Machine Learning with Applications 7 (2022) 100247

Contents lists available at ScienceDirect

Machine Learning with Applications

journal homepage: www.elsevier.com/locate/mlwa

Fatigue damage detection of aerospace-grade aluminum alloys using feature-based and feature-less deep neural networks

Susheel Dharmadhikari [a], Amrita Basak [b,*]

[a] Department of Mechanical Engineering, Pennsylvania State University, 309 Reber Building, University Park, PA 16802, United States
[b] Department of Mechanical Engineering, Pennsylvania State University, 233 Reber Building, University Park, PA 16802, United States

ARTICLE INFO

Keywords:
Fatigue crack detection
Deep learning
Ultrasonics
Structural health monitoring
Neural networks
Autoencoders

ABSTRACT

Fatigue damage is one of the most common causes of failure in aerospace structural components. While numerical modeling and laboratory-scale experimentation provide much insight to the physics of failure evolution, it is extremely challenging to account for all variabilities that a component may be subjected to during on-field operation. Human-supervised continuous monitoring of such components using sensors, therefore, provides a much-needed alternative for reliable operation of these components. By leveraging the concepts in deep learning, such human supervision can be assisted with an automated pre-trained deep network for damage detection. To that end, this article studies two distinct deep neural network (DNN) architectures for fatigue damage detection in aluminum using ultrasonic time-series data obtained from a novel customized fatigue testing apparatus. The first DNN, called as a feature-based network, is built by using two predefined features viz. frequency domain amplitude and autocorrelation from the ultrasonic data as the inputs. The second DNN, called as a feature-less network, uses the ultrasonic data as-is without any pre-processing and relies on the black-box features generated during training. The capability of fatigue crack detection for both DNN architectures is evaluated at two distinct stages of fatigue failure. The feature-less network is observed to outperform the feature-based network with an accuracy of 94.26% and 98.94% for the two stages. The result indicates that feature-less DNNs, owing to their construction, can formulate better, albeit black-box features, and simplify the process of choosing customized signal processing methods for similar problems.

1. Introduction

Fatigue failure is one of the most critical damage mechanisms observed in metals (Sofronas, 2012). The components that are subjected to fatigue damage, often seem to be operating under safe loads. However, due to periodic loading, a cumulative damage build-up causes internal changes to the material structure and may go unnoticed if not targeted during maintenance. This internal damage manifests itself in the form of a crack and leads to a catastrophic fracture (Suresh, 1998). With an effort to avoid such occurrences, fatigue damage has been extensively studied leading to comprehensive modeling and computational approaches to predict fatigue lives (Sangid, 2013; Stephens, Fatemi, Stephens, & Fuchs, 2000; Suresh, 1998). However, while in operation such components are subjected to several undeterministic factors such as unpredictable environmental conditions, corrosion, inadequate handling, among others (Molent & Dixon, 2020; Sofronas, 2012). It is, therefore, often advisable to assist the analytical predictions with sensor-based data-driven tools. In fact, in a recent review paper, Yuan, Zargar, Chen, and Wang (2020) argues that with growing complexity of a structure, the paradigm of anomaly detection is weighted towards data-driven techniques with system-physics operating in an assisting role. Coupling this insight with the recent developments in deep learning (Goodfellow, Bengio, Courville, & Bengio, 2016), the problem of data-driven fatigue crack detection has been reinvigorated in a new light in the past few years.

In general, the prevalent literature for metal fatigue crack detection (or crack detection in metals under a broader domain), can be divided into two approaches based on the type of sensors used for probing, viz., imaging devices and time-series-based sensors (Munawar, Hammad, Haddad, Soares, & Waller, 2021; Zhao, Yan, Chen, Mao, Wang, & Gao, 2019). Accordingly, the methods to incorporate deep learning models have also varied with a number of review articles collectively summarizing the vast extent of both approaches (Bao, Chen, Wei, Xu, Tang, & Li, 2019; Munawar et al., 2021; Salehi & Burgueño, 2018; Zhao et al., 2019). For instance, through one review paper, Zhao et al. (2019) has shown the applications of Auto-Encoders, Restricted Boltzmann Machines, Deep Boltzmann Machines, Convolutional Neural Nets (CNN), and Recurrent Neural Nets. Imaging-based devices have mainly relied on the use of traditional machine learning tools or CNNs for training models (Munawar et al., 2021; Yuan et al.,

* Corresponding author.
E-mail addresses: sud85@psu.edu (S. Dharmadhikari), aub1526@psu.edu (A. Basak).

https://doi.org/10.1016/j.mlwa.2021.100247
Received 9 September 2021; Received in revised form 1 November 2021; Accepted 27 December 2021
Available online 29 December 2021

2020). Kong and Li (2018) proposed a crack detection framework through motion tracking using computer vision for steel components. For aluminum alloys, Zhang, Wang, Wang, Zhang, Chen, and Meng (2021) proposed a computer-vision-based decision tree approach to crack detection. Munawar et al. (2021) reviewed the recent trends in image-based crack detection and commented about the focus of transfer learning in CNN-based approaches. Accordingly, through AlexNet, VGGNet13, and ResNet18, Yang, Chen, Li, and Huang (2021), successfully demonstrated the efficacy on crack detection. Similarly, Dung, Sekiya, Hirano, Okatani, and Miki (2019) successfully showed the application of VGGNet13 in detecting cracks on welded joints of steel bridges. Although incorporating transfer learning, has compensated for the shortcoming of low testing data, imaging-based methods suffer from a major drawback in real-time setting. For complex structures such as automobiles, aerospace equipment, or nuclear sites, sophisticated microscopes cannot be implemented beyond a lab-based study (Sofronas, 2012). These applications, therefore, stay limited to larger components.

Time-series based sensors have found success in eliminating such shortcomings. Guided-wave based sensors such as ultrasonics (Bhattacharya, Dharmadhikari, Basak, & Ray, 2021), acoustic emissions (Karimian, Modarres, & Bruck, 2020; Papazian, Nardiello, Silberstein, Welsh, Grundy, Craven, et al., 2007), strain gauges (Mylonas, Tsialiamanis, Worden, & Chatzi, 2022), laser vibrometers (Leong, Staszewski, Lee, & Scarpa, 2005) among many others have shown remarkable successes in identifying smaller cracks during fatigue failures. Processing such time-series data with deep learning has also shown encouraging results (Azimi, Eslamlou, & Pekcan, 2020; Yuan et al., 2020). The primary focus, while dealing with time series data for deep learning, has been towards demonstrating success through appropriate feature generation. For instance, Tripathi, Anowarul, Agarwal, and Prasad (2019) showed the capability of fatigue crack detection using ultrasonic sensors by building a machine learning model on three manually defined features. Xu, Yuan, Chen, and Ren (2019) used a CNN-based approach on guided-wave data to identify fatigue cracks in aerospace structures. Nima and Shila (2020) built a deep learning model on features extracted through wavelet transforms to detect cracks in a rotor-bearing system. Using a Bayesian Graph Neural Net, Mylonas et al. (2022) demonstrated the process of strain-based damage localization. Aria, Lopez Droguett, Azarm, and Modarres (2020) proposed a deep learning framework by combining imaging-based data and time-series-based sensor network for crack detection in aluminum 7075 alloy. Ebrahimkhanlou and Salamone (2018) used an Auto-Encoder to process acoustic emission data whereas Ewald, Groves, and Benedictus (2019) developed a DeepSHM structure to process ultrasonic signals using CNNs. Needless to say, the applications of deep learning to this field are being researched with a strong impetus.

Perusing through the literature, some common themes among all above studies can be extracted. All studies have focused on either showing deep learning capabilities at different stages of failure, or have targeted towards quantifying the capabilities in terms of crack lengths. For example, Aria et al. (2020), Lim and Sohn (2018), Xu et al. (2019) have shown accuracies in predicting the remaining useful life which correlates to different stages in failure. Similarly, Aria et al. (2020) and Lim and Sohn (2018) have shown successful detection for crack lengths of 250 μm and 1.5 mm, respectively. In addition to that, from a data analysis perspective, almost all articles have significantly focused on explaining the various methods for feature extraction from the available data to fit into a deep learning structure. With these observations in the context of deep learning and fatigue crack detection, two main questions can be posed:

- How capable are deep learning models at predicting fatigue failures at different stages?
- Can such models simplify the prevalent manual feature selection process?

Undeniably, the questions are quite broad and a comprehensive answer through one research article seems implausible. Nevertheless, with disjoint studies, the answers can be formulated in time. In a smaller domain, however, this research article addresses the two aforementioned questions. This article expands the application of deep learning for fatigue damage detection in aerospace-grade aluminum alloys using high-frequency ultrasonic time-series data. As a step toward this objective, Al7075-T6 specimens are tested on a custom fatigue testing apparatus. While this apparatus is integrated with two imaging microscopes, the specimens are instrumented with a pair of ultrasonic transducers. During fatigue testing the data from the ultrasonic transducers are continuously recorded using a data acquisition system and, thereafter, segmented into *healthy* and *cracked* regimes using the visual information obtained from both microscopes. The usage of two microscopes, having significantly different resolutions, facilitates the segmentation of ultrasonic signal in two different ways. Such a segmentation is important from a prognostics perspective – the earlier the detection, the better it is to plan maintenance tasks or take precautionary measures (Farrar & Lieven, 2007). This research, thus, aims at quantifying the capability of detecting damage at multiple stages of fatigue damage (Dharmadhikari, Keller, Ray, & Basak, 2021c). To analyze the data for damage detection, two categories of deep neural networks (DNN) are constructed, viz. a feature-based DNN and a feature-less DNN. The feature-based DNN is built upon features that are manually extracted from the ultrasonic time-series data. The feature-less network is built using an encoder-inspired framework which can compare the capability of the DNN to formulate features by itself. The performance of both the types of networks is quantified for two distinct stages of the crack detection enabled by the dual microscope system. Using the terminology from a recent review paper on time-series classification (Fawaz, Forestier, Weber, Idoumghar, & Muller, 2019), the two aforementioned deep networks can be categorized into discriminative models which are either relying on human-engineered inputs or end-to-end deep learning. Through the analysis reported in this article, two main contributions are achieved:

- Fatigue crack detection with an accuracy of 94.26% is achieved at the first stage (45% fatigue life), and 98.94% is achieved at the second stage (58% fatigue life) with feature-less networks.
- The feature-less networks are observed to outperform the feature-based networks.

Both of these conclusions offer important insights when compared to the existing literature. From a mechanics standpoint, the damage detection at first stage is shown to occur before the appearance of any cracks unlike the depictions in similar research (Aria et al., 2020; Lim & Sohn, 2018). The crack detection shown in this paper is considered to be originating in the 'crack coalescence' regime as compared to the 'crack propagation' regime observed in the prevalent literature (Dharmadhikari, Bhattacharya, Ray, & Basak, 2021b). From a data analysis perspective, by comparing the proposed approach to similar studies with symbolic analysis (Bhattacharya et al., 2021; Dharmadhikari et al., 2021b), an improvement of about 10%–15% is observed. From the success of the feature-less networks, a step towards the 'domain-invariant' approaches (Fawaz et al., 2019) can be pursued for metal fatigue crack detection. However, as often observed with DNNs, the black-box nature of the networks limit the physical understanding of the process. Although not studied in this article, deriving physics-based correlations to such approaches can expedite their reliable implementation in safety-critical environments. In the future, with proven reliability on such data-driven damage detection systems, the reliance on human-supervision, and human-error thereof, can be minimized. And the field can approach towards truly automated damage detection.

The article is divided into five sections including the current one. The second section explains the experimental method used for generating the required data. The third section describes the nature of the data and elaborates the data labeling criteria. The fourth section details the

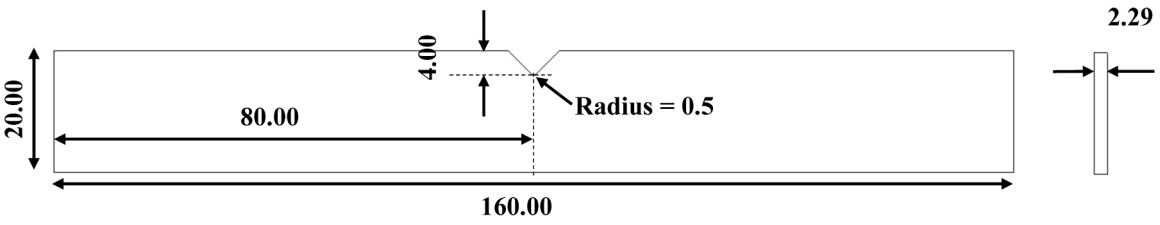

Fig. 1. V-notch specimen design (all dimensions are in mm).

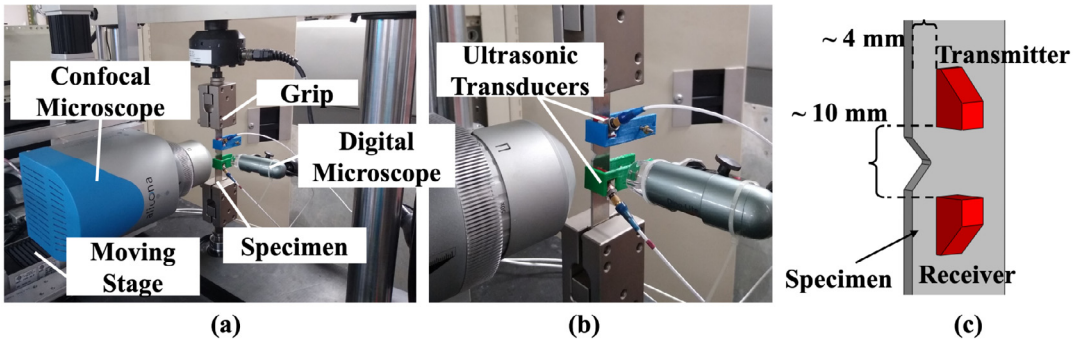

Fig. 2. (a) and (b) Custom fatigue testing apparatus highlighting the important components. (c) Positioning of the ultrasonic sensors on the specimen.

DNN strategies used for the analysis followed by the fifth section documenting and discussing the results. Finally, the main outcomes from the study are summarized in the final section along with recommendations for future work. The conclusion also comments on the shortcomings of the study, particularly in regards to the black-box nature of deep learning models.

2. Experimental method

The experiments are performed on V-notched specimens made of an aluminum alloy, Al7075-T6. The alloy is chosen due to its applications in the aerospace industry owing to its excellent mechanical properties. The design of the specimen, according to ASTM standard E466 (ASTM, 2015), is shown in Fig. 1. The V-notched design on the edge of the specimen incorporates a stress concentration near the base of the notch and is meant to enable a controlled initiation of fatigue cracks for imaging purposes (Dharmadhikari & Basak, 2021; Dharmadhikari et al., 2021c). The specimens are fabricated by waterjet machining from cold-rolled Al7075-T6 sheet obtained from McMaster-Carr. The fatigue testing apparatus is shown in Fig. 2(a) and (b). The apparatus consists of four components, viz. (i) the MTS testing equipment, (ii) the confocal microscope mounted on a moving stage, (iii) the digital microscope, and (iv) the ultrasonic sensors. The MTS testing equipment is a servo-hydraulic setup that is rated at 25 kN. The specimens are mounted on the equipment and are subjected to a tensile–tensile fatigue loading at 20 Hz with a maximum load of 4 kN, and a stress ratio (min. stress (MPa)/max. stress (MPa)) of 0.5. During the test, periodical data collection is conducted through the microscopes and the ultrasonic sensors. The tests are controlled through an MTS controller using an automated routine available in the Multi-Purpose TestWare® software suite. Through this routine, the MTS system is controlled to stop every 500 cycles at the maximum load.

Fig. 2(c) schematically shows the positioning of the ultrasonic transmitter and receiver used during the fatigue test. For all tests, the positions of the transducers are identical. For the current experiments, the sensors (Olympus, Shinjuku, Japan) have a base frequency of 10 MHz and the received signals are sampled at 100 MHz. The data collected from the sensors is in the form of time-series. Although ultrasonic sensors are excellent for damage detection, it is challenging to pinpoint the exact instance of crack appearance just by analyzing the time series data. Therefore, in a laboratory-based testing, visual sensors such as microscopes provide the necessary calibration insight. For the current experiment, two microscopes (digital and confocal) of varied resolutions are used in two perpendicular orientations (Fig. 2(a)) to enable earlier crack detection (Dharmadhikari et al., 2021c). A schematic of the two microscopes with their respective orientations is shown in Fig. 3(a). The confocal microscope with a higher resolution is focused on the base of the notch as fatigue cracks have been found to originate inside the notch root (Dharmadhikari et al., 2021c). The microscope (Model: IF-SensorR25) belongs to the InfiniteFocus series of Bruker Alicona (Graz, Austria). The region of interest for the microscope coupled with the actual images are shown in Fig. 3(b). The microscope has sufficient resolution to capture a crack opening displacement (COD) of ~ 3 μm as shown in Fig. 3(c). The digital microscope which is more commonly employed in traditional fatigue analyses (Papazian et al., 2007) is focused on the side of the specimen as depicted in Fig. 3(a) The microscope model used in the study is Dino-Lite Premier, a commonly used USB microscope. As compared to the confocal microscope, this microscope has a much lower resolution and is able to capture a crack length ~ 0.1 mm. Fig. 3(d) and (e) show the region of interest for the digital microscope and a cracked surface respectively.

3. Dataset description, classification, and featurization

3.1. Ultrasonic signal and binary classification

A total of 15 specimens are tested using the experimental protocol explained in Section 2. The specimens exhibit an average fatigue life of 60,000 cycles. During the testing of each specimen, roughly 30,000 ultrasonic signals are collected. When all signals are stacked together sequentially for one specimen, the attenuation of the signal is apparent as shown in Fig. 4(a). Zoomed in windows (highlighted with 1, 2, and 3) illustrating the sequential concatenation of 5 representative signals are shown in Fig. 4(b)–(d). By inspecting the composite signal in Fig. 4(a), it is evident that to accurately pinpoint the instant of damage (crack) initiation, additional visual information provided by the microscopes is essential to segregate the signal data into *healthy* and *cracked* regimes.

Since the experimental apparatus is instrumented with two microscopes of varied resolutions, two independent instances of damage

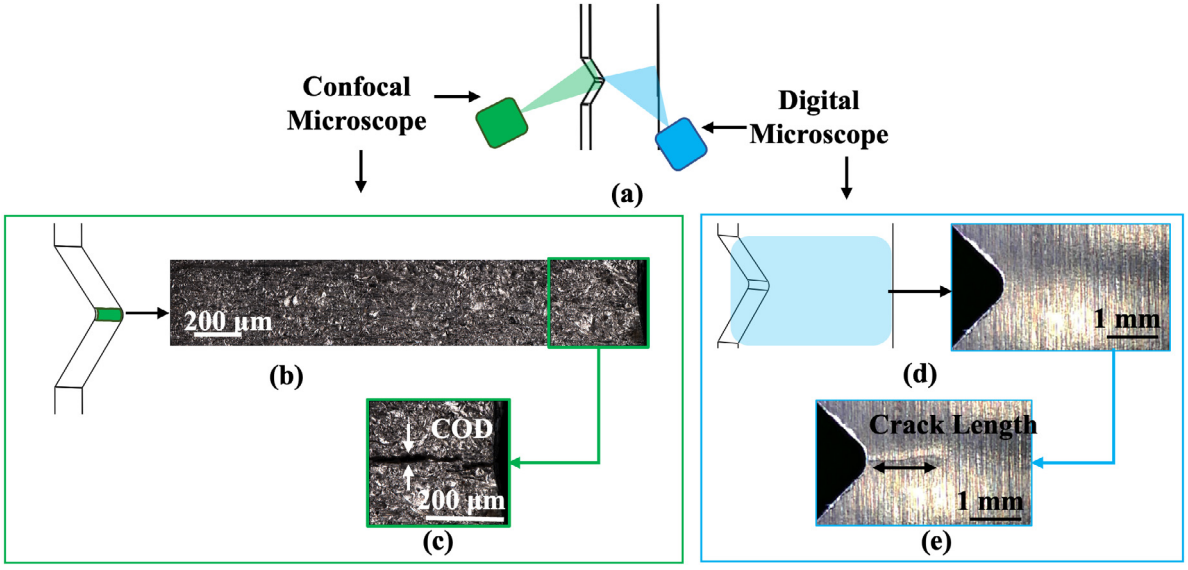

Fig. 3. (a) A schematic of the specimen along with the regions of interest for the digital and confocal microscope. (b) The region of interest of the confocal microscope with a collage of six individual images covering the entire notch root of an actual specimen. (c) A representative image of the 6th segment (extreme right of (b)) of the collage showing the formation of a fatigue crack in a cracked specimen and indicating the crack opening displacement (COD). (d) A typical image captured by the digital microscope and the corresponding area on the specimen. (e) A cracked version of the specimen as captured by the digital microscope indicating the crack length.

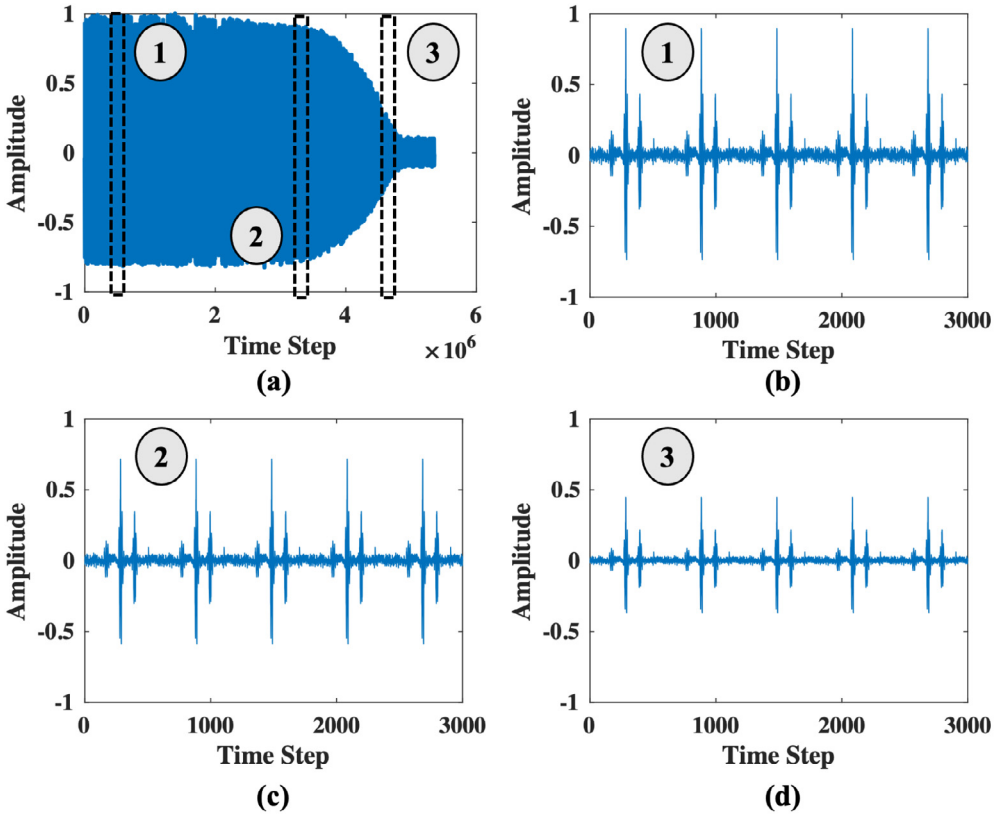

Fig. 4. (a) A concatenated signal during the entire fatigue experiment of a representative specimen showing the attenuation behavior. (b), (c), and (d) Zoomed in windows containing 5 signals each for the highlighted sections 1, 2, and 3 in (a).

initiation can be characterized (Dharmadhikari et al., 2021c), thus bifurcating the complete signal into three regimes, viz. healthy, Stage 1 (S_1), and Stage 2 (S_2) (Fig. 5(a)). The transition from healthy to S_1 damage corresponds to a crack appearance on the confocal microscope. This is visualized by microscopic cracks on the confocal images (Fig. 5(b)). Due to its high resolution, the cracks observed through this microscope show a COD of $\sim$ 3 µm. At this stage, there is no indication of damage through the digital microscope (Fig. 5(b)). However, as the fatigue damage continues to accumulate, the cracked state starts to become more prominent. Eventually, the transition to Stage 2 (S_2) damage is observed from the appearance of a crack on the digital microscope (Fig. 5(c)). As compared to confocal, the digital microscope is of significantly lower resolution. Therefore, the corresponding cracks measure up to a length scale of $\sim$ 0.2 mm. From an

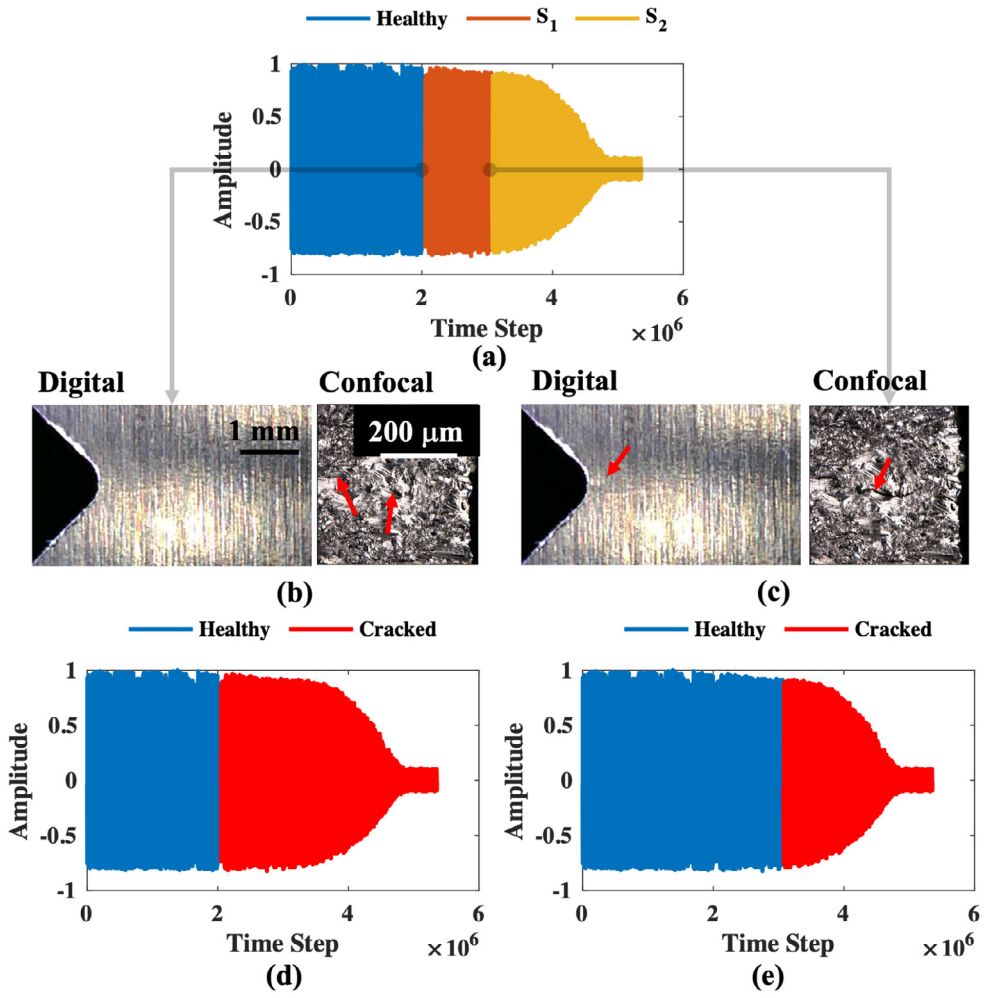

Fig. 5. (a) Segregation of the time-series data into three regimes, viz. healthy, S_1, and S_2. (b) Images from digital and confocal microscopes with cracks seen on confocal images (arrows). (c) Images from both microscopes showing the emergence of cracks. (d) and (e) A binary classification of the ultrasonic data for the S_1 and S_2 stages, respectively.

applied mechanics perspective, S_1 belongs to the "short crack" regime while S_2 belongs to the "long crack" regime (Stephens et al., 2000). From a fatigue life standpoint, the S_1 damage corresponds to 45% of the fatigue life, whereas S_2 damage corresponds to 58% of the fatigue life (Dharmadhikari et al., 2021b).

The labeled classification of the signals for both S_1 and S_2 the stages is done separately by using the respective transition points. A representative comparison of the labeled data for both the stages is shown in Fig. 5(d) and (e). For S_1 prediction (Fig. 5(d)), all the signals beyond the S_1 transition point are treated as *cracked*, whereas for the S_2 regime (Fig. 5(e)), only the points beyond the S_2 transition are labeled as *cracked*. A total of 448,939 signals are accrued through the testing of 15 specimens. For S_1 labeled data, the distribution of the signals between the two classes is well balanced, with 45% healthy and 55% cracked signals. For S_2 data, however, the portion of the data from the cracked class is reduced, and almost 71% healthy and 29% cracked signals are available. Even with the skewness of this distribution, with 29%, a total of 130,000 signals are available for training which proves to be a rich resource for DNN-based evaluations.

3.2. Featurization of the ultrasonic signal data

It is often the case that several important features from time-series store critical information which may hold physics-based insights on the system at hand (Ying, Garrett Jr, Oppenheim, Soibelman, Harley, Shi, et al., 2013). Under such situations, developing a classifier that is built upon such features may come in handy but with an additional task of pre-processing the raw data. With ultrasonics, it is expected that the amplitude and phase information between the accrued data over time may contain the damage signatures (Rose, 2014). Using this insight, following two features are defined.

Amplitude Response in the Frequency Domain: The frequency response (a fast Fourier transform or FFT) of one signal is shown in Fig. 6(a). The maximum value (highlighted with a red circle in Fig. 6(a)) corresponds to the contribution of the base frequency of 10 MHz. Due to the interaction of the ultrasonic waves with the specimen boundaries, the signal received by the ultrasonic receiver may also contain other frequencies. However, these frequencies are insignificant compared to the base frequency which has the most dominant contribution while analyzing the FFT behavior (Matlack, Kim, Jacobs, & Qu, 2015). Therefore, to study the impact of damage on the amplitude attenuation, the maximum frequency in Fig. 6(a) is considered for further analysis. The corresponding feature (denoted by FFT_{max}) is a normalized version of this maximum value such that its behavior for each specimen starts with a unit magnitude as shown in Fig. 6(b) and (c). The curves are plotted against the normalized fatigue life for each specimen. The normalized fatigue life is computed by dividing the number of cycles by the total fatigue life of each specimen, thus ensuring a unified measure for comparison across several specimens.

Autocorrelation Function: Although amplitude attenuation is visually apparent while looking at the concatenated signal, the impact of phase

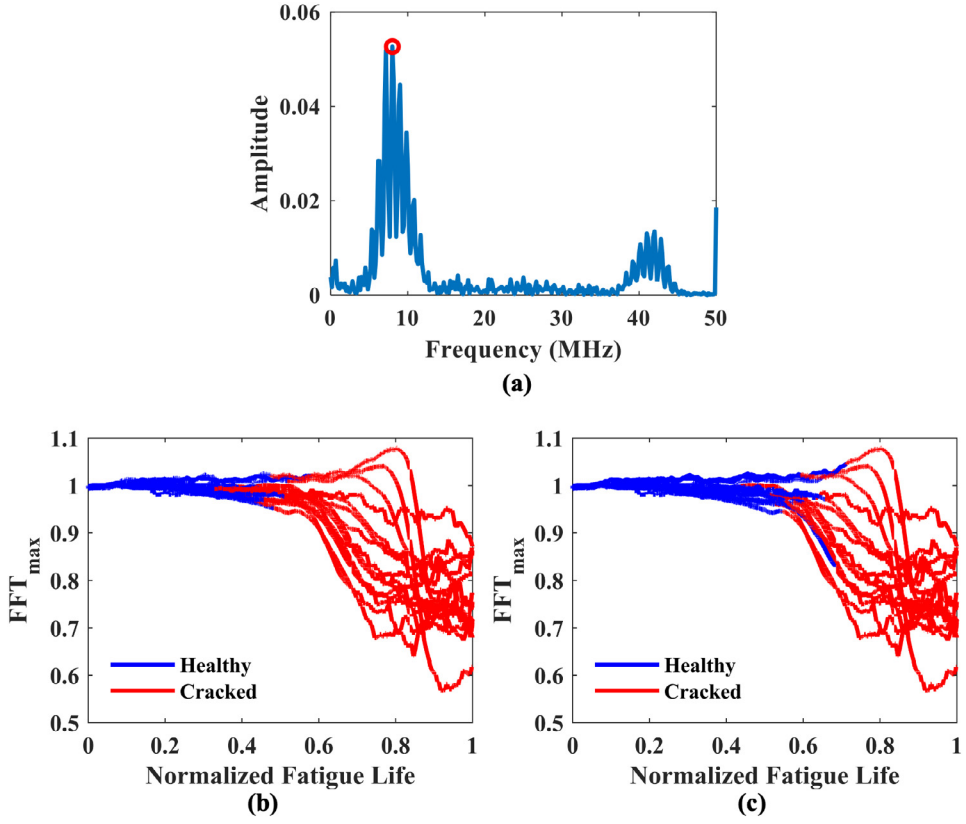

Fig. 6. (a) Frequency response of the raw signal in Fig. 4(b). The variation of FFT_{max} against the normalized fatigue life for the (b) S_1, and (c) S_2 stage. The transition between healthy and cracked states is shifted from ~ 45% fatigue life for the S_1 stage to a ~ 58% fatigue life for the S_2 stage.

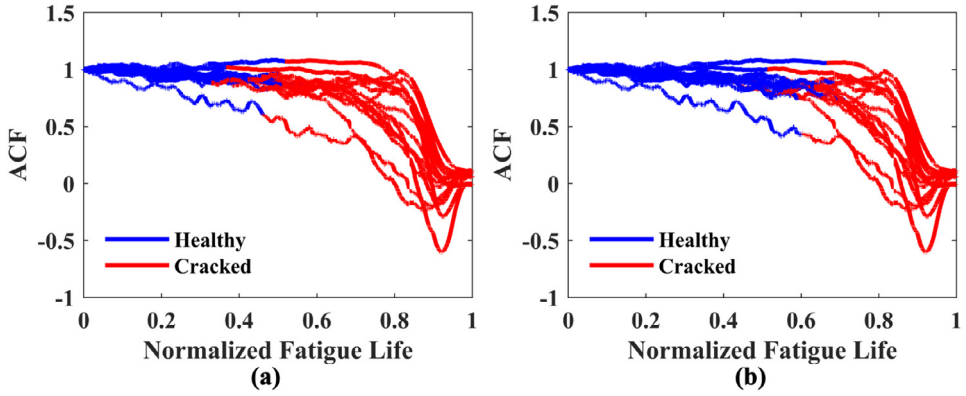

Fig. 7. The variation of ACF against normalized fatigue life for the (a) S_1, and (b) S_2 stage. The transition from healthy to cracked state for both stages shows a shift identical to FFT_{max}.

shift is lost while simply analyzing FFT_{max}. To capture these effects, the autocorrelation function (ACF) proves to be useful. The computation of ACF is done as follows. For one specimen, first, the reference raw signal (the first one) is denoted by x_0, and any subsequent ith signal is denoted by x_i. Therefore, using the definition of ACF, the magnitude at the ith instance can be computed using Eq. (1). x_i, μ_{x_i}, and σ_{x_i} are the amplitude, mean, and variance of the ith signal and have the units of signal amplitude. Similarly, x_0, μ_{x_0}, and σ_{x_0} are the amplitude, mean, and variance of the first reference signal. The morphing of the normalized ACF for all 15 specimens throughout its fatigue life is shown in Fig. 7.

$$ACF_i = \frac{1}{N-1}\sum\left(\frac{x_0-\mu_{x_0}}{\sigma_{x_0}}\right)\left(\frac{x_i-\mu_{x_i}}{\sigma_{x_i}}\right) \quad (1)$$

4. Design of deep neural networks

The premise of this study is to assess the capability of deep neural networks (DNN) to tackle the proposed classification problem of fatigue crack detection. Fig. 8 shows a broad schematic overview of the analysis adopted in the current work. Depending on the classification of the training data used, two DNN architectures can be hypothesized. The first one is a *feature-based* deep network that uses pre-defined features as its inputs. And, the second one is a *feature-less* deep network that is built through the use of raw unprocessed data.

4.1. Feature-based DNN (DNN_{FB})

Using ACF and FFT_{max} as inputs, a DNN abbreviated as DNN_{FB} is constructed. Fig. 9 depicts the structure of the DNN. The figure shows

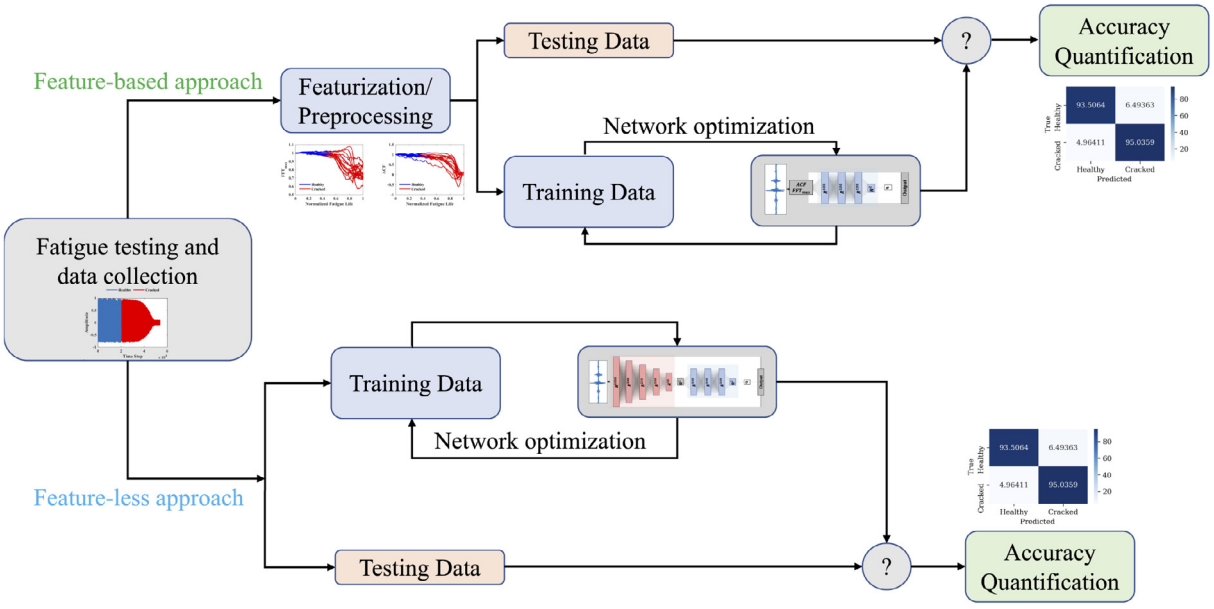

Fig. 8. A schematic of the current approach to train and test a deep neural network and the intended application for fatigue crack detection.

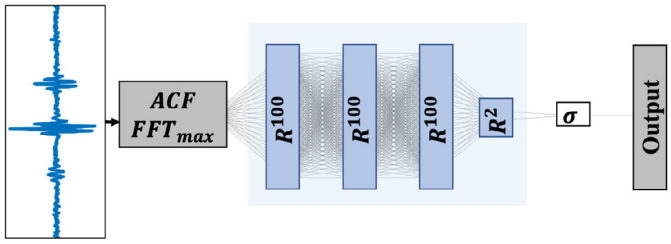

Fig. 9. Structure of the feature-based DNN (DNN_{FB}).

that ACF and FFT_{max} are extracted from the raw signal and used as inputs to the network. Following the input, the DNN has three hidden layers with 100 neurons each (R^{100}). The activation function used for each of these layers is the rectified linear unit (ReLU) (Goodfellow et al., 2016). The penultimate layer contains two neurons feeding into a sigmoid activation function. The network is fully connected. Due to binary classification, binary cross entropy is chosen as the objective function. The best results are obtained through the Adam optimizer (adaptive moment estimation) (Géron, 2019) with a learning rate of 0.001. The formulation is coded in TensorFlow using Python. The computations are conducted on an Intel®Xeon(R) Gold 5217 CPU @ 3.00 GHz with a NVIDEA Quadro RTX 6000 GPU.

4.2. Feature-less DNN (DNN_{FL})

One of the main limiting condition for the feature-based networks is the decision in the choice of features. With system knowledge, and exhaustive search, there can be some features that may be preferred (Ying et al., 2013). But, what if the choice of these features can be posed into an optimization problem that chooses the best formulation to maximize the efficiency of classification? To answer this question, the creation of the two inputs used in DNN_{FB} is delegated to a deeper neural network that uses just the raw signal without any preprocessing as its input. Through subsequent dimensionality reduction, a latent space of two features is extracted and fed into a DNN that has exactly the same structure as DNN_{FB}. The idea is inspired from the encoder principle of the autoencoder framework (Goodfellow et al., 2016). The structure of this network is shown in Fig. 10(a). Since the network does not require a pre-defined feature as an input, it is named as a feature-less network for subsequent discussion and to identify the latent space, it is abbreviated as DNN_{FL-2D}. To exemplify the relative ease with DNN in expanding the latent space to higher dimensions, two more networks with 10 (DNN_{FL-10D}) and 100 ($DNN_{FL-100D}$) latent neurons are also constructed (Fig. 10(b) and (c)). Similar to DNN_{FB}, a ReLU activation function is chosen for all the layers except the penultimate one which has the sigmoid activation function. Binary cross entropy is chosen as the objective function, and again the best results are obtained through the ADAM optimizer with a learning rate of 0.001. It is important to note that with constructing feature-less DNNs, not only manual intervention in choosing features is circumvented, but also significant pre-processing is eliminated since these networks are built to handle the raw data.

4.3. Performance metrics

Since the aim of all classifiers studied in this article is binary classification, their performance is best visualized through a confusion matrix with a focus on three metrics, viz. accuracy, sensitivity (true positive rate), and specificity (true negative rate) (Géron, 2019). As compared to the traditional *positive–negative* nomenclature prevalent in machine learning literature (Géron, 2019), in this context, a *positive* instance is identical to the *cracked* state, and vice-versa. Therefore, sensitivity is the ratio of accurately diagnosing the truly cracked data, and specificity is the ratio of accurately identifying the truly healthy data. Based on the volume of the data available, a 75%–25% train-test split is used for subsequent analysis. From the training dataset, 10% is chosen for validation. The data distribution among the healthy and cracked instances in the training and testing sets is kept the same as in the original dataset.

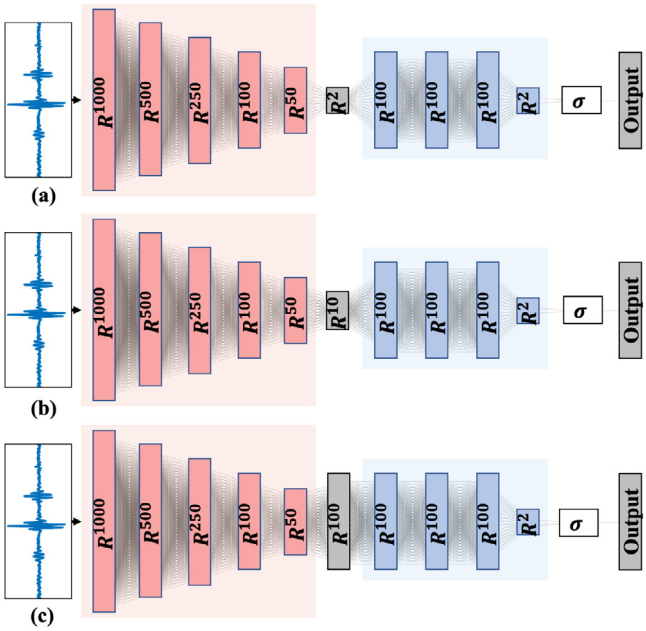

Fig. 10. Structure of the feature-less DNNs with (a) 2 (DNN_{FL-2D}), (b) 10 (DNN_{FL-10D}), and (c) 100 ($DNN_{FL-100D}$) latent neurons.

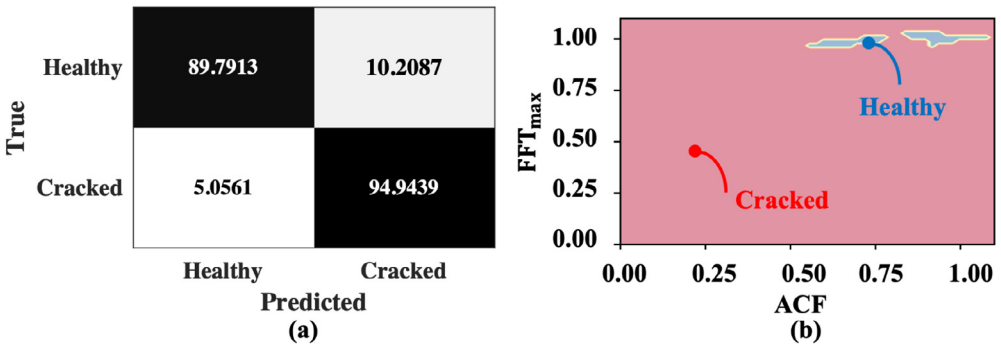

Fig. 11. (a) Confusion matrix, and (b) Decision boundary for $DNN_{FB}^{S_1}$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

4.4. Naming convention

In total, there are four different DNN's that are defined above, viz. DNN_{FB}, DNN_{FL-2D}, DNN_{FL-10D}, and $DNN_{FL-100D}$. By training each of this network on both sets of labeled data (S_1 and S_2), there are eight differently trained networks. To differentiate between the training data used for each DNN, a superscript indicating the training data is added. With this convention, for DNNs trained with S_1 labeled data, the nomenclatures are $DNN_{FB}^{S_1}$, $DNN_{FL-2D}^{S_1}$, $DNN_{FL-10D}^{S_1}$, and $DNN_{FL-100D}^{S_1}$, and for the ones trained with S_2 labeled data, the nomenclatures are $DNN_{FB}^{S_2}$, $DNN_{FL-2D}^{S_2}$, $DNN_{FL-10D}^{S_2}$, and $DNN_{FL-100D}^{S_2}$.

5. Results and discussion

5.1. Feature-based DNNs

The performance of $DNN_{FB}^{S_1}$ on the testing data is shown through the confusion matrix in Fig. 11(a). An overall accuracy of 92.36% is achieved with a sensitivity of 94.94% and specificity of 89.79%. Because of the 2-dimensional nature of the input, the decision boundaries of the trained network are easier to visualize and are shown in Fig. 11(b). In the feature space, the small regions occupied by the blue islands correspond to the healthy state, and the surrounding red space is the cracked state. The disjoint nature of the decision boundary is a numerical outcome of the available data. Physically, it would be impossible to have such discontinuous region of healthy feature space. For the S_2 regime, the performance of $DNN_{FB}^{S_2}$ is shown in Fig. 12(a). Since the detection is later than the S_1 regime, it is a comparatively easier problem and the performance of this network is better with an overall accuracy of 96.2%. This improvement is mainly due to the specificity of the network which is remarkably better by about 8%. The significant disparity between both the stages is better visualized through the widening of the decision boundaries as shown in Fig. 12(b). The wider boundaries indicate that as the detection is delayed, a larger chunk of the feature space becomes available for the healthy state. Again, the disjoint islands are numerical artifacts arising due to the nature of this particular data.

5.2. Feature-less DNNs

The performance of $DNN_{FL-2D}^{S_1}$ and $DNN_{FL-2D}^{S_2}$ is shown in Fig. 13. For the S_1 regime, the overall accuracy is 94.26% - about 2% improvement than $DNN_{FB}^{S_1}$. As observed from the matrix, this change is mainly due to a marked improvement in the specificity of the DNN. For the S_2 regime, $DNN_{FL-2D}^{S_2}$ outperforms $DNN_{FB}^{S_2}$ in all the metrics by having a remarkably high overall classification accuracy (98.94%). For DNN_{FL-10D} and $DNN_{FL-100D}$, similar improvements are observed. The confusion matrices for both of these networks are

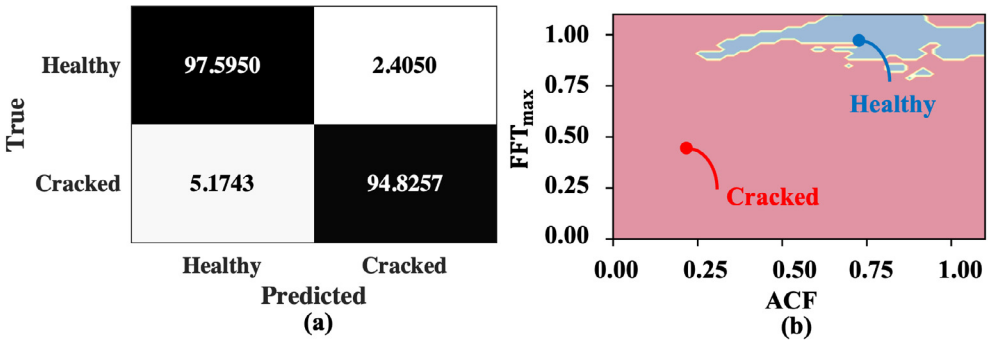

Fig. 12. (a) Confusion matrix, and (b) Decision boundary for $DNN_{FB}^{S_2}$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

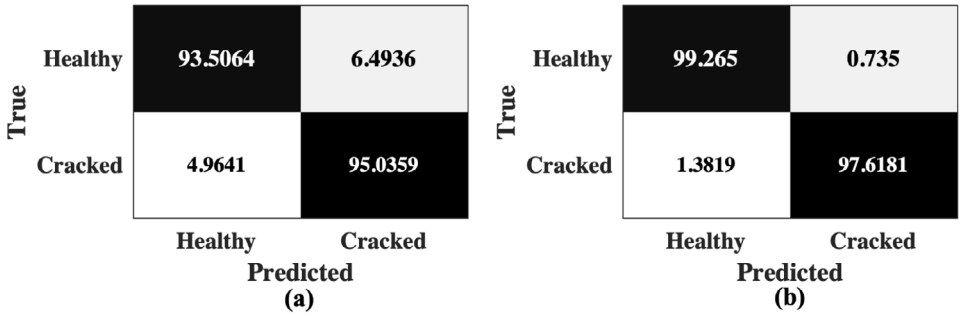

Fig. 13. Confusion matrices for (a) $DNN_{FL-2D}^{S_1}$, and (b) $DNN_{FL-2D}^{S_2}$.

not explicitly shown, instead a comparison of the performance of all the networks is summarized in Fig. 14. The comparison offers multiple insights between the two different networks types, and amongst different feature-less networks for both the regimes. For the S_1 regime, the best feature-less DNN outperforms the feature-based DNN by about 2%. However, a negligible difference in sensitivity is observed across all DNNs implying that all the networks are equally good in identifying the cracked state data. The major differentiating factor in the performance is specificity with the feature-less DNNs clearly emerging the superior classifiers when it comes to identifying the healthy state data. Among the feature-less DNNs for the S_1 regime, no strong correlation to the number of latent neurons can be interpreted. $DNN_{FL-2D}^{S_1}$ is observed to be marginally outperforming the other two. For the S_2 regime, the feature-less DNNs are unequivocally better in all the three aspects as compared to the feature-based DNNs outperforming them by about 2%. Unlike the S_1 regime, a distinct improvement is observed in the specificity of the feature-less networks. Among the feature-less DNNs, the sensitivity seems to be degrading with increase in latent neurons with $DNN_{FL-2D}^{S_2}$ showing the best performance and $DNN_{FL-100D}^{S_2}$ showing the worst. The specificity shows exactly opposite behavior with a marginal increase corresponding to increase in latent neurons.

As compared to the past literature, these results show better performance on two fronts, viz, (i) from a mechanics perspective, and (ii) from a data analysis perspective. From the mechanics standpoint, the successful capability of crack detection in the S_1 regime with ultrasonics exceeds the demonstrations seen in literature (Dharmadhikari et al., 2021b). Often, with such sensors, the crack length scales are limited to few hundred microns (Aria et al., 2020). However, in this case, damage detection is achieved in the 'crack coalescence' regime which occurs before the appearance of any crack. By using ultrasonics, past studies have often relied on nonlinear techniques (Matlack et al., 2015) for detection at such smaller length scales. But a successful demonstration with this study identifies a limiting case with aluminum 7075 for which linear ultrasonics can also prove efficient. From a data analysis standpoint, it is often difficult to compare with contemporary literature due to varied experimental protocols. However, as compared to symbolic analysis on similarly accrued data (Dharmadhikari et al., 2021b), this deep learning study has shown an improvement of about 10%-15%.

The significance of superior performances of the feature-less DNNs also helps in comparing between manually chosen features, and DNN-based feature definitions. From the above analysis, it is unequivocally evident that DNNs are capable of formulating complex albeit black-box features that outperform manually chosen definitions. One can argue that with better system knowledge, more relevant features can be chosen. However, the time and effort in choosing and pre-processing the features can be eliminated by training better DNNs. Moreover, the feature-less DNNs used in this analysis operate on raw unprocessed data as compared to the pre-processed filtered data for feature-based networks, thereby enhancing the impact of feature-less DNN performance. Since varying the number of latent neurons is simpler than selecting appropriate features, the users can choose to have as many features as needed for the analysis. Finally, with the above advantages, it can be concluded that some shortcomings due to the lack of expert supervision can be overcome with such feature-less networks.

6. Conclusions and future work

The article primarily shows the application of a deep learning-based strategy for fatigue damage detection at two stages (S_1 and S_2) using ultrasonic time-series data. Two main DNN architectures are used for the analysis. The first DNN, a feature-based network, uses two features (ACF and FFT_{max}) as its input. This network shows as overall accuracy of 92.36% in classifying S_1 labeled data and 96.2% in classifying S_2 labeled data. The second network is built using an encoder-inspired framework that automatically generates features from the raw data without manual intervention. Depending on the dimension of the latent space for the encoder, three DNNs with 2, 10, and 100 dimensions are studied. Through these feature-less networks, an accuracy of 94.26% and 98.94% is achieved for the S_1 and S_2 stages, respectively. On an average, therefore, the performance of the feature-less networks is observed to be better than the feature-based DNNs. The accuracies across

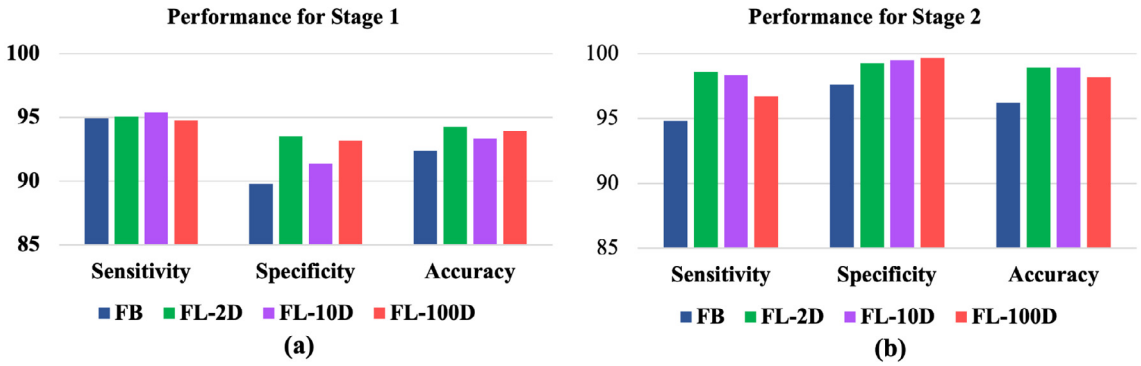

Fig. 14. Performance comparison across the four DNN's for (a) S_1, and (b) S_2 stage.

all the analyses are compared in Fig. 14. These observations also lead to the interpretation that the feature-less networks are better capable of formulating complex features that triumph the manual selections. Owing to their construction, it can be inferred that the feature-less DNNs pose an optimization problem to find the best features and in the process span a vast space of possible formulations. Whereas manually chosen features are always limited to exploring a constricted functional space that is pre-defined. The feasibility of constructing such feature-less DNNs also leads to end-to-end deep learning models which follow 'domain-invariance' (Fawaz et al., 2019). Such invariance can facilitate a common data analysis method across all damage detection problems irrespective of the sensing tools.

However, further developments in the proposed research need to be pursued, mainly on two fronts. First, the feasibility of employing such black-box features in safety-critical environments need to demonstrated beyond a lab-based setting. And second, efforts need to be taken towards gaining perfect classifications between healthy and cracked data. With several evidences demonstrating that such DNN models exhibit excellent performance in identifying anomalies in a controlled setting, research towards testing on uncontrolled environments is gaining traction (Karvelis, Georgoulas, Kappatos, & Stylios, 2021; Yuan et al., 2020) thereby approaching real-time applications. To tackle the black-box nature, a middle ground is being pursued through physics-informed neural nets which offer better interpretation of the network performance (Chen & Liu, 2021). Hence, further research into understanding the correlations between the physics of damage and latent feature space in such deep networks needs to be pursued. Towards gaining better (perfect) accuracies, a probable approach is through the use of better time-series analysis models such as recurrent neural networks (RNNs) or long short-term memory networks (LSTMs) which take into account the history of the data. In this article, a classification problem is pursued to identify anomalies for intermittent and randomly collected data. However, through a continuous analysis of the available data, in theory, forecasting models can be developed to predict imminent damage. A main limiting factor towards such modeling approaches can be the lack of sufficient data. Nevertheless, with the available resources, the authors intend to develop such forecasting models. Moreover, the crack detection problem in this article is reduced to binary classification. There can be several paths of developing it towards multi-class damage detection. The first and obvious approach can be to have a three class problem with the states *healthy*, S_1, and S_2. A second approach would be from the remaining useful life (RUL) perspective where the data is segregated in to several instances based on the RUL. Such bifurcations have been pursued in the past, albeit for larger components (Aria et al., 2020; Lim & Sohn, 2018). Eventually, the developed methods will be applied to additively manufactured aluminum alloys e.g., AlSi10Mg (Dharmadhikari & Basak, 2021a).

CRediT authorship contribution statement

Susheel Dharmadhikari: Experiments, Data analysis, Writing – original draft. **Amrita Basak:** Funding acquisition, Writing - review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The work reported in this paper is supported in part by the Department of Mechanical Engineering at the Pennsylvania State University, United States, University Park, PA 16802. Any opinions, findings, and conclusions in this paper are those of the authors and do not necessarily reflect the views of the supporting institution. The authors would like to thank Dr. Clyde Giles, and Mr. Ankur Mali for their help with deep learning analysis through the course IST 597 at Penn State.

References

Aria, A., Lopez Droguett, E., Azarm, S., & Modarres, M. (2020). Estimating damage size and remaining useful life in degraded structures using deep learning-based multi-source data fusion. *Structural Health Monitoring*, *19*(5), 1542–1559.

ASTM, D. (2015). E466-15, standard practice for conducting force controlled constant amplitude axial fatigue tests of metallic materials. *ASTM International*.

Azimi, M., Eslamlou, A. D., & Pekcan, G. (2020). Data-driven structural health monitoring and damage detection through deep learning: State-of-the-art review. *Sensors*, *20*(10), 2778.

Bao, Y., Chen, Z., Wei, S., Xu, Y., Tang, Z., & Li, H. (2019). The state of the art of data science and engineering in structural health monitoring. *Engineering*, *5*(2), 234–242.

Bhattacharya, C., Dharmadhikari, S., Basak, A., & Ray, A. (2021). Early detection of fatigue crack damage in ductile materials: A projection-based probabilistic finite state automata approach. *ASME Letters in Dynamic Systems and Control*, *1*(4), Article 041003.

Chen, J., & Liu, Y. (2021). Probabilistic physics-guided machine learning for fatigue data analysis. *Expert Systems with Applications*, *168*, Article 114316.

Dharmadhikari, S., & Basak, A. (2021). Energy dissipation metrics for data-driven fatigue damage detection in the short crack regime. In *ASME turbo expo*.

Dharmadhikari, S., & Basak, A. (2021a). Evaluation of early fatigue damage detection in additively manufactured AlSi10Mg. In *2021 International Solid Freeform Fabrication Symposium*.

Dharmadhikari, S., Bhattacharya, C., Ray, A., & Basak, A. (2021b). A data-driven framework for early-stage fatigue damage detection in aluminum alloys using ultrasonic sensors. *Machines*, *9*(10), 211.

Dharmadhikari, S., Keller, E., Ray, A., & Basak, A. (2021c). A dual-imaging framework for multi-scale measurements of fatigue crack evolution in metallic materials. *International Journal of Fatigue*, *142*, Article 105922.

Dung, C. V., Sekiya, H., Hirano, S., Okatani, T., & Miki, C. (2019). A vision-based method for crack detection in gusset plate welded joints of steel bridges using deep convolutional neural networks. *Automation in Construction*, *102*, 217–229.

Ebrahimkhanlou, A., & Salamone, S. (2018). Single-sensor acoustic emission source localization in plate-like structures using deep learning. *Aerospace*, *5*(2), 50.

Ewald, V., Groves, R. M., & Benedictus, R. (2019). Deepshm: a deep learning approach for structural health monitoring based on guided lamb wave technique. *Vol. 10970*, In *Sensors and smart structures technologies for civil, mechanical, and aerospace systems 2019* (p. 109700H). International Society for Optics and Photonics.

Farrar, C. R., & Lieven, N. A. (2007). Damage prognosis: the future of structural health monitoring. *Philosophical Transactions of the Royal Society of London A (Mathematical and Physical Sciences)*, *365*(1851), 623–632.

Fawaz, H. I., Forestier, G., Weber, J., Idoumghar, L., & Muller, P.-A. (2019). Deep learning for time series classification: a review. *Data Mining and Knowledge Discovery*, *33*(4), 917–963.

Géron, A. (2019). *Hands-on machine learning with scikit-learn, keras, and tensorflow: concepts, tools, and techniques to build intelligent systems*. O'Reilly Media.

Goodfellow, I., Bengio, Y., Courville, A., & Bengio, Y. (2016). *Vol. 1*, *Deep learning*. (2), MIT press Cambridge.

Karimian, S. F., Modarres, M., & Bruck, H. A. (2020). A new method for detecting fatigue crack initiation in aluminum alloy using acoustic emission waveform information entropy. *Engineering Fracture Mechanics*, *223*, Article 106771.

Karvelis, P., Georgoulas, G., Kappatos, V., & Stylios, C. (2021). Deep machine learning for structural health monitoring on ship hulls using acoustic emission method. *Ships and Offshore Structures*, *16*(4), 440–448.

Kong, X., & Li, J. (2018). Automated fatigue crack identification through motion tracking in a video stream. *Vol. 10598*, In *Sensors and smart structures technologies for civil, mechanical, and aerospace systems 2018* (p. 105980V). International Society for Optics and Photonics.

Leong, W., Staszewski, W. J., Lee, B., & Scarpa, F. (2005). Structural health monitoring using scanning laser vibrometry: III. Lamb waves for fatigue crack detection. *Smart Materials and Structures*, *14*(6), 1387.

Lim, H. J., & Sohn, H. (2018). Online fatigue crack quantification and prognosis using nonlinear ultrasonic modulation and artificial neural network. *Vol. 10598*, In *Sensors and smart structures technologies for civil, mechanical, and aerospace systems 2018* (p. 105981L). International Society for Optics and Photonics.

Matlack, K. H., Kim, J.-Y., Jacobs, L. J., & Qu, J. (2015). Review of second harmonic generation measurement techniques for material state determination in metals. *Journal of Nondestructive Evaluation*, *34*(1), 273.

Molent, L., & Dixon, B. (2020). Airframe metal fatigue revisited. *International Journal of Fatigue*, *131*, Article 105323.

Munawar, H. S., Hammad, A. W., Haddad, A., Soares, C. A. P., & Waller, S. T. (2021). Image-based crack detection methods: A review. *Infrastructures*, *6*(8), 115.

Mylonas, C., Tsialiamanis, G., Worden, K., & Chatzi, E. N. (2022). Bayesian graph neural networks for strain-based crack localization. *Vol. 9*, In *Data science in engineering* (pp. 253–261). Springer.

Nima, R., & Shila, F. (2020). Crack classification in rotor-bearing system by means of wavelet transform and deep learning methods: an experimental investigation. *Journal of Mechanical Engineering, Automation and Control Systems*, *1*(2), 102–113.

Papazian, J. M., Nardiello, J., Silberstein, R. P., Welsh, G., Grundy, D., Craven, C., et al. (2007). Sensors for monitoring early stage fatigue cracking. *International Journal of Fatigue*, *29*(9–11), 1668–1680.

Rose, J. L. (2014). *Ultrasonic guided waves in solid media*. Cambridge University Press.

Salehi, H., & Burgueño, R. (2018). Emerging artificial intelligence methods in structural engineering. *Engineering Structures*, *171*, 170–189.

Sangid, M. D. (2013). The physics of fatigue crack initiation. *International Journal of Fatigue*, *57*, 58–72.

Sofronas, A. (2012). *Case histories in vibration analysis and metal fatigue for the practicing engineer*. John Wiley & Sons.

Stephens, R. I., Fatemi, A., Stephens, R. R., & Fuchs, H. O. (2000). *Metal fatigue in engineering*. John Wiley & Sons.

Suresh, S. (1998). *Fatigue of materials*. Cambridge University Press.

Tripathi, G., Anowarul, H., Agarwal, K., & Prasad, D. K. (2019). Classification of micro-damage in piezoelectric ceramics using machine learning of ultrasound signals. *Sensors*, *19*(19), 4216.

Xu, L., Yuan, S., Chen, J., & Ren, Y. (2019). Guided wave-convolutional neural network based fatigue crack diagnosis of aircraft structures. *Sensors*, *19*(16), 3567.

Yang, C., Chen, J., Li, Z., & Huang, Y. (2021). Structural crack detection and recognition based on deep learning. *Applied Sciences*, *11*(6), 2868.

Ying, Y., Garrett Jr, J. H., Oppenheim, I. J., Soibelman, L., Harley, J. B., Shi, J., et al. (2013). Toward data-driven structural health monitoring: application of machine learning and signal processing to damage detection. *Journal of Computing in Civil Engineering*, *27*(6), 667–680.

Yuan, F.-G., Zargar, S. A., Chen, Q., & Wang, S. (2020). Machine learning for structural health monitoring: challenges and opportunities. *Vol. 11379*, In *Sensors and smart structures technologies for civil, mechanical, and aerospace systems 2020*. International Society for Optics and Photonics, Article 1137903.

Zhang, L., Wang, Z., Wang, L., Zhang, Z., Chen, X., & Meng, L. (2021). Machine learning based real-time visible fatigue crack growth detection. *Digital Communications and Networks*.

Zhao, R., Yan, R., Chen, Z., Mao, K., Wang, P., & Gao, R. X. (2019). Deep learning and its applications to machine health monitoring. *Mechanical Systems and Signal Processing*, *115*, 213–237.