced by VI may lead to either incorrectly identifying potentially pernicious defects (i.e., false alarms), or ignoring real defects (i.e., missed alarms). While false alarms may cause unnecessary maintenance and loss of plant availability, missing an existing critical defect may potentially degrade product quality & reliability as well as loss of plant availability, including even catastrophic accidents [2–5].

In many industrial processes, ultrasonic testing (UT) is one of the most extensively used non-destructive

techniques for estimating the internal damage and defects of mechanical structures. Often structural damage is detected from the time series of UT signals, where the attenuation of measured signals leads to detection of damage (e.g., evolution of cracks). Figure 1 a, b exhibits typical signal responses at the receiver end of an ultrasonic transducer for undamaged structure and damaged structure, respectively.
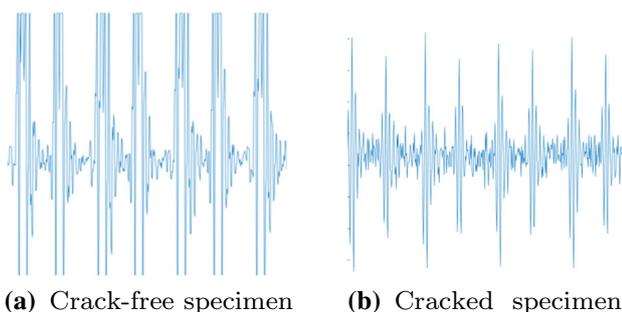
Several techniques and methods of data interpretation and signal processing have been applied on the ensemble of UT signals to extract the pertinent features. Along this line, the current paper examines the effectiveness of signal processing methods, namely, discrete wavelet transform (DWT), Hilbert transform (HT), and spectral energy, for performance enhancement of damage detection & classification algorithms, where the extracted features are obtained from low frequencies, high frequencies, signal envelope, and signal energy of UT signals. The performance of the damage monitoring system is measured in terms of *accuracy*, *sensitivity*, *specificity*, and *precision*; and its efficacy is determined by the area under the curve (AUC) of receiver operating characteristics (ROC) [6], the Matthews correlation coefficient (MCC) [7], and the F-Measure [8]. Figure 2 presents the proposed stages for monitoring the structure: [9–12]

This paper proposes an automated system of online fatigue damage assessment, which makes use of UT with the following:

1. Discrete wavelet transform (DWT)-based and Hilbert transform (HT)-based analysis for fatigue-damage detection, and
2. Neural networks (NN)-based fatigue-damage classification.

From the above perspectives of data analysis, the novelty of this paper is delineated as follows:.

1. *Data classification for damage detection in mechanical structures*: The data to be classified are acquired by synchronizing the images of a digital microscope with the corresponding UT signals; and the data classification strategy is built upon the onset of an evolving crack in the surface of the U-notch, as seen in Fig. 5, where the UT signals before the crack appearance are postulated to belong to the *healthy* class, while UT signals at and after the crack appearance are classified as *unhealthy*.

2. *Reduced human involvement in the decision of damage detection*: The automated real-time evaluation of fatigue damage, supported by computer-assisted analysis of nondestructive testing (NDT), is expected to significantly alleviate the shortcomings of visual inspection (VI); consequently, the resulting damage monitoring system could become largely independent of direct human decisions. In this context, various researchers have shown that the discipline of machine learning has the potential of building smart systems that are capable of thorough quality checks of manufactured products down to the finest points of inspection details [13, 14, 15].

3. *Development of a robust neural network-based decision support system*: The performance of feature extraction from UT signals is measured in terms of Accuracy, Sensitivity, Specificity, and Precision. The efficacy of these models is examined by various analytical tools such as Matthews correlation, Area Under Curve (AUC), and F-Measure. The best feature is determined based on the highest performance model.

The paper is organized into five main sections including the present one. The second section presents a description of the laboratory apparatus that serves as the data generator for validation of the methodology of fatigue-damage detection and classification, proposed in this paper. The third section illustrates the methodology including an overview of discrete wavelet transform (DWT), Hilbert transform (HT), and neural networks (NN). The fourth section discusses the results of experimental validation of the proposed methodology. Finally, the fifth section summarizes and concludes the paper with recommendations for future research.

2 Description of the experimental apparatus

The laboratory-scale experimental apparatus, which serves to validate the proposed methodology of fatigue-damage detection and classification, is illustrated in Fig. 3a. The apparatus is built upon a computer-controlled and computer-instrumented fatigue-testing machine[1], which is supplied with ultrasonic sensors[2] for generation of

(a) Crack-free specimen   **(b)** Cracked specimen

Fig. 1 Typical profiles of ultrasonic signals at the receiver transducer

---

[1] Manufacturer: MTS®, Berlin, New Jersey, USA.

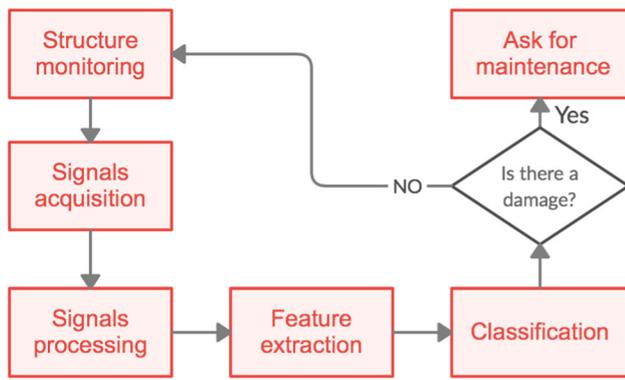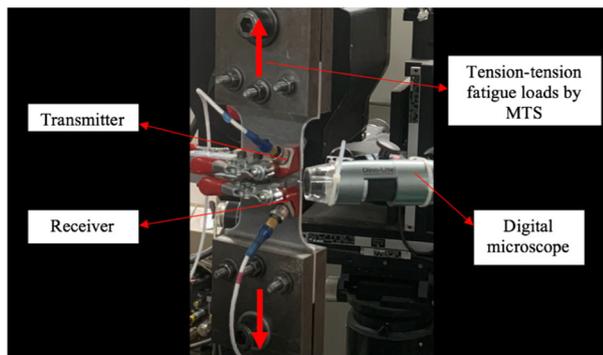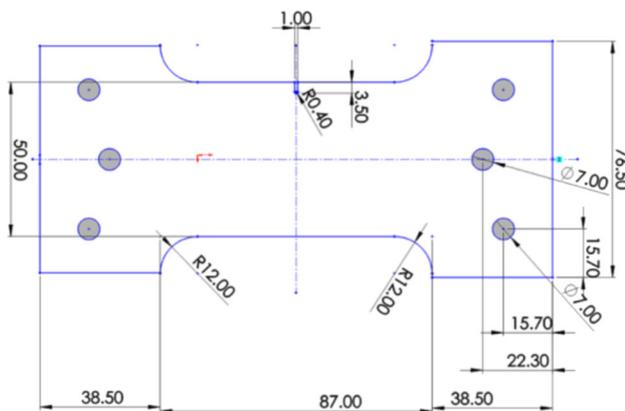[2] Manufacturer: OLYMPUS®, Shinjuku, Tokyo, Japan.

Fig. 2 Strategy of a real-time damage monitoring system

(a) Picture of the experimental apparatus.

(b) CAD drawing of an AL7075-T6 aluminum alloy specimen (all dimensions are in *mm* and are not to scale)

Fig. 3 The fatigue testing apparatus and ancillaries

ultrasonic test (UT) signals and a digital microscope[3] for generation of images for visual inspection (VI).

The main purpose here is to obtain ensembles of fatigue test data for assessment of the damage state of the structure under consideration (e.g., test specimens in Fig. 3b). Under medium-cycle to high-cycle fatigue loading of (ductile-alloy) machinery structures, bulk of the component service

––––––––––––
3 Manufacturer: QUESTAR®, New Hope, Pennsylvania, USA.

life is consumed in the crack initiation phase. Thus, the onset of fatigue cracks provides the necessary information for reduction in the probability of unanticipated failures as well as for control of the machinery performance; the objective here is to enhance both availability and reliability of machinery operation at a mitigated maintenance cost.

In the work reported in this paper, experiments have been conducted with *five* identical[4] test specimens (that are made of AL7075-T6 aluminum alloy) to build an automated system for fatigue damage detection. The dimensions of these specimens are: 50 mm width and 3 mm in thickness. As shown in Fig. 3b, a (1 × 3.5 mm) U-notch is incorporated at the edge. In the reported work, all specimens are subjected to tension-tension load cycles, generated by the fatigue testing machine in Figure 3a. A typical applied target set-point in this paper is a tensile-tensile load fluctuation, where the peak load is 11,000 N and the valley load is 6,000 N, i.e., the nominal stress fluctuates between 40 MPa (valley) and 73 MPa (peak).
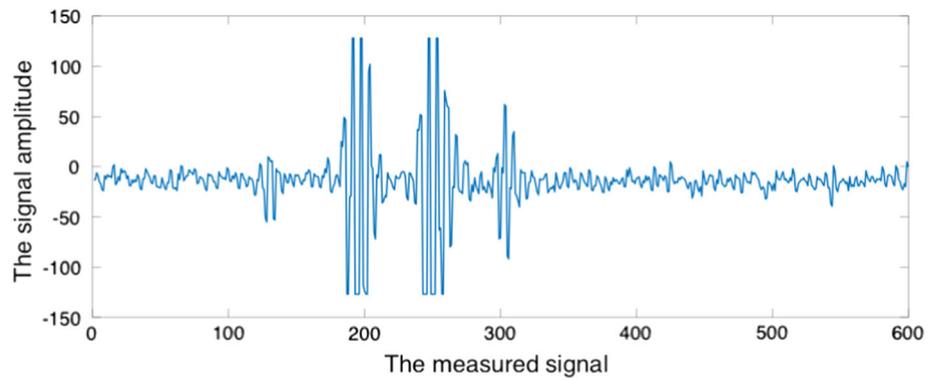
2.1 Ultrasonic testing

In the ultrasonic testing (UT) probes (see Fig. 3a), high-frequency acoustic pulses (i.e., 15 MHz ultrasonic waves) are injected into each specimen by a piezo-electric transducer (i.e., the transmitter), and are received after propagating through the material by another piezoelectric transducer (i.e., the receiver), which is placed on the other side of the transmitter, as seen in Fig. 3(a). In the experiments, UT sensors (i.e., transducers) are electrically contacted to the data acquisition device that has the capabilities of signal conditioning and analog to digital conversion (ADC). The ADC device is connected to the laboratory computer to collect the time series of measured UT signals. Figure 4 presents a typical UT signal, measured and recorded by the computer in the fatigue testing apparatus system.

The onset of a fatigue crack is characterized by the response of the UT signal which is measured after the crack has propagated through the specimen. The response of UT signal is (possibly) influenced by the preexisting flaws in a test specimen (e.g., grain boundaries, voids, and inclusions), which are located on the path of the propagated signals. The effects of the preexisting flaws on the response of a UT signal are considered to be very gradually evolving and are therefore quasi-static over the crack initiation phase of the structure. On the other hand, the amplitude of a UT signal tends to decrease significantly when the crack

––––––––––––
4 The reason for using five (apparently) identical specimens is to build both consistency and credibility of experimental results. The rationale is that, due to the uncertainties accruing from internal defects and the machining process, the fatigue life of similar test specimens may significantly duffer.

Fig. 4 A typical profile of an ultrasonic signal

Fig. 5 Examples of typical UT signal and microscope image
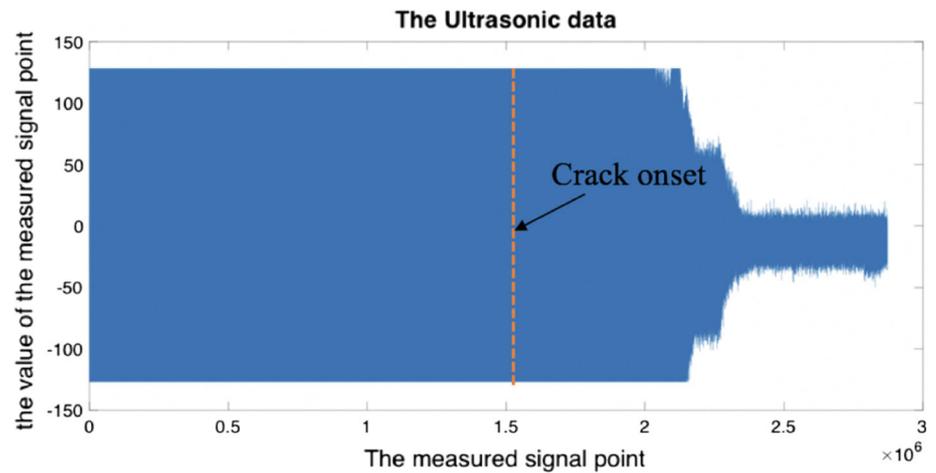
(a) Typical profile of a measured ultrasonic test (UT) signal.

(b) Typical visual inspection (VI) image of a crack onset.

propagation starts through the specimen, because a significant part of the UT signal is reflected back and hence the full-strength signal is not transmitted to the receiver [16–18].

2.2 Digital microscope

The role of the digital microscope (DM) is to provide deformation images on the specimen surface, which serves as the ground truth for quantification of fatigue damage derived from the UT signal attenuation; here the DM
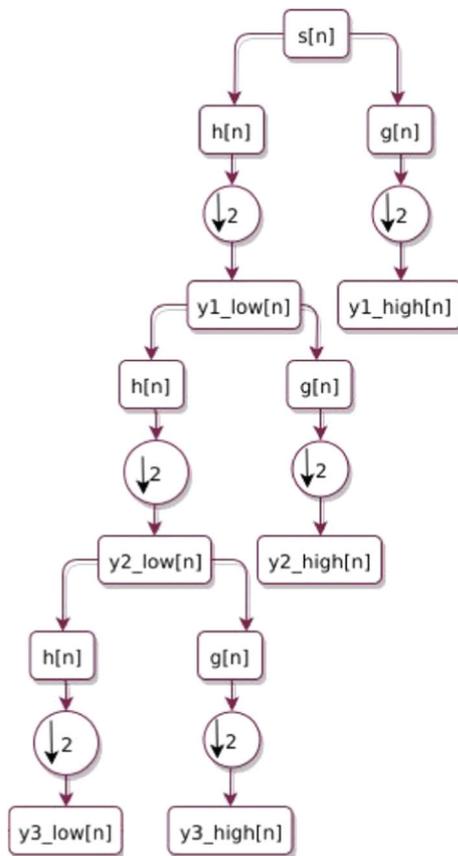
Fig. 6 Third-level dyadic representation of DWT on a two-channel filter bank

images have been taken in loose synchronism with the UT signals. The image resolution of the DM is 640×480 pixels, and the range of variable magnification of these images is 10-200X.

3 Methodology

In order to detect and classify fatigue damage in mechanical structures, the ultrasonic test (UT) signals are first preprocessed and analyzed for feature extraction. Then, the extracted features are used as inputs to the neural networks (NN)-based classifiers.

3.1 Feature extraction

Feature extraction is used to transform data (e.g., UT signals) into a set of features such that these features are distinctive properties of the patterns [19, 20] so that discrimination among the classes of these patterns becomes a feasible task. In the work reported in this paper, four different types of features have been identified for extraction

from UT signals: (i) Low frequency; (ii) High frequency; (iii) Signal envelope; and (iv) Signal energy.

The methodologies to extract the above features are explained below.

3.1.1 Discrete wavelet transform

Bulk of the pertinent information in UT signals exist in nonstationary forms and also have 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 signals have been decomposed by digital wavelet transform (DWT) [9, 10] that is suitable for analyzing nonstationary signals. Figure 5(a) displays a typical profile of (non-stationary) UT signals.

Multiresolution analysis (MRA) [9, 10], 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. The selected wavelet decomposition level that provides the best detection of the crack onset is observed to be a third-level wavelet decomposition with the Daubechies wavelet basis function, db10. The approximation and detail parts of the signal are expressed as:

$$\text{Approximation part: } y_{lo}[k] = \sum_n x[k]h[2n - k] \qquad (1)$$

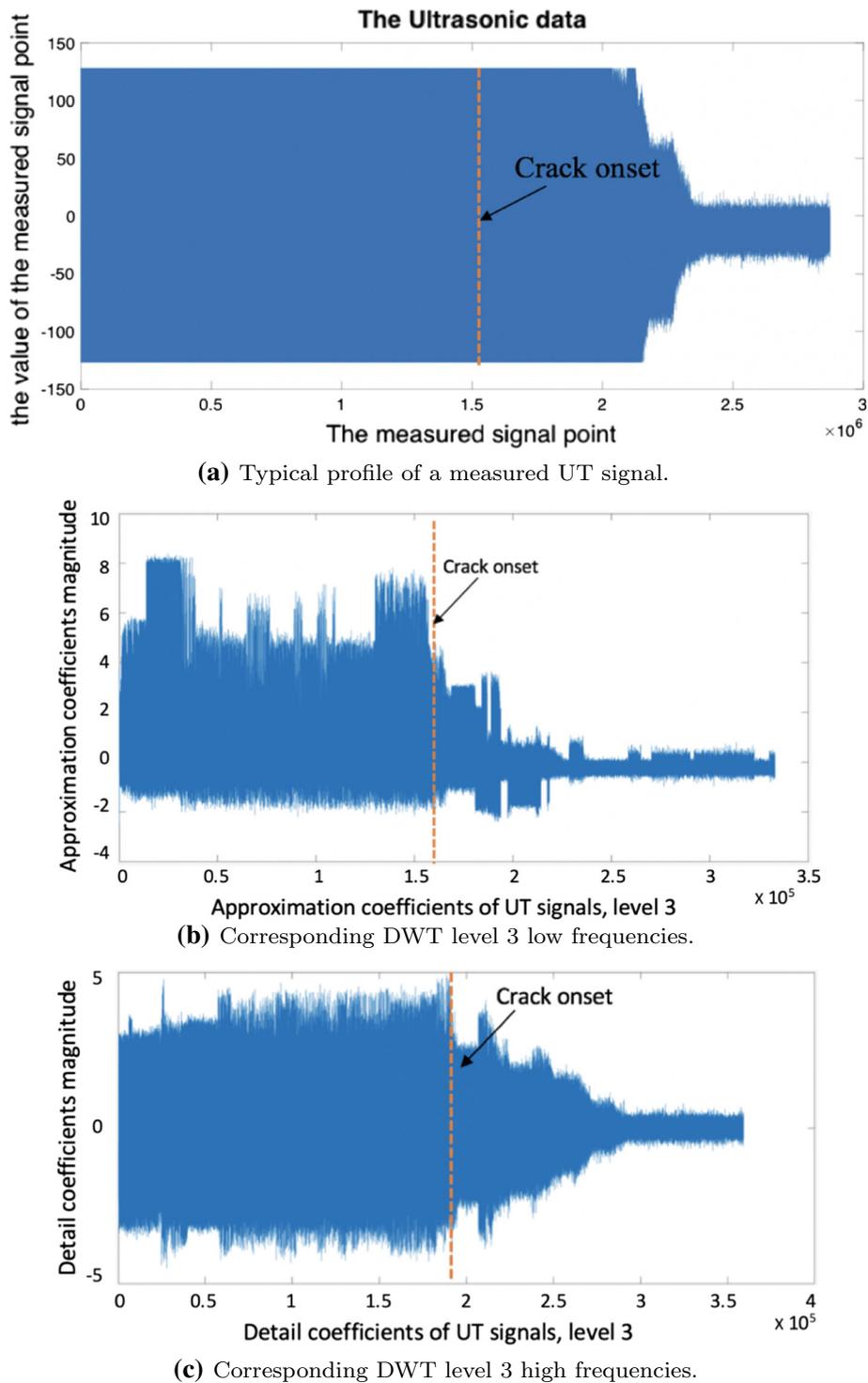$$\text{Detail part: } y_{hi}[k] = \sum_n x[k]\,g[2n - k] \qquad (2)$$

where y_{lo} and y_{hi} represent the outputs of the low-pass and high-pass filters, respectively, after down-sampling by 2. The extracted features using DWT are high-frequency components and low-frequency components [12, 21–24].

Figure 6 presents a schematic diagram of the DWT process analysis for decomposition of non-stationary signals. Fig. 7a, b, and c present the profile of a typical UT signal, the DWT level 3 low frequencies and high frequencies of UT signals, respectively.

3.1.2 Hilbert transform

Another possible method of analyzing UT signals is Hilbert transform (HT) [25]. The core concept of HT-based detection of signal envelopes is built upon the (complex-valued) analytical signal, generated from the input data. The real part of the (complex-valued) analytical signal is the original UT signal, while the imaginary part is Hilbert transform of the raw UT signal [26, 27]. In this paper, the HT method

Fig. 7 Detection for a fatigue crack by DWT analysis

(a) Typical profile of a measured UT signal.

(b) Corresponding DWT level 3 low frequencies.

(c) Corresponding DWT level 3 high frequencies.

characterizes the upper and lower boundaries of UT signals in order to provide a good visual representation of signal wave-forms for crack detection; here, the time-domain bounds of the UT signal identify the fatigue crack in early stages of crack propagation, as seen in Fig. 8(b).

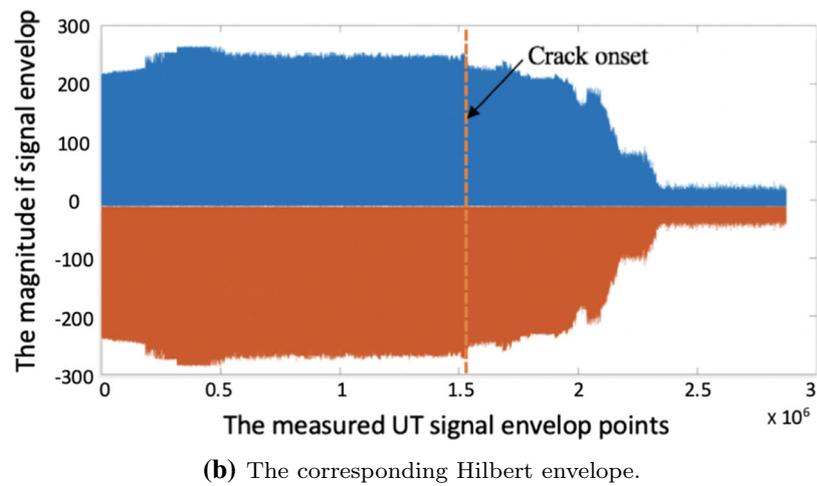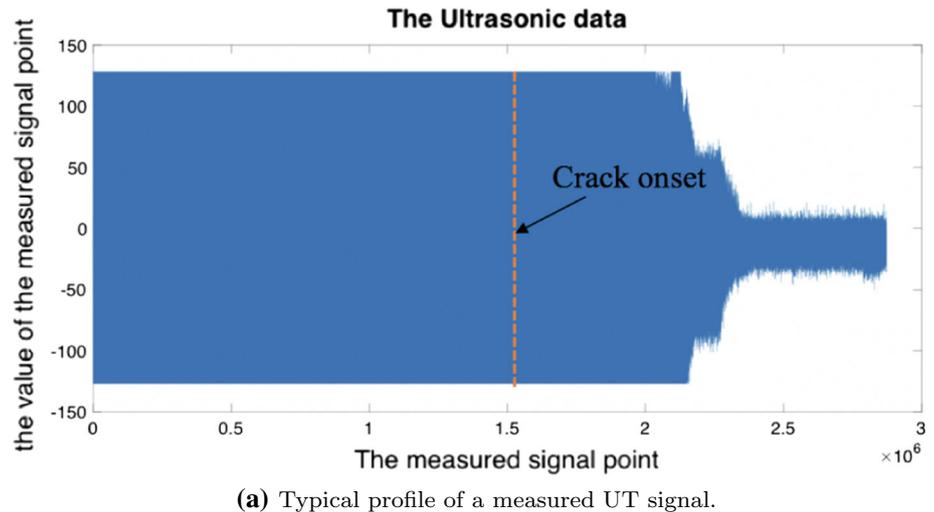The Hilbert transform, $\hat{x}(t)$, of the UT signal, $x(t)$, is obtained by the convolution of $x(t)$ with the Hilbert transform kernel $h(t)$, as illustrated below:

$$\hat{x}(t) = \int_{-\infty}^{\infty} x(\tau)\, h(t-\tau)\, d\tau \tag{3}$$

Fig. 8 Detection of a typical fatigue crack by HT analysis

(a) Typical profile of a measured UT signal.

(b) The corresponding Hilbert envelope.

The signal envelope of the UT signal $x(t)$ is used as an extracted feature here, which is defined as:

$$env_x(t) \triangleq \sqrt{|x(t)|^2 + |\hat{x}(t)|^2} \qquad (4)$$

3.1.3 The signal energy

In the discipline of signal processing, signal energy is widely used for quantifying the signal strength. In this paper, a decreasing trend in the signal strength serves to identify the onset of fatigue crack. It follows from Fig. 1a and b that, by applying Eq. (5) on UT signals, the signal energy, $E(t)$, of a crack-free specimen at the receiver transducer is seen to be greater than that of a cracked specimen, as seen in Fig. 9b.

$$E(t) = \int_{-\infty}^{\infty} |x(t)|^2 \, dt \qquad (5)$$

3.2 Neural network-based classification

Neural network (NN) is a computational tool that attempts to mimic the logic of a human brain. In its simplest form, NN is composed of a set of nodes and a set of connections that link the neurons in individual layers in the NN. Hence, an NN tends to imitate the most vital mechanism of the brain, which is the neural association. In essence, the operating principle of an NN is largely dependent on the connections, called synapses, between nodes and the distributed (nonlinear) activation functions to create different types of NN. Neural networking (NN) is a computational method that is designed to adjust and self-program the software by using learning algorithms. The mathematical derivations/equations of neural network (NN) are given in one of the authors' previous publications [28] and are therefore not repeated here. The feed-forward neural network with backpropagation is considered to be one of the most common types of NN.
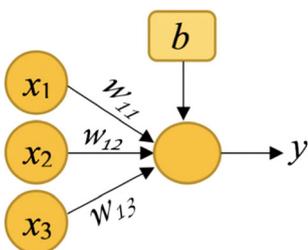
Fig. 9 Detection of signal energy for a fatigue crack
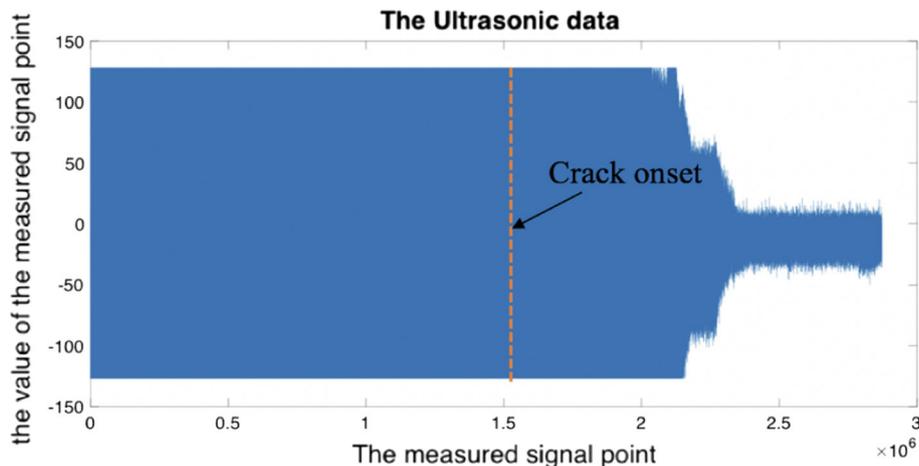
(a) Typical profile of a measured UT signal.
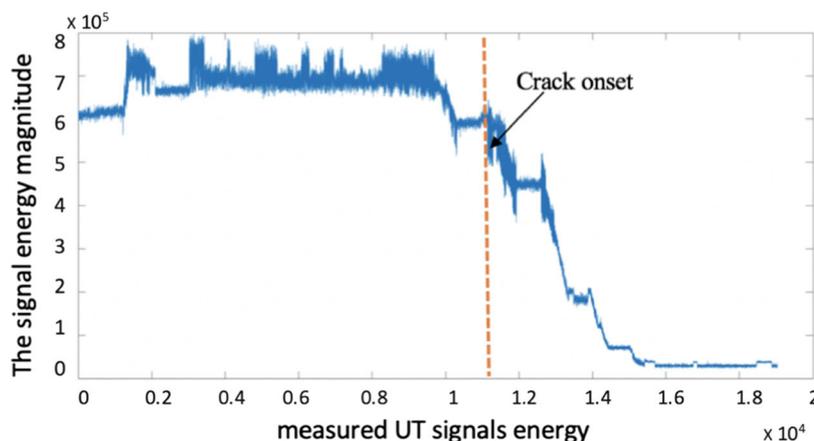
(b) The corresponding energy of the UT signal.

Fig. 10 The mathematical structure of a typical neuron

In the above context, NN has been used in several application areas, such as engineering, economics, and medicine. In its simplest form, NN is composed of a set of neurons (known as processing units), which works by building interconnections between neurons in various layers of NN; neurons are characterized by the weights that represent the connections between the neurons and thus compares the impact of one neuron relative to another. It is well known that different types of neuron connections create different types of NN systems, which are broadly categorized as follows:

1. *The NN architecture* that is represented by NN layers and a number of neurons in each layer. For example, single layer neural network presents only two layers which are the input layer and output layer, while the shallow neural network includes three layers which are the input layer, the hidden layer, and the output layer. This study applies the shallow neural network.

2. *The mathematical structure of a neuron*, which determines how a neuron should be activated. Figure 10 presents the mathematical structure of an activation function in a neuron, where, (x_1, x_2, x_3) are the input data, (w_{11}, w_{12}, w_{13}) are weights, and b is the bias. For each node, the input data are multiplied by weights, and added with a bias b as:

$$a = w * x + b = w_{11}x_1 + w_{12}x_2 + w_3x_{13} + b$$
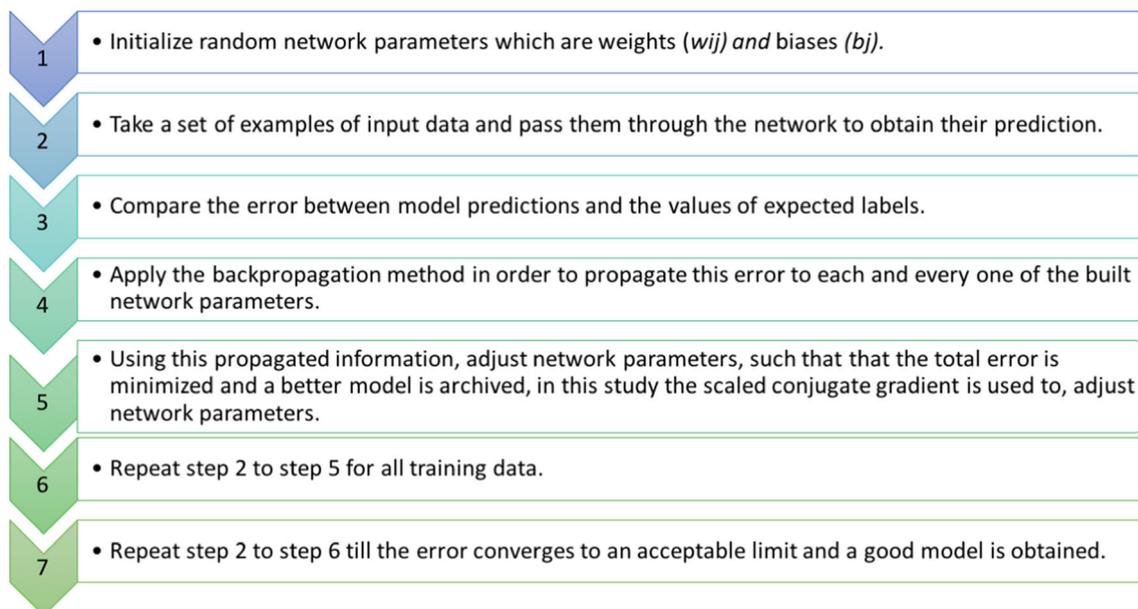$$y = \phi(a) = \phi(wx + b) \tag{6}$$

Fig. 11 The procedure of neural network analysis

Fig. 12 The structure of a confusion matrix

where ϕ is the activation function [29–32], which is used to introduce a nonlinearity into the neuron's output. In this paper, the (smooth and nonlinear) sigmoid function, which serves as the activation function, provides an analogue output between 0 and 1, The sigmoid function is expressed by:

$$y = \frac{1}{1 + e^a} \qquad (7)$$

3. *The learning mechanism*: In the learning phase of an NN model, the error between the outputs and the target values is minimized by adjusting weights (w_{ij}) of the NN model, hence, the model performance is improved. The number of NN model epochs is determined from the number of the learning phase cycles, where each cycle of the learning phase accounts for one epoch. At each epoch, the outputs of NN are compared with the target set, and according to the estimated error, the back-propagation algorithm is performed, where the estimated error is passed in the reverse direction of the NN architecture (from the output layer to the input layer) for re-adjusting the values of the weights. This procedure is repeated continuously for the next epochs until the optimal error is achieved. In this study, the scaled conjugate gradient is applied to adjust the weights of the NN architecture. For more details, the reader is referred to [28, 33–35] .

Figure 11 presents a general procedure of NN analysis, which describes the essential steps for the learning phase.

3.2.1 Performance evaluation

The task of the classifier in this paper is *binary classification*, whose outputs belong to one of the four possible options:

1. True positive (TP): TP is defined when the sample is positive and it is classified as positive.
2. False negative (FN): FN is defined when the sample is positive and it is classified as negative.
3. True negative (TN): TN is defined when the sample is negative and it is classified as negative.
4. False positive (FP): FP is defined when the sample is negative and it is classified as positive.

Figure 12 shows the associated confusion matrix. Note: A confusion matrix is a table that is often used to visualize the performance of a classification model [12].

The performance evaluation of a NN classifier is achieved by the following metrics:

1. *Accuracy* of a classifier is defined as the ratio of correct predictions to total predictions made:

Fig. 13 Undamaged and damaged (dashed area) classes of a UT signal
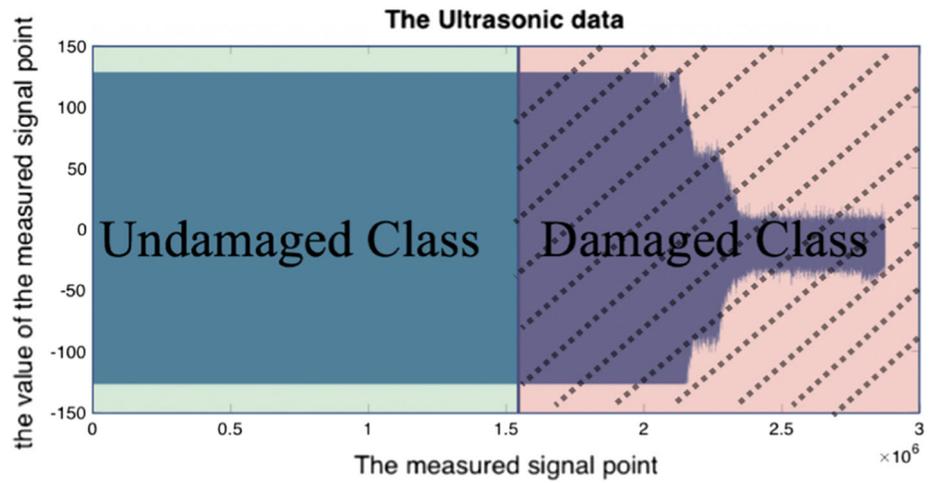
Fig. 14 Undamaged and damaged classes at low frequencies of a UT signal
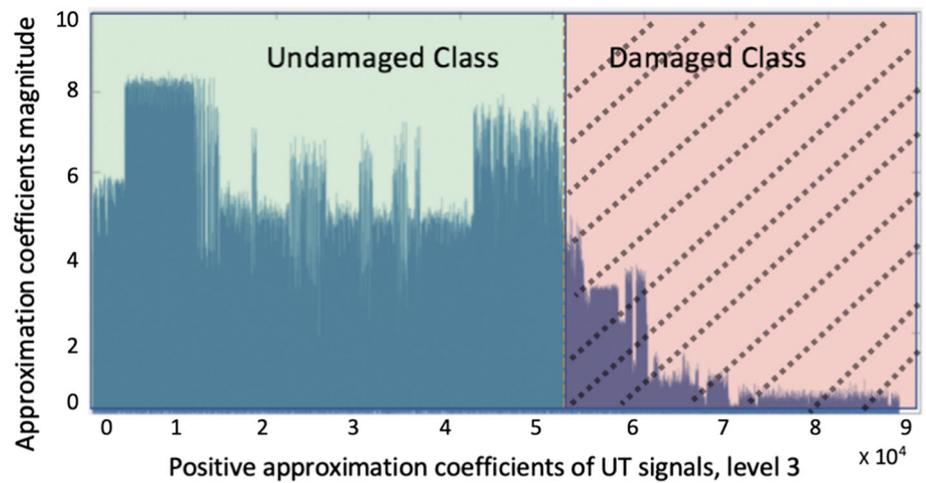
Fig. 15 Undamaged and damaged classes at high frequencies of a UT signals

$$Accuracy \triangleq \frac{TP + TN}{TP + TN + FP + FN} \quad (8)$$

2. *Sensitivity* of a classifier is defined as the ratio of correct positive predictions to the total number of positives:

Fig. 16 Undamaged and damaged classes of the signal envelope of a UT signal

Fig. 17 Undamaged and damaged classes of UT signal energy

$$Sensitivity \triangleq \frac{TP}{TP + FN} \quad (9)$$

3. *Specificity* of a classifier is defined as the ratio of correct negative predictions to the total number of negatives:

$$Specificity \triangleq \frac{TN}{TN + FP} \quad (10)$$

4. *Precision* of a classifier is defined as the ratio of correct positive predictions to the total number of the samples classified as positive:

$$Precision \triangleq \frac{TP}{TP + FP} \quad (11)$$

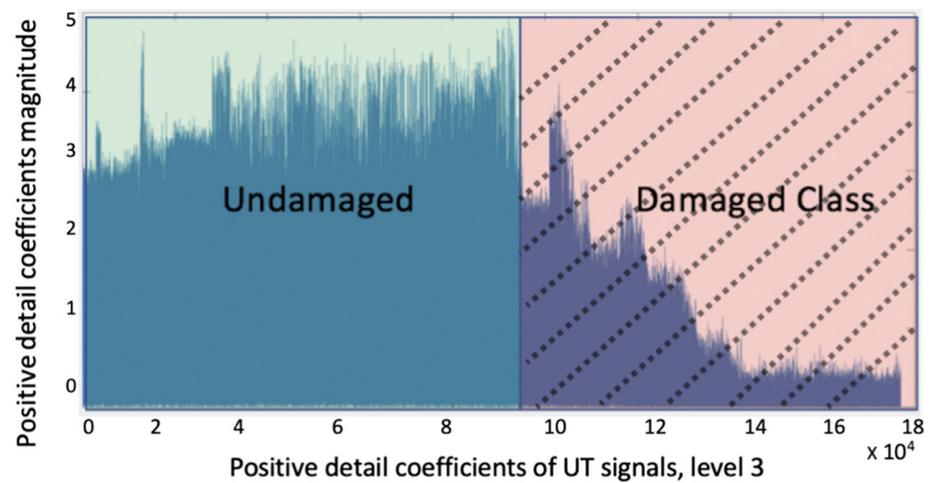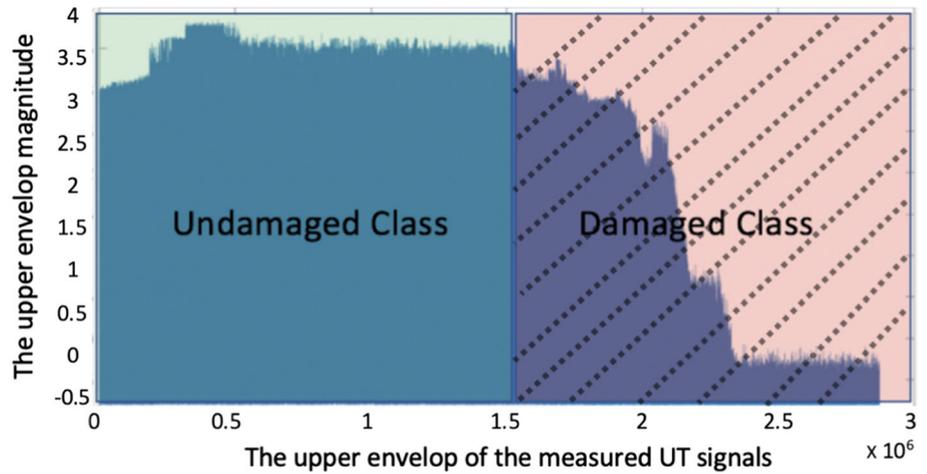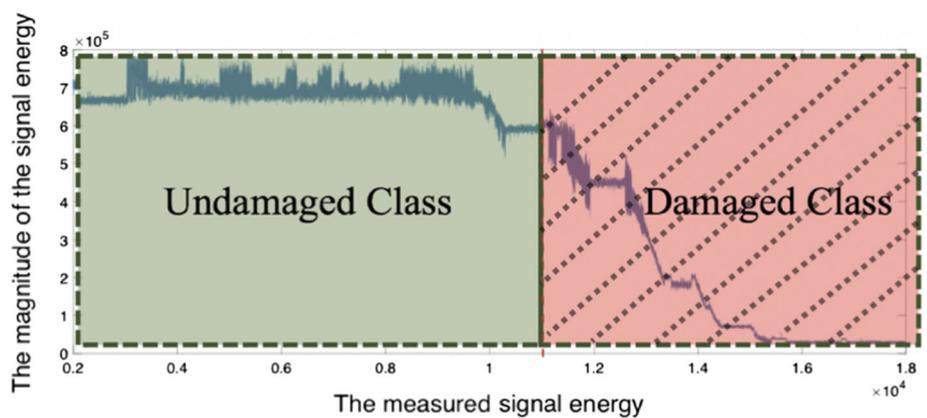5. *F-measure* of a classifier is defined as the weighted average of Precision and Sensitivity:

$$F - measure \triangleq \frac{2 * Precision * Sensitivity}{Precision + Sensitivity} \quad (12)$$

Thus, F-measure takes both false negatives and false positives into account.

6. *Receiver Operator Characteristic (ROC)* helps to quantify the best threshold value. The graph of ROC

is generated by plotting the Sensitivity (y-axis) against the False Positive Rate = (1- Specificity) (x-axis).

7. *Area under the curve (AUC)* uses the logistic model to provide the rate of successful classification. Also, it is used to compare the ROC curve of one model to that of another model.

$$AUC \triangleq \frac{Sensitivity + Specificity}{2} \quad (13)$$

8. *Matthews Correlation Coefficient (MCC)* correlates the predicted observed binary classifications to generate a score that ranges in the interval $[-1, +1]$, where $MCC = +1$ implies that the samples are perfectly classified, and $MCC = -1$ implies that the samples are perfectly mis-classified. The MCC is normalized as:

$$MCC \triangleq$$

$$\frac{TP * TN - FP * FN}{\sqrt{(TP + FP) * (TP + FN) * (TN + FP) * (TN + FN)}} \quad (14)$$

Fig. 18 Feature extracted at low frequencies of a UT signal
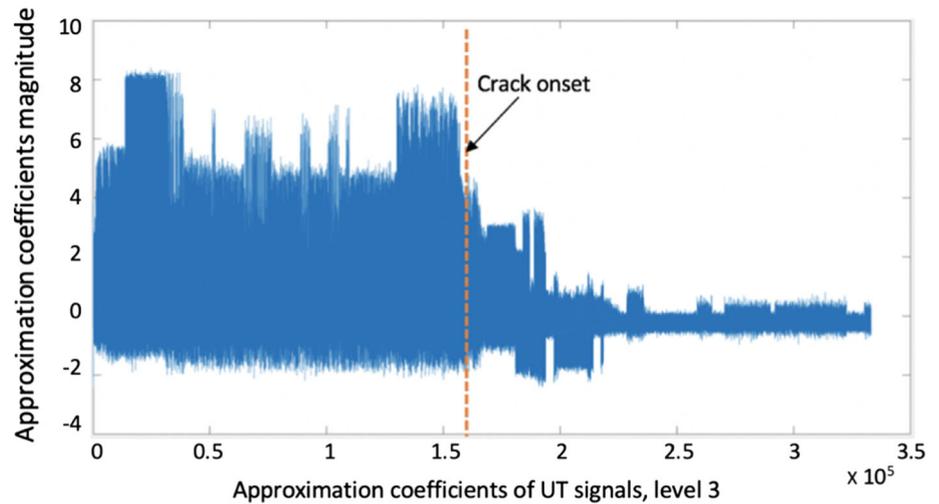
(a) Normalized approximation of UT and DWT coefficients at Level 3.
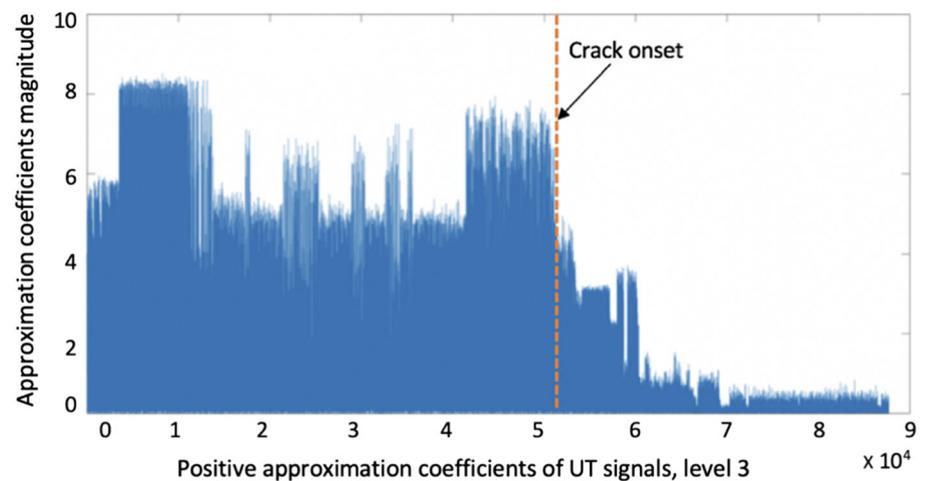
(b) Positive elements of normalized approximation coefficients of UT and DWT coefficients at Level 3.

3.3 Pattern classification

The principle of pattern recognition has been applied in this paper for damage detection and classification based on experimental data [36, 37]. The methodology of pattern classification in neural networks relies on a pair of variables $\{\phi, \lambda\}$, where the vector ϕ represents the extracted feature, and the vector λ contains the corresponding labels of ϕ. In this way, the features of UT signals are classified as the crack that appears on the notch of the specimen, where the visual inspection by the digital microscope is synchronized performed with the ultrasonic testing to provide a ground truth for the ultrasonic fatigue crack detection. All features that are identified before the crack onset belong to the undamaged class, and those that occur after the crack onset belong to the damaged class. Figures 13, 14, 15, 16 and 17 illustrate the undamaged class and damaged class (dashed area) for the original UT

signals, low frequencies of UT, high frequencies of UT, the signal envelope of UT, and the signal energy of UT, respectively.

The proposed procedure of fatigue damage classification is briefly described in the following steps:

1. Feature extraction: the low frequencies and high frequencies of UT signals were obtained by DWT. The signals envelope was obtained from UT signals by HT, and the last feature is the signal energy of UT which is obtained by applying Eq. 5.

2. *Data preparation*: The generated UT signals are split into two categories based on the appearance of crack on the surface notch. The undamaged class contains all extracted features before the onset of the crack, while the damaged class includes all extracted features after the crack onset. However, before data splitting, two

Fig. 19 Feature extracted at high frequencies of a UT signal
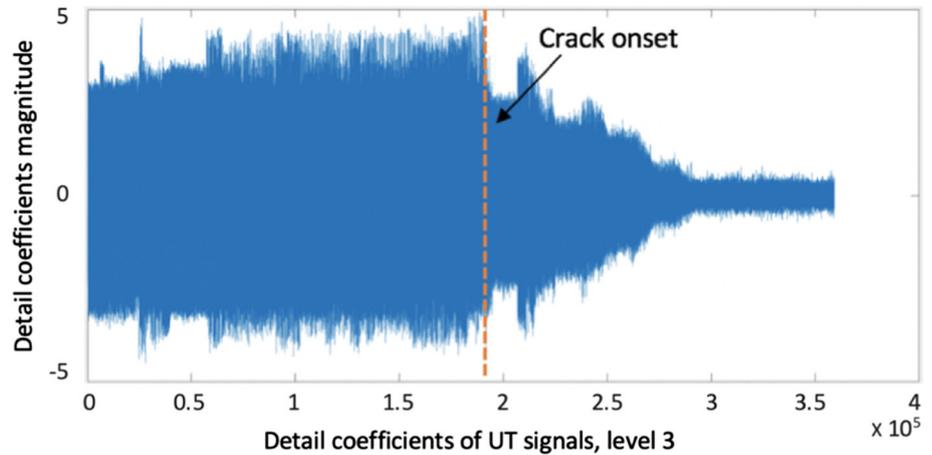
(a) Normalized details of wavelet coefficients at Level 3.
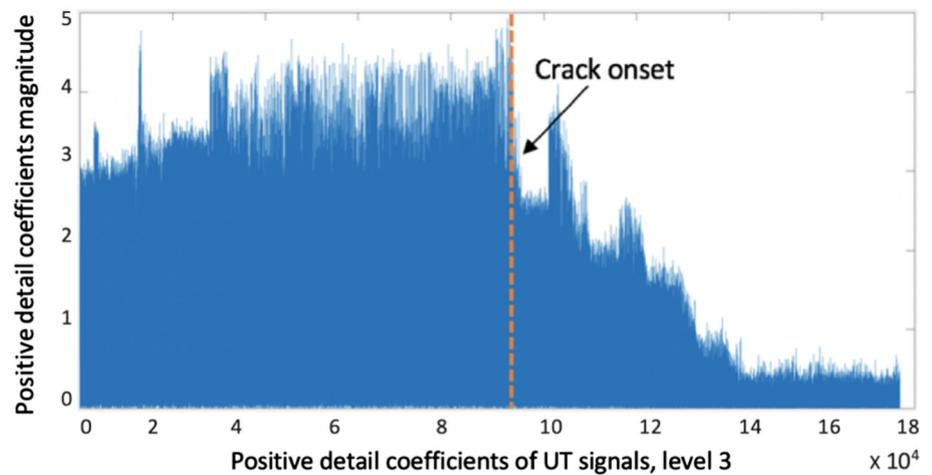
(b) Positive elements of normalized details of wavelet coefficients at Level 3.

essential steps must be executed after extracting the features from UT signals.

(a) *Data normalization*: The reason for normalizing the measured features is that their responses for all tested specimens are of similar texture, but their values may differ from one to another. The normalization is defined as a z-score, and it transforms the measured features to have a mean of zero and a standard deviation of one. The measured features are normalized as:

$$z_n = \frac{x_n - \bar{x}}{\sigma} \tag{15}$$

where $\bar{x}$ is the mean and σ is the standard deviation.

(b) *Data selection*: In this step, the significant information that perfectly represents both classes is chosen. For example, as shown in Fig. 18a, most of the positive components of the third approximation level, low frequencies of UT signals, represent the undamaged class and damaged class perfectly. Hence, the positive components of low frequencies are selected as input data for NN model, see Fig. 18b. Furthermore, the high frequencies and the signals envelop are almost symmetric at the x-axis, as shown in Figs. 19a, 20a, hence the positive components of high frequencies and signals envelop are selected as an input data for NN model, see Figs. 19b, 20b. This step also provides the advantage of reduced learning time of the NN model by removing the unnecessary input data.

(c) *Data classification*: The neural network pattern recognition toolbox of Matlab has been used to classify the extracted feature and to provide the best feature for fatigue crack detection.

Fig. 20 Feature extracted from the envelope of a UT signal
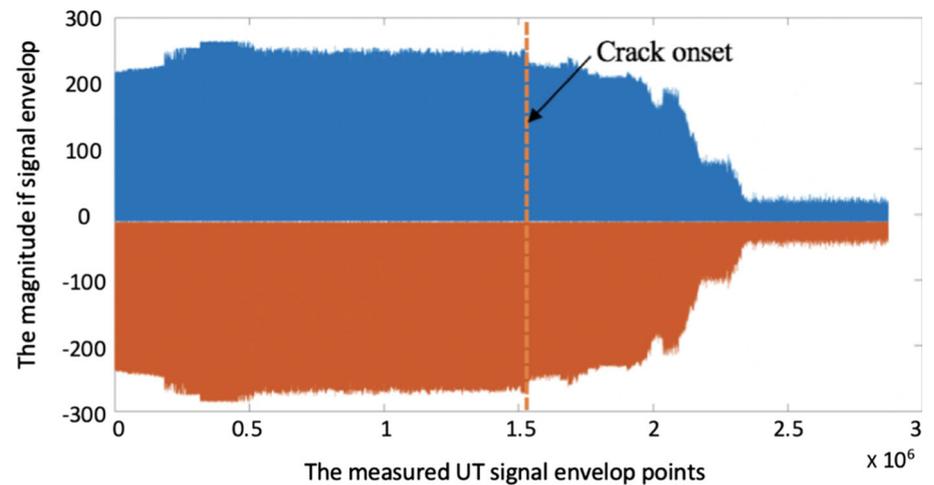
(a) A normalized signal envelope at Level 3.
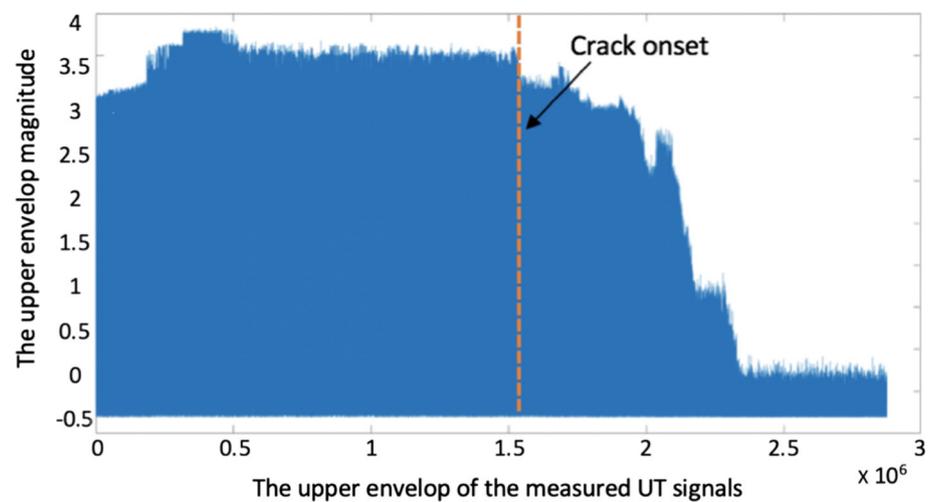
(b) Positive elements of a normalized envelope at Level 3.

4 Results and discussion

This section presents and discusses the experimental results for validation of neural network (NN) models and feature extraction methods. Upon feature extraction by various methods, the best feature of UT signals is selected based on its combination with NN-based classification. The performance of each feature classifier is measured by *accuracy, sensitivity, specificity*, and *precision* of each classifier. This is accomplished in terms of Matthews correlation coefficient (MCC), area under the curve (AUC) of ROC, and F-measure on a two-layer feed-forward network network that is constructed with ten Sigmoid hidden neurons and two Softmax output neurons by using the neural network toolbox of MATLAB. The input data of NN are split as follows: 70% of the input data for training, 15% for validation, and 15% for testing.

4.1 Performance of neural network model

The ground truth of the original experimental data (i.e., time series of UT signals and images from the digital microscope) is used as a reference to measure the performance of each combination of feature and NN model. As seen in Fig. 13, a time series of UT signals is partitioned into undamaged and damaged classes, where the input data of NN using UT signals is a vector that contains 2,873,379 elements; the fatigue crack onset occurred around the element 1,641,001. Hence, the first 1,641,001 elements are labeled as the undamaged class, while the remaining elements of the vector are labeled as the damaged class in Fig. 13.

Figure 21 presents the confusion matrix of UT signals using NN, where the correct classifications are indicated by green squares, while red squares present incorrect classifications. The right-most column of the confusion matrix presents the percentages of all the examples predicted to
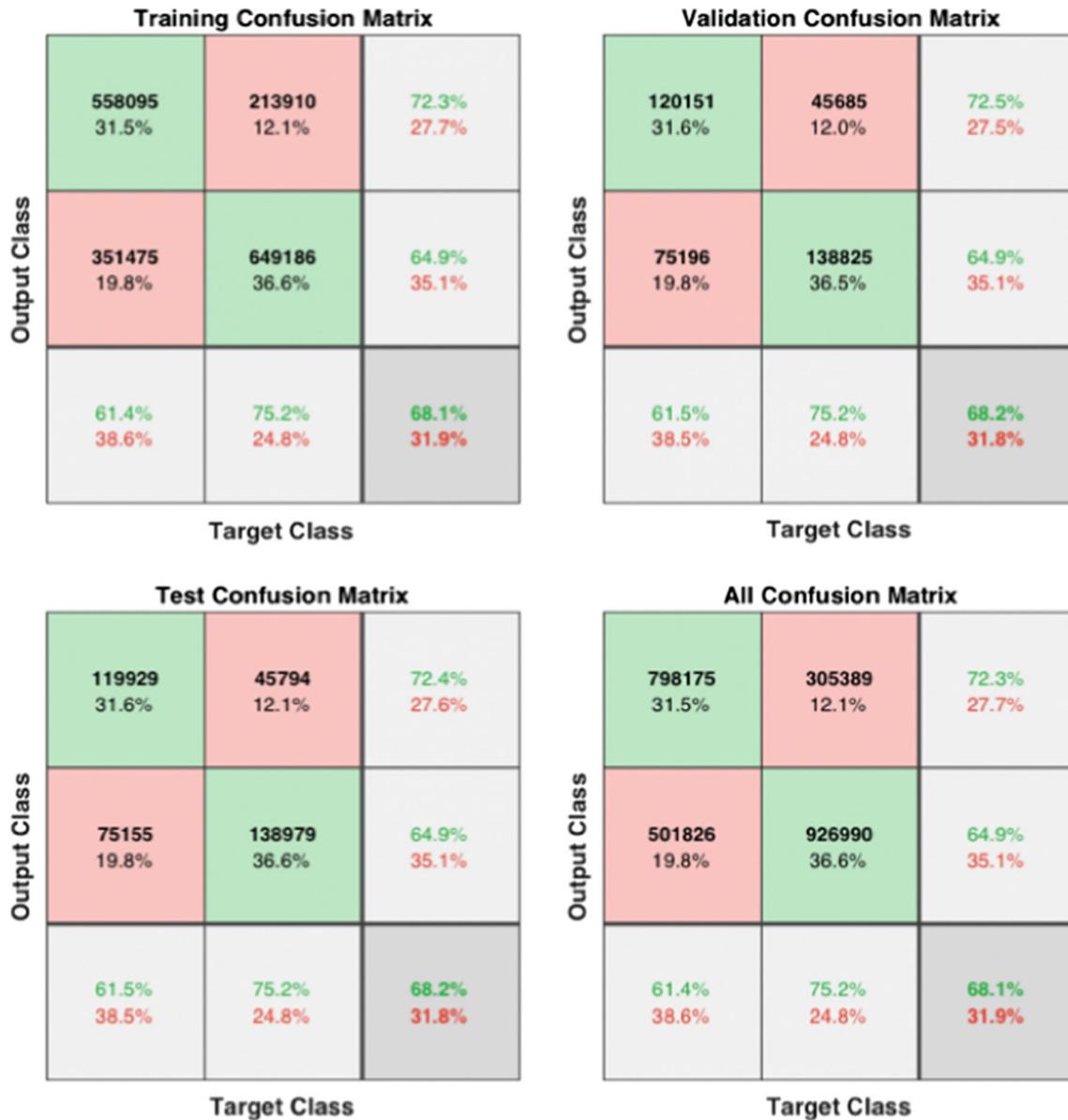
Fig. 21 Confusion matrices derived from a UT signal. In all four subplots, the three regions on each of the abscissa (from left to right) and the ordinate (from top to bottom) indicate the results of undamaged class, damaged class, and overall performance, respectively

Table 1 Classification via sole usage of the NN model

	Training	Validation	Testing	All
Accuracy	0.681	0.682	0.682	0.681
Sensitivity	0.614	0.615	0.615	0.614
Specificity	0.752	0.752	0.752	0.752
Precision	0.723	0.725	0.724	0.723
F-measure	0.664	0.665	0.665	0.664
AUC	0.683	0.684	0.683	0.683
MCC	0.369	0.370	0.370	0.369

belong to each class that are correctly and incorrectly classified. The last row of the confusion matrix shows the percentages of all the examples belonging to each class that are correctly and incorrectly classified. The cell in the bottom right of confusion matrix shows the overall accuracy. Table 1 lists the performance of the NN classification using different criteria, where accuracy of the classifier indicates that more than 30% of the measured UT signals are incorrectly classified, and that of MCC is ~ 0.37 which implies that the correlation between the ground truth and predicted responses is insufficient. (It is noted that the desired MCC should be close to 1.) Hence, the UT classification by solely using the NN-based model is

Training Confusion Matrix

757426 40.5%	275934 14.7%	73.3% 26.7%
251386 13.4%	586620 31.3%	70.0% 30.0%
75.1% 24.9%	68.0% 32.0%	71.8% 28.2%

Output Class / Target Class

Validation Confusion Matrix

162303 40.5%	59296 14.8%	73.2% 26.8%
53749 13.4%	125659 31.3%	70.0% 30.0%
75.1% 24.9%	67.9% 32.1%	71.8% 28.2%

Output Class / Target Class

Test Confusion Matrix

162026 40.4%	59643 14.9%	73.1% 26.9%
54111 13.5%	125227 31.2%	69.8% 30.2%
75.0% 25.0%	67.7% 32.3%	71.6% 28.4%

Output Class / Target Class

All Confusion Matrix

1081755 40.5%	394873 14.8%	73.3% 26.7%
359246 13.4%	837506 31.3%	70.0% 30.0%
75.1% 24.9%	68.0% 32.0%	71.8% 28.2%

Output Class / Target Class

Fig. 22 Confusion matrices of a signal envelope model. In all four subplots, the three regions on each of the abscissa (from left to right) and the ordinate (from top to bottom) indicate the results of undamaged class, damaged class, and overall performance, respectively

Table 2 Performance measurements of a HT model

	Training	Validation	**Testing**	All
Accuracy	0.718	0.708	0.709	0.709
Sensitivity	0.751	0.640	0.640	0.640
Specificity	0.680	0.788	0.790	0.790
Precision	0.733	0.779	0.782	0.781
F-measure	0.742	0.703	0.704	0.704
AUC	0.715	0.714	0.715	0.715
MCC	0.432	0.430	0.432	0.432

Table 3 Performance measurements of a DWT high-frequencies model

	Training	Validation	Testing	All
Accuracy	0.752	0.753	0.748	0.752
Sensitivity	0.749	0.750	0.746	0.749
Specificity	0.756	0.756	0.751	0.755
Precision	0.794	0.797	0.787	0.793
F-measure	0.771	0.773	0.766	0.771
AUC	0.752	0.753	0.748	0.752
MCC	0.502	0.503	0.495	0.501

Table 4 Performance measurements of a DWT low-frequencies model

	Training	Validation	**Testing**	All
Accuracy	0.728	0.727	0.727	0.728
Sensitivity	0.700	0.695	0.700	0.699
Specificity	0.756	0.758	0.755	0.756
Precision	0.738	0.737	0.740	0.738
F-measure	0.718	0.715	0.719	0.718
AUC	0.728	0.727	0.727	0.727
MCC	0.456	0.454	0.455	0.455

Table 5 Performance measurements of a typical signal-energy model

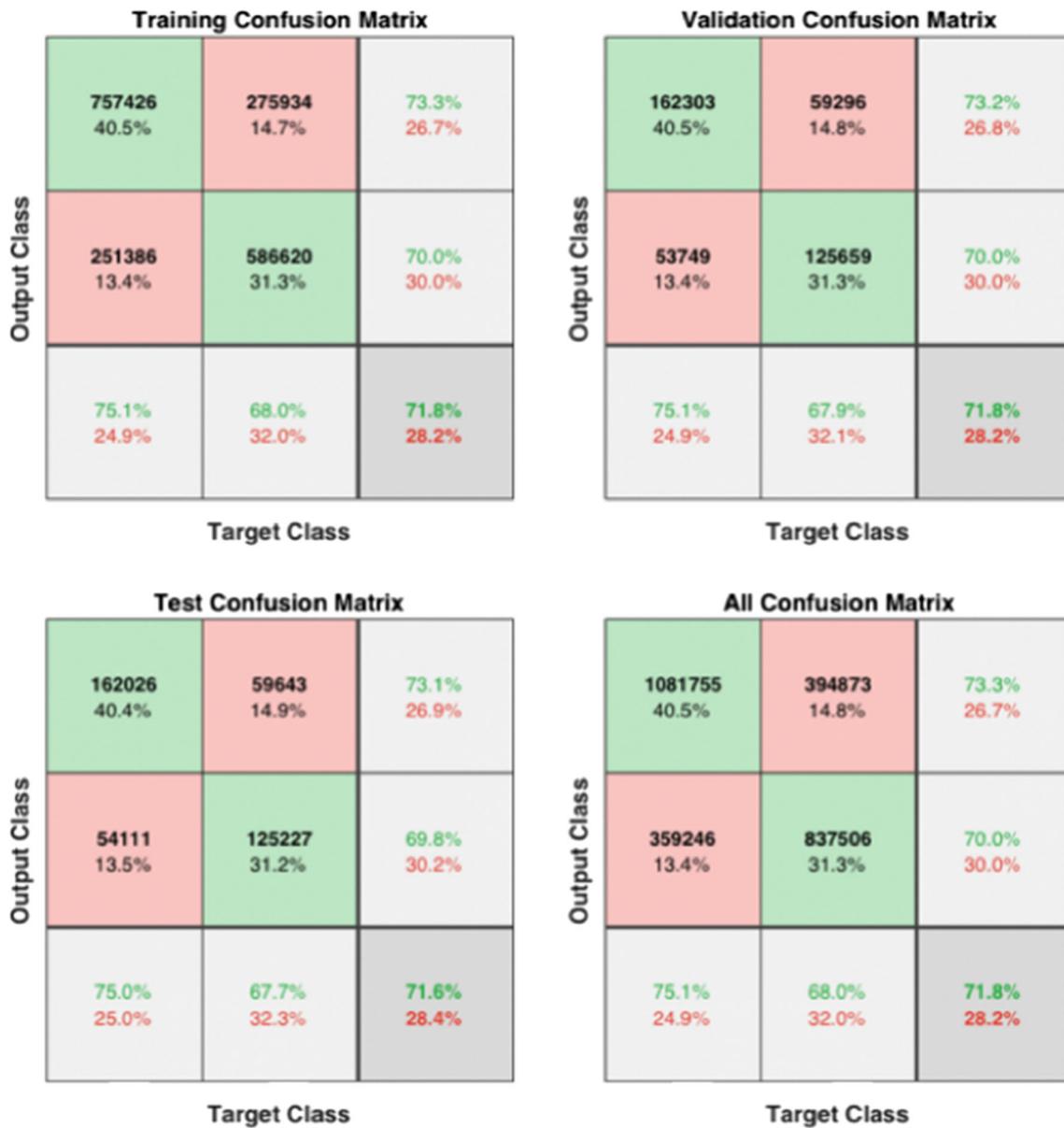|	Training	Validation	Testing	All
Accuracy	0.985	0.987	0.986	0.985
Sensitivity	0.998	0.998	0.999	0.998
Specificity	0.868	0.875	0.879	0.871
Precision	0.985	0.987	0.985	0.985
F-measure	0.992	0.993	0.992	0.992
AUC	0.933	0.937	0.939	0.935
MCC	0.916	0.921	0.927	0.919

Fig. 23 Confusion matrices of a typical high-frequencies model. In all four subplots, the three regions on each of the abscissa (from left to right) and the ordinate (from top to bottom) indicate the results of undamaged class, damaged class, and overall performance, respectively

Training Confusion Matrix

40306 35.6%	15020 13.3%	72.9% 27.1%
15768 13.9%	42059 37.2%	72.7% 27.3%
71.9% 28.1%	73.7% 26.3%	72.8% 27.2%

Output Class — Target Class

Validation Confusion Matrix

8719 35.8%	3120 12.8%	73.6% 26.4%
3487 14.3%	9001 37.0%	72.1% 27.9%
71.4% 28.6%	74.3% 25.7%	72.8% 27.2%

Output Class — Target Class

Test Confusion Matrix

8714 35.9%	3287 13.6%	72.6% 27.4%
3291 13.6%	8962 37.0%	73.1% 26.9%
72.6% 27.4%	73.2% 26.8%	72.9% 27.1%

Output Class — Target Class

All Confusion Matrix

57739 35.7%	21427 13.2%	72.9% 27.1%
22546 13.9%	60022 37.1%	72.7% 27.3%
71.9% 28.1%	73.7% 26.3%	72.8% 27.2%

Output Class — Target Class
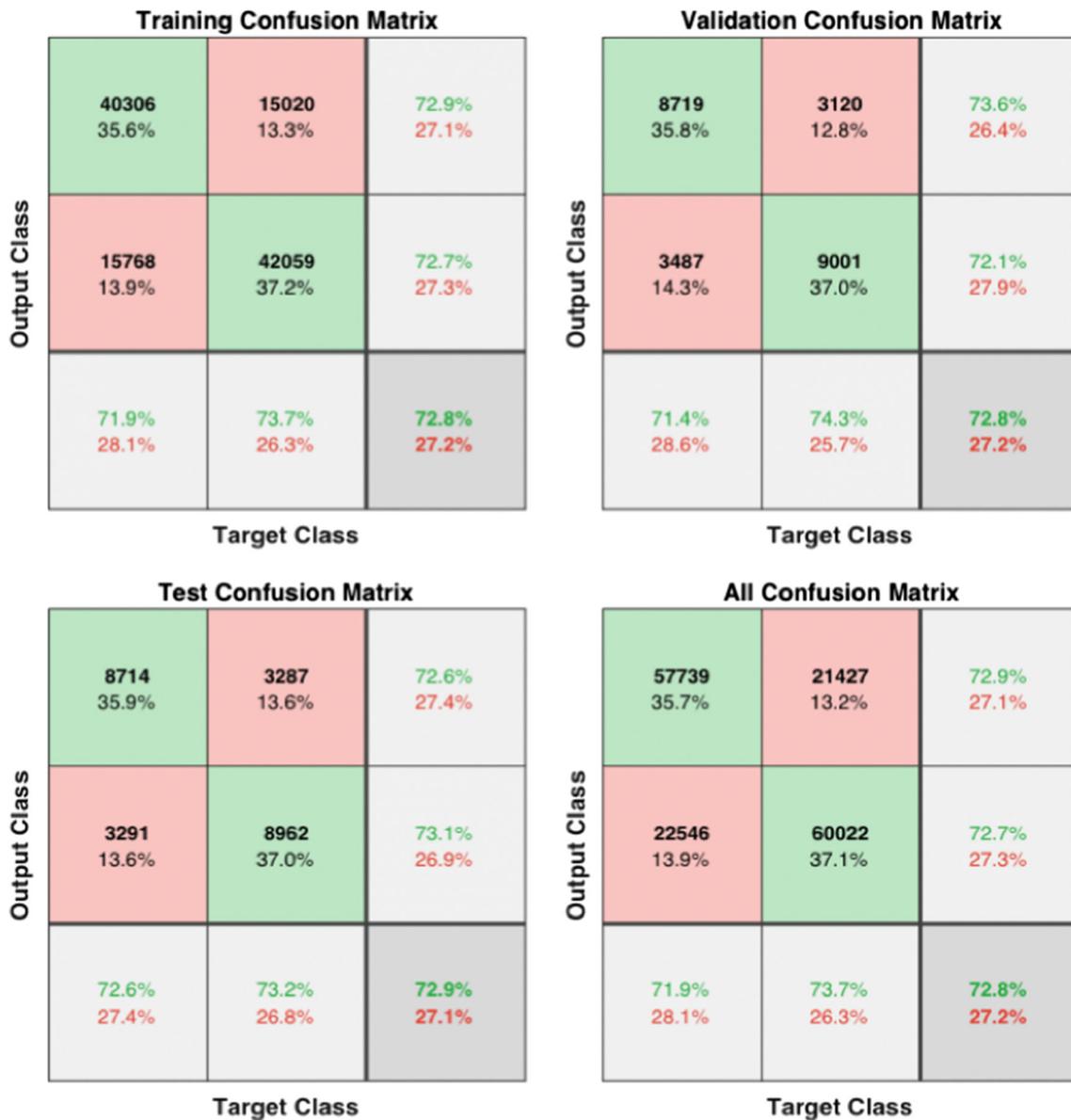
Fig. 24 The confusion matrices of a typical low frequencies model. In all four subplots, the three regions on each of the abscissa (from left to right) and the ordinate (from top to bottom) indicate the results of undamaged class, damaged class, and overall performance, respectively

inadequate for both detection and classification of fatigue damage, especially for the applications that have potentially grave consequences (e.g., human safety, colossal financial loss, and environmental contamination). The next subsection addresses how the classification performance can be improved by combining an extracted feature with the NN model.

4.2 Combination of the NN model with features

Signal envelope (see Eq. (4)) is the first feature, which is combined with the NN model, as seen in Fig. 16. The input data of NN using HT is a vector of (1×2873379), where

the first 1,641,001 elements are classified as the undamaged class, and the remaining elements are classified as the damaged class. It follows from Fig. 22 and Table 2 that the accuracy of classification is improved to $\sim 5.43\%$, and the improvement in the correlations between the target responses and predicted responses is almost negligible, ≈ 0.06. Hence, the performance of NN model using HT feature is still inadequate.

The second and the third features are DWT level 3 low frequencies and high frequencies. Equations 1 and 2 are used to extract these futures, as shown in Figs. 14, 15. The size of the input data of NN is (1×359189) for both features. The first 185,000 elements of the input data are

Fig. 25 Confusion matrices of a typical signals-energy model. In all four subplots, the three regions on each of the abscissa (from left to right) and the ordinate (from top to bottom) indicate the results of undamaged class, damaged class, and overall performance, respectively

classified as an undamaged class, while the last 174,189 are classified as a damaged class. From Tables 3 & 4, the improvement in the accuracy as compared with UT classifier is 6.90% & 10.42% for a low-frequencies feature and high-frequencies feature, respectively, and MMC is still insufficient for both features. However, the performance of the high-frequencies model is slightly better than NN model of low- frequencies. The desired performance of NN model is still not archived by using the low-frequencies feature & the high-frequencies feature.

The last feature is the signal energy, the measured signals energy are shown in Fig. 17. The input data size of the NN model is (1 × 19029). The first 9,900 elements are classified as an undamaged class, while the last 9,129 elements are classified as a damaged class. The NN model of this feature achieves an overall accuracy of 98.5 % which means the improvement in the classifier accuracy as compared with the UT classifier is increased by almost 44.46 %, and MCC is $\approx$ 0.920.92, as illustrated in Table 5, which manse the target responses are strongly correlated to the predicted responses. Therefore, the signal energy classifier is the best classifier among the other UT selected features (Figs. 22, 23, 24 and 25).

The performance comparison of NN systems

(a) The performance of NN system as measured by four methods.

The comparison of NN systems effectiveness

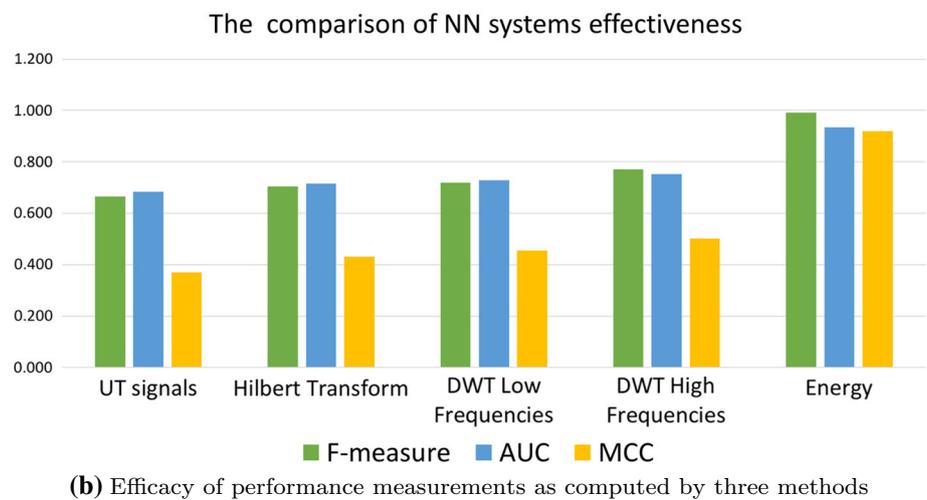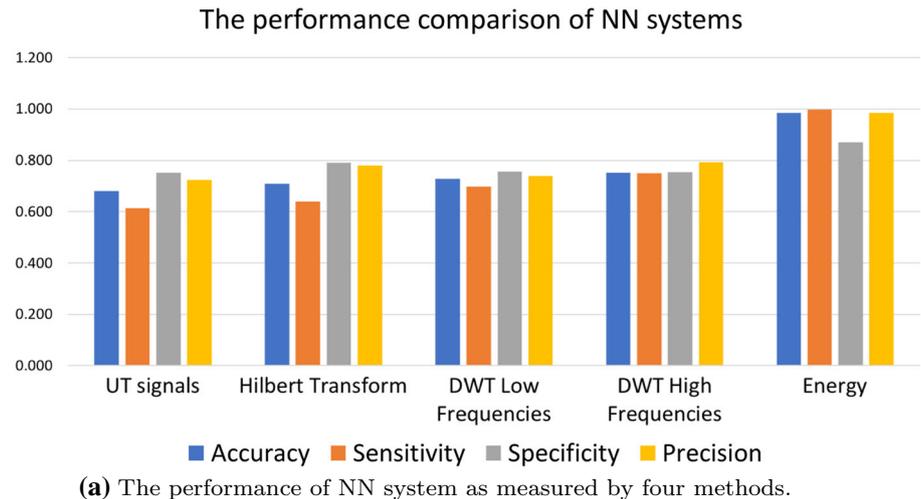**(b)** Efficacy of performance measurements as computed by three methods

Figure 26a compares the measurements tools for evaluating the performance of each feature classifier, and Fig. 26b compares the effectiveness of these tools for all NN classifiers.

5 Summary, conclusions, and future work

This paper has investigated several methods of signal processing to characterize the fatigue damage in AL7075-T6 aluminum alloy structures on a laboratory apparatus, where the main topics of research are feature extraction and pattern classification by making use of the tools of neural networks (NN). To this end, ultrasonic test (UT) signals and images from a digital microscope have been used to enhance the performance of fatigue damage detection & classification, where all features before the onset of fatigue crack on the surface notch of tested specimens are classified as the undamaged class, while all features after the onset of fatigue crack are classified as the damaged class. The feature extraction methods are constructed from

Hilbert transform (HT)-based signal envelope, discrete Wavelet transform (DWT)-based analysis at both high-frequency and low-frequency spectra, and signal energy.

The sole usage of a neural network (NN)-based model is found to be inadequate for detection and classification of fatigue damage. Therefore, this NN model is combined with features that are extracted from time series of experimental data (i.e., UT signals). The performance of these NN-based systems has been evaluated in terms of accuracy, sensitivity, specificity, and precision. Specifically, performance of the extracted features has been quantified by computing area under curve (AUC) of receiver operating characteristics (ROC) [6], Matthews correlation coefficient (MCC) [7], and the F- measure [8].

The results of this investigation reveal that the best extracted feature is the signal energy, where the classification accuracy of the NN model using this feature reaches 98.5%, and that of MCC is almost 0.92. The remaining extracted features show a slight improvement in the performance of their respective models as compared to the sole usage of the NN-based model. It is concluded that, the key to the construction of a successful NN-based model is accurate

feature extraction that has the potential of effectively detecting the fatigue damage onset in mechanical structures.

While there are many areas of both theoretical and experimental research that should be undertaken before its commercial application, the following topics are suggested for future research:

1. Examination of the detailed information at different levels of DWT.
2. Investigation of signal contents at different stages in the crack initiation regime.
3. Image analysis by more advanced apparatuses of optical metrology (e.g., confocal microscopy).

Declarations

Conflicts of interest We have no conflicts of interest to disclose.

References

1. Suresh S (1998) Fatigue of materials. Cambridge University Press, Cambridge, UK
2. Campbell FC (2008) Elements of metallurgy and engineering alloys. ASM International, The Netherlands
3. Worden K, Farrar CR, Manson G, Park G (2007) The fundamental axioms of structural health monitoring. Proc Royal Soc A: Math Phys Eng Sci 463(2082):1639–1664
4. Kumar P, Prashant K (2009) Elements of fracture mechanics. Tata McGraw-Hill Education, US
5. See JE, Drury CG, Speed A, Williams A, Khalandi N (2017) The role of visual inspection in the 21st century, in: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, Vol. 61, SAGE Publications Sage CA: Los Angeles, CA, , pp. 262–266
6. Poor H (1994) An introduction to signal detection and estimation. Spriger-Verlag, New York, NY, USA
7. Chicco D, Jurman G (2020) The advantages of the matthews correlation coefficient (mcc) over f1 score and accuracy in binary classification evaluation. BMC Genomics 21(1):1–13
8. Boughorbel S, Jarray F, El-Anbari M (2017) Optimal classifier for imbalanced data using matthews correlation coefficient metric. PloS one 12(6):e0177678
9. Kaiser G (1994) A friendly guide to wavelets. Birkhauser, Boston, MA, USA
10. Mallat S (2009) A wavelet tour of signal processing: the sparse way, 3rd edn. Academic Publishers, Amsterdam, The Netherlands
11. Alqahtani H, Ray A (2021) Forecasting and detection of fatigue cracks in polycrystalline alloys with ultrasonic testing via discrete wavelet transform. J Nondest Eval Diagn Prognostics Eng Syst 4:1–15
12. D'Angelo G, Rampone S (2016) Feature extraction and soft computing methods for aerospace structure defect classification. Measurement 85:192–209
13. Paolanti M, Romeo L, Felicetti A, Mancini A, Frontoni E, Loncarski J (2018) Machine learning approach for predictive maintenance in industry 4.0, in: 2018 14th IEEE/ASME International Conference on Mechatronic and Embedded Systems and Applications (MESA), IEEE, , pp. 1–6
14. Susto GA, Schirru A, Pampuri S, McLoone S, Beghi A (2014) Machine learning for predictive maintenance: a multiple classifier approach. IEEE Trans Ind Inform 11(3):812–820
15. Alqahtani H, Bharadwaj S, Ray A (2021) Classification of fatigue crack damage in polycrystalline alloy structures using convolutional neural networks. Eng Fail Anal 119:104908
16. Gholizadeh S (2016) A review of non-destructive testing methods of composite materials. Proc Struct Integr 1:50–57
17. Garnier C, Pastor M-L, Eyma F, Lorrain B (2011) The detection of aeronautical defects in situ on composite structures using non destructive testing. Compos Struct 93(5):1328–1336. https://doi.org/10.1016/j.compstruct.2010.10.017
18. Shull PJ (2002) Nondestructive evaluation: theory, techniques, and applications. CRC Press, Boca Raton, Florida, USA
19. Ghojogh B, Samad MN, Mashhadi S A, Kapoor T, Ali W, Karray F, Crowley M Feature selection and feature extraction in pattern analysis: A literature review, arXiv preprint arXiv:1905.02845
20. Vithlani P, Kumbharana C (2015) Structural and statistical feature extraction methods for character and digit recognition. Int J Comput Appl 120(24):0975–8887
21. Alonso GA, Gutiérrez JM, Marty J-L, Muñoz R, Implementation of the discrete wavelet transform used in the calibration of the enzymatic biosensors, Discrete Wavelet Transforms-Biomedical Applications
22. Rioul O, Vetterli M (1991) Wavelets and signal processing. IEEE Signal Proc Mag 8(4):14–38
23. Oliveira MO, Bretas AS (2009) Application of discrete wavelet transform for differential protection of power transformers, in. IEEE Buchat PowerTech IEEE 2009:1–8
24. Mane AR, Biradar P, Shastri P (2015) Review paper on feature extraction methods for eeg signal analysis. Department of electronics and telecommunication engineering. VPCOE/Savitribi Phule University, IJEEBS, pp 2349–6967
25. Cohen L (1995) Time-frequency analysis. Prentice-Hall PTR, Upper Saddle River, NJ, USA
26. Klingspor M(2015) Hilbert transform: Mathematical theory and applications to signal processing
27. Kschischang FR (2006) The Hilbert transform. University of Toronto, Toronto, Canada 83:277
28. Alqahtani H, Ray A (2020) Neural network-based automated assessment of fatigue damage in mechanical structures. Machines 8(4):85
29. Coolen AC (1998) A beginner's guide to the mathematics of neural networks. Concepts for neural networks. Springer, NY, pp 13–70
30. Zakaria M, Al-Shebany M, Sarhan S (2014) Artificial neural network: a brief overview. Int J Eng Res Appl 4:7–12
31. Islam M, Chen G, Jin S (2019) An overview of neural network. Am J Neural Netw Appl 5(1):7–11
32. A. Zayegh, N. Al Bassam (2018) Neural network principles and applications, in: Digital Systems, IntechOpen
33. Bishop CM (2006) Pattern recognition and machine learning. Springer, Berlin
34. Møller MF (1993) A scaled conjugate gradient algorithm for fast supervised learning. Neural Netw 6(4):525–533
35. Dkhichi F, Oukarfi B (2014) Neural network training by gradient descent algorithms: application on the solar cell. Int J Innov Res Sci Eng Technol 3:15696–15702
36. horpade S, Ghorpade J, Mantri S (2010) Pattern recognition using neural networks. Int J Comput Sci Inform Technol (IJCSIT) 2(6):92
37. Ghotkar VN, Indurkar P, Chilke B (2015) The pattern recognition using spiking neural network and neural network. Int J 5:ISO 690