Gas turbine failure classification using acoustic emissions with wavelet analysis and deep learning

M.S. Nashed [a], J. Renno [b,*,1], M.S. Mohamed [a,2], R.L. Reuben [c,3]

[a] *Institute for Infrastructure & Environment, School of Energy, Geoscience, Infrastructure and Society, Heriot-Watt University, Edinburgh, United Kingdom*
[b] *Department of Mechanical & Industrial Engineering, College of Engineering, Qatar University, Doha, Qatar University, Qatar*
[c] *Department of Mechanical, Process and Energy Engineering, School of Engineering & Physical Sciences, Heriot-Watt University, Edinburgh, United Kingdom*

ARTICLE INFO

ABSTRACT

Compared to vibration monitoring, acoustic emission (AE) monitoring in gas turbines is highly sensitive to changes that do not involve whole-body motion, such as wear, rubbing, and fluid-induced faults. AE signals captured by suitably mounted sensors can potentially provide early indications of abnormal turbine operation before such abnormalities manifest in structural vibration or emitted airborne noise. However, developing an online fault detection system requires extensive real-time data treatment to extract appropriate features and indicators from raw AE records. To build such a system for industrial turbines, researchers need to understand the AE-generating mechanisms associated with turbine operation and the sources of background noise. In this study, we aim to develop such an understanding using a small-scale turbine whose operational conditions can be modified safely to reflect both normal and faulty conditions. Our signal processing approach involves first extracting a time-series envelope using an averaging time selected to enhance major features and eliminate irrelevant noise. We then generate time–frequency features using a continuous wavelet transform, which are used to train a deep convolutional neural network to classify gas turbine conditions. The resulting model demonstrates high accuracy in classifying two normal running conditions and two faulty conditions at various turbine speeds. Overall, the proposed methodology offers a powerful tool for gas turbine condition monitoring, and we make all associated data available in open-source format to facilitate further research in this field.[4]

1. Introduction

The most critical components in a gas turbine are found in its combustion area where they would be subjected to a combination of high temperature, centrifugal and aerodynamic forces. Such conditions represent some of the biggest challenges in modern materials development and component design. Over the course of a gas turbine's operating lifespan, various types of faults can develop between the commissioning and any planned maintenance or overhaul including compressor fouling (Cyrus et al., 2004; Kurz et al., 2001; Lu, Wang, Jia, & Qi, 2016), blade erosion (Mann, 1998; Metwally & Hamed, 1995), blade fatigue (S. K. Bhaumik, Sujata, M., Venkataswamy, M. A., Parameswara, M.A., 2006; j. Hou, Wicks, B.J., Antoniou, R.A., 2002; Tong,

et al., 2020), blade creep (Bhaumik et al., 2002; Mazur et al., 2005), and seal rubbing (Bentley, 1974; YU, April 2002). These faults can significantly impact the performance and operating lifespan of the gas turbine and many of them are associated with mechanisms which generates acoustic emissions (AE) directly or indirectly. For example, AE can be directly generated by any contact between the rotor and stator, damage in bearings or rotor crack propagation. Indirect AE can arise from various sources such as abnormal dynamic loading, wear, accretion or corrosion of flow surfaces, misalignment, lubrication degradation, oil starvation, or foreign object damage. It has been reported that some faults in turbines are better detected with AE than other techniques such as vibration monitoring. For example, rubbing (Hall. L. D. and Mba. D, 2004; Y. He, et al., 2021; D. Mba, Cooke, A., Roby, D., and Hewitt, G., 2004; D. Mba, Hall, L. D., 2002) and damage in bearings (Caesarendra,

Nomenclature

Abbreviation/Symbol	Description
AE	acoustic emission
ANN	artificial neural network
AUC	area under the ROC curve
CNN	convolutional neural network
CWT	continuous wavelet transform
E	energy of signal in [V^2.s]
FN	number of false negative classifications
FP	number of false positive classifications
FPT	free power turbine
GG	gas generator
ML	machine learning
ROC	receiver operating characteristic
RPS	rotation per second
SFFT	short-time Fourier transform
t	time in seconds [s]
TN	number of true negative classifications
TP	number of true positive classifications
v	amplitude of raw acoustic emission signal in volts [V]

et al., 2016; D. Hou, Qi, Wang, & Han, 2022) have been reported to be better detected through AE. Boundary/supporting conditions can also significantly influence the accuracy of fault diagnosis results (D. Zhou, et al., 2021). Therefore, processing the measured signals to distinguish diagnostically significant temporal and spectral features from the AE signal collected by the sensors in an operational turbine unit is crucial for accurately detecting faults.

Board's seminal work (Board, 2000) was one of the earliest applications of AE for gas turbine condition monitoring. Board found that the cage rotational frequency of a roller bearing (105.8 Hz) was dominant in the AE spectrum. However, it should be noted that Board's study (Board, 2000) primarily focused on the AE associated with ancillary turbine equipment such as roller element bearings. While subsequent research has further explored AE monitoring of gearboxes and rolling element bearings, these topics are not directly relevant to the current study. Douglas et al. (Douglas et al., 2004) have observed the compressor blade passing frequency in an operating laboratory gas turbine and also noted changes in the AE spectrum associated with an artificial blade fault. Following this, Nashed et al. (Nashed, Steel, & Reuben, 2014), published the first systematic study of fluid-induced acoustic emission in gas turbines, focusing on "normal running" conditions. The currnet work builds on this baseline understanding of the physical sources of AE in gas turbines.

When diagnosing mechanical faults from a "raw" signal, several methods are commonly used. Typically, the raw signal consists of data acquired at a low multiple (2.5x to 5x) of the frequency of the structure-borne stress wave(s) (0.1–1 MHz). To extract relevant information from the signal, averaging is often employed. Averaging can be performed in analogue (prior to acquisition) or digital format. In the latter case, very fast capture and storage hardware is essential and several commercial systems are available to facilitate this. For example Hall and Mba (Hall. L. D. and Mba. D, 2004) calculated the AE envelope using a low order smoothing filter applied to the root-mean-square (RMS) of the collected AE signal by individual sensors attached to the journal bearing of a large-scale turbine. A significantly larger amplitude fluctuation was observed at one position which was attributed to rubbing shaft whirl in the vicinity of that sensor. Mba et al. (D. Mba, Cooke, A., Roby, D., and Hewitt, G., 2004) later focused on the industrial application of (D. Mba, Hall, L. D., 2002) and the transmission of AE signals in such complex structures. While the approaches mentioned above have been successful in specific applications, generalizing them to cases that involve multiple

sources and sensors can be challenging.

Detecting failures using AE requires pre- and post-processing of data and analytical tools to extract useful features that can be directly linked to a fault (Z. Chen, Xia, Li, & Pan, 2022; Fan, 2007; Shahkar & Khorasani, 2019; Shanbhag, Meyer, Caspers, & Schlanbusch, 2020). This can be particularly challenging in industrially significant applications where the sources of AE that are diagnostically important can have specific spectral and temporal characteristics and may originate from various parts of the machine or structure of interest. Additionally, noise from sources that are not diagnostically relevant can pose a challenge to identifying the diagnostically significant AE. Recent advances in machine learning (ML) algorithms offer an opportunity to develop online monitoring and failure prediction systems that can potentially identify any changes in the AE signal patterns without requiring analytical intervention by humans. Moreover, using ML for online monitoring and failure prediction is compliant with ISO 13379 (König, Jacobs, Stratmann, & Cornel, 2021). Surucu et al. (Surucu, Gadsden, & Yawney, 2023) recently published a comprehensive review of ML techniques for condition monitoring applications and provided comparative analysis and insights into the advantages and limitations of each technique. Among ML techniques, convolutional neural networks (CNNs) is a supervised deep learning method that processes topological datasets and was initially inspired by the visual cortex (Hubel & Wiesel, 1968). Rui et al. (R. He, Li, Chen, Chen, & Liu, 2020) used a generative adversarial network (GAN) to monitor industrial processes. GANs are a type of CNNs that comprises two networks: a generator network that is tasked with generating data that is like the ground-truth and a discriminator network that is tasked with identifying the "forged" data (hence their adversarial nature). Jiang et al. (Jiang, Xia, Wang, Zhang, & Xi, 2022) proposed using multi-layer joint distribution-based CNNs in order to facilitate feature fusion between layers and to enhance the transferability of features between the layers of the CNN. They demonstrated the proposed approach on benchmark datasets. To improve the memory function for dynamic data of roller bearing condition monitoring, Xia et al. (Xia, et al., 2021) proposed combining different training algorithms and recurrent neural networks to provide a multistage fault diagnosis classifier.

In this work we demonstrate the ML approach for the particularly complex case of a gas turbine. In the interest of generating a wide dataset quickly, a laboratory-scale gas turbine was run under normal running conditions (with and without load) and under (two) abnormal conditions. AE signals were recorded at the various speeds for each of the four operating conditions. Then, the envelope of each recorded signal was computed in the time-domain using an averaging time selected to enhance the major features in the signal and supress the irrelevant noise. Time-frequency features are then obtained using the continuous wavelet transform (CWT). This transforms every AE signal into an image that is labelled with one of the four operating conditions. These images are used to train a convolutional neural network to classify the turbine operating condition. Although the CWT have been used with CNNs in condition monitoring applications (Liao, Ragai, Huang, & Kerner, 2021), the signals were not demodulated to reduce the size of the obtained images. In this work, we demodulate the AE signals and use the CWT to provide a computationally efficient implementation of CNNs. In this regard, this work fills a gap in the literature of using AE and CNNs for the online monitoring of gas turbines.

The paper is organized as follows. The experimental setup and the testing campaign are presented in Section 2. Section 3 presents the ML approach used in this paper and the classification strategy developed herein. Results are presented and discussed in Section 4. Conclusions for this work are drawn in Section 5.

2. Experimental set-up and test procedure

A schematic and a photo of the gas turbine used in this work are shown in Fig. 1. The P.9000 unit, manufactured by Cussons Technology
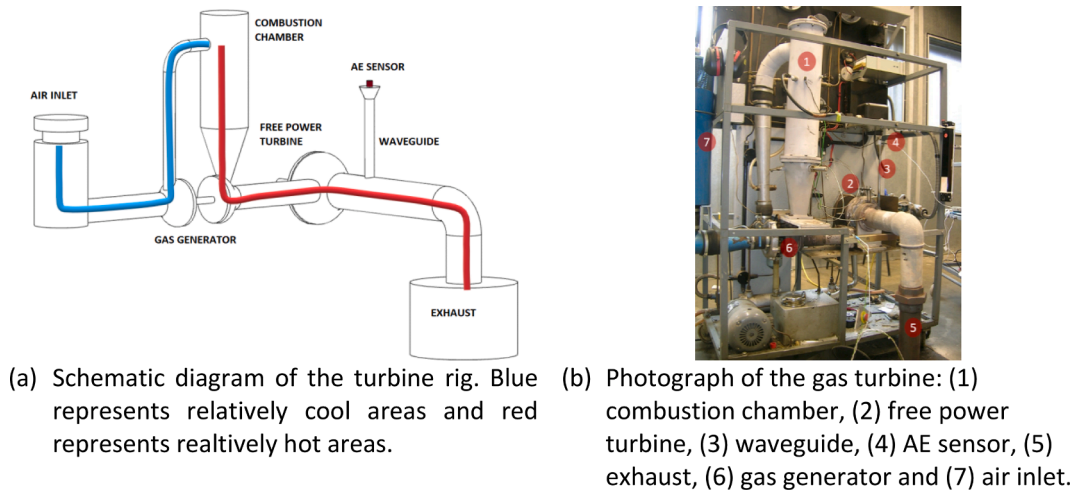
(a) Schematic diagram of the turbine rig. Blue represents relatively cool areas and red represents realtively hot areas.

(b) Photograph of the gas turbine: (1) combustion chamber, (2) free power turbine, (3) waveguide, (4) AE sensor, (5) exhaust, (6) gas generator and (7) air inlet.

Fig. 1. Experimental setup used to collect AE data of normal and faulty operations.

Ltd® (Cussons, 2004), comprises of a gas generator (GG) and a free power turbine (FPT). The GG has two stages: a compressor and a turbine mounted back-to-back on a short shaft supported on a journal bearing. The FPT is a single stage radial turbine that operates over the range 170 to 600 revolutions per second (RPS) and can generate a maximum power of approximately 4 kW. The FPT is connected to an alternator that produces electricity, which can be dissipated through heat lamps or made available via normal plug sockets. The fuel flow rate is adjusted manually. Both the fuel and the air flow rates are measured using flowmeters.

To detect acoustic emissions from the gas turbine, a Micro-80D AE sensor (manufactured by Physical Acoustics®) was used. The sensor was attached to a waveguide that was welded onto the turbine exhaust. The GG and FPT both contain journal bearings which produce continuous AE signals. Therefore, the sensor was specifically placed on the waveguide to avoid recording bearing-induced AE, which is primarily structure-borne and transmitted along the shafts to the bearing housings. The gas-borne sources of interest, on the other hand, are expected to originate from the flow through confined spaces, combustion, combustion instability, and standing waves within the various passages of the turbine.

A tachometer was used to record shaft speed on the FPT synchronously with the AE sensor. The tachometer consisted of a slotted disc with 34 teeth of circumferential extent of $5°$ and one with the extent of $15°$, which allows the determination of shaft speed and rotational position along with the AE signals. The AE data was acquired using a data acquisition system based on an in-house built desktop PC with a 12-bit, National Instruments®, PCI-6115 Data Acquisition Board. The board could record raw AE signals at a rate up to 5 MHz for up to four channels simultaneously and had a total onboard memory of 32 MB.

During all the tests, the raw AE data acquisition configuration was consistent for each operating condition, recording 20 sets of raw AE data, each with a duration of 0.03 s at a sampling rate of 5 MHz. Four different types of operating conditions were tested, consisting of two normal running conditions and two faulty conditions. To ensure consistency, each test (i.e., at each turbine condition) was performed three times. In total, 2578 AE signals were recorded, and these were used for training the CNN as will later be explained.

2.1. Normal operation without load

No loads were applied in this set of tests (i.e., the turbine was idling). To operate the turbine with no load, the driving belt between the FPT and the alternator was removed, and the turbine speed was changed by incrementally increasing the fuel flow. The change in fuel flow results in

changes in both the GG and FPT speeds. The speed was varied between 120 and 350 RPS and 20 records of AE (each of length 0.03 s as noted above) were acquired at each speed. This speed range was the widest possible range over which reasonably stable operation of the turbine could be achieved. To avoid capturing gas turbine transient conditions, after each change in speed, the turbine was left to settle down (typically within 20 s) and measurements were not taken until the turbine had settled at the new condition. This experiment alongside other experiments (i.e., different turbine condition) was repeated a total of three times to ensure consistency.

2.2. Normal operation with load

To achieve a change in load, the speed of the GG was kept constant, whilst the alternator load was increased since higher alternator loads result in lower FPT speeds. In this test, the signals were expected to be more complex, but also more relevant to real operating conditions, hence the need for the idling benchmark. The other testing parameters such as the speed range, the length of AE recordings and their number were the same as those described in subsection 2.1.

2.3. Damaged impeller with two levels of intensity

This set of tests were carried out under the same conditions as the unloaded normal running tests to examine the effect of two induced faults on the AE signature. Blade damage was induced in the 10-blade FPT impeller by grinding off about 20 mm from the tips of opposing blades. The choice to damage only opposing blades was made primarily for safety reasons, but also because it minimises the introduction of radial imbalance which might otherwise have an independent effect on the AE signature. First, two opposing blades were damaged (D2) and then four opposing blades (D4) as shown in Fig. 2(b) and Fig. 2(c), respectively. In the second set of results the damage on the four blades was not circumferentially symmetrical, which was deliberate. As in the normal running no-load test, the connecting belt between the FPT and the alternator was removed and the speed of the turbine was changed between 120 RPS and 350 RPS by changing the fuel flow, which was recorded along with other thermodynamic aspects of the turbine rig. All data acquisition parameters were kept the same as in subsection 2.1.

3. Machine learning (ML)

This section covers a brief overview of Artificial Neural Networks (ANNs) and Convolutional Neural Networks (CNNs), followed by a description of the ML approach that will be used in this work. The entire
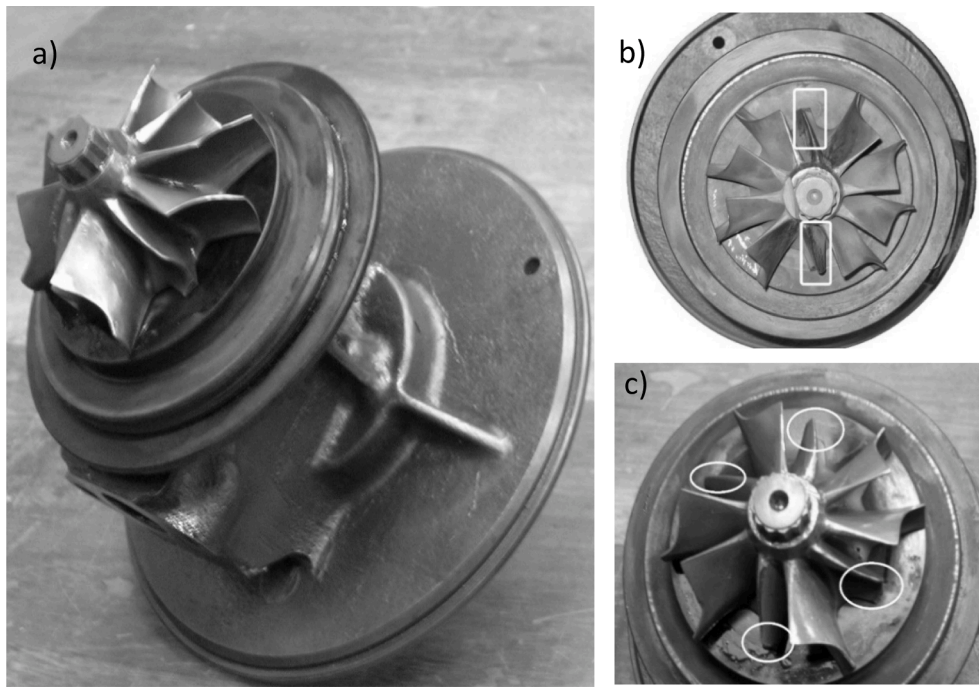
Fig. 2. Free power turbine impeller with damaged blades, (a): impeller body, (b): two damaged blades (D2), (c): four damaged blades (D4).

approach is implemented in MATLAB® and is openly available at a public repository[4].

3.1. Artificial and convolutional neural networks

Artificial neural networks (ANNs) are a set of computational tools inspired by the biological nervous system (Haykin & Lippmann, 1994). The main processing units, called neurons or perceptrons, are organized in layers starting with an input layer that receives pattern of features and culminating in an output layer that provides the network's "interpretation" of the pattern. The network is "trained" to recognise patterns, and the trained network with the adjusted weights and biases of the neurons forms a specific ANN model that can be used to interpret signals.

The specific type of ANN used here is the convolutional neural network (CNN) which have shown great ability to solve difficult pattern recognition tasks, such as those in medical imaging (Hameurlaine, Moussaoui, & Safa, 2019; Shen, Wu, & Suk, 2017) or traffic prediction for autonomous vehicles (Ranjan, Bhandari, Khan, Hong, & Kim, 2021; Ranjan, Bhandari, Zhao, Kim, & Khan, 2020). A CNN consists of a particular collection of layers; convolutional layers, fully connected layers and pooling layers, the convolution layers being the most important. The convolutional layers are only connected to a few neurons located within a small rectangular area in the previous layer which allows the network to focus on small low-level features in the first hidden layer then assemble them into larger higher-level features in the next hidden layer and so on. The pooling layers are similarly not fully connected; each neuron being connected to the outputs of a limited number of neurons in the previous layer aggregating these for onward transmission.

Residual Networks (ResNets) are CNNs that were recently introduced by He et al. (K. He, Zhang, Ren, & Sun, 2015). This class of CNNs achieved a top-five error rate in the ILSVRC challenge (Russakovsky, et al., 2015). ResNets consist of several stacked residual (or skip) units and

there are many versions based on the number of layers. In this research, ResNet-50 (which comprises 49 convolutional layers plus an output layer) is used. ResNet-50 was found to perform best with the test data in this work. The key to deep learning networks, such as ResNets is the use of skip connections where the signal is fed into a set of input layers and the architecture includes deep stacks of simple "residual units". Each residual unit is composed of two convolutional layers, with Batch Normalization (Sergey & Christian) and ReLU (J. Chen & Liu, 2021) activation using 3×3 kernels and preserving spatial dimensions. Skip units have two benefits; to force the network to learn the model and the residual at the same time, and to accelerate of the training process due to "in process" comparison of identity functions and target functions
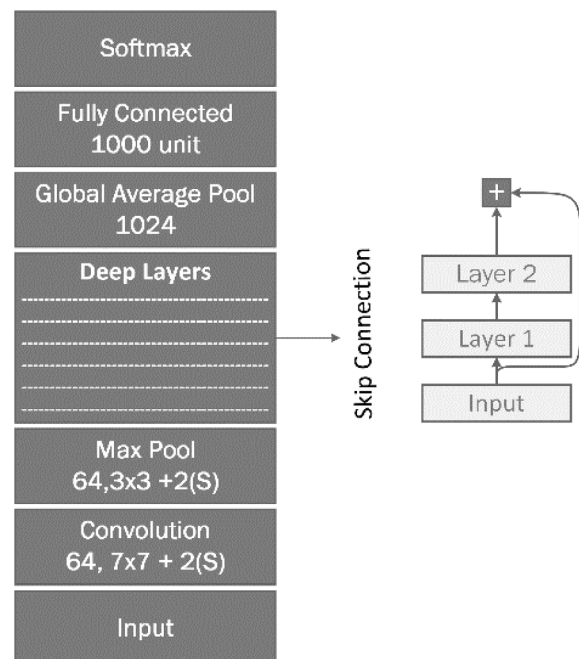

Fig. 3. Schematic of the structure of a ResNet (K. He, et al., 2015).

[4] https://github.com/MShadiNashed/Gas-Turbine-Failures-Classification-Using-Acoustic-Emissions-with-Wavelet-analysis-and-Deep-Learning.

(Fig. 3).

These properties of ResNets make them suitable for any task that can benefit from the depth of the network and the reduction in computational expense.

3.2. Fault classification development strategy

Fig. 4 depicts the four stages of the fault classification development strategy. The first stage involves signal processing and analysis of the AE data to design the most suitable time–frequency transform function. Demodulated frequency analysis is performed by calculating the signal envelope and then extracting the frequency components of interest within that envelope using a time–frequency method. At this stage, a comparison between the CWT and the short-time Fourier transform SFFT for raw and enveloped signals is performed to determine which transform reveals more information about the signals. Computational costs of both transform are also evaluated. Lastly, both the transform resolution and the frequency range of interest are assessed. Furthermore, the transform resolution and the frequency range of interest are assessed. The enveloped signal with the CWT provides the maximum amount of detail in the raw signal that may otherwise be indiscernible due to noise and the broad frequency range of interest. Employing the CWT and considering the signal envelope is not only useful for capturing details but can also significantly reduce computational costs for signal processing compared to using the SFFT. Additionally, as a time-–frequency analysis technique, the CWT is effective for analysing non-stationary signals (Zhou, et al., 2020) and it has been used successfully in various ML applications for image processing (Jiang, Xia, Wang, Fang, & Xi, 2022).

The second stage is then to apply the wavelet transform to all the AE signals so that they are converted into images and labelled based on their category, i.e., fault or no-fault. In the third stage we run a benchmark test to estimate the best CNN network to use with the data produced in the previous stage. During this stage different types of convolutional neural network are evaluated to check which network has a lower loss and achieves the best accuracy. ResNet-50 was found to be the best performing network. In the final stage the best performing network is tuned to achieve the best possible performance. To this end, the hyperparameters of ResNet-50 including learning rate, the mini-batch size and the dropout rate are adjusted.

4. Results and discussion

In the framework of fault identification and classification, the quality of the model is characterized by a set of typical metrics. Sokolova and Lapalme (Sokolova & Lapalme, 2009) demonstrated that for such problems, the most suitable metrics are the accuracy, recall, specificity, precision, F1 score, receiver operating characteristic, confusion matrix and the area under the curve. These performance metrics will be briefly described below where TP is the number of true positives, TN is the number of true negatives, FP is the number of false positives and FN is the number of false negatives.

- **Accuracy** measures how many observations, both positive and negative, are correctly classified:

$$\frac{TP + TN}{TP + FN + TN + FP}100\% \tag{1}$$

- **Recall (Sensitivity)** measures how many observations out of all positive observations are classified as positive:

$$\text{Recall} = \frac{TP}{TP + FN}100\% \tag{2}$$

- **Specificity** measures how many observations out of all negative observations are classified as negative:

$$\frac{TN}{TN + FP}100\% \tag{3}$$

- **Precision** measures how many observations predicted as positive are in fact positive:

$$\text{Precision} = \frac{TP}{TP + FP}100\% \tag{4}$$

- **F1 score** is the harmonic mean between precision and recall:

$$2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}100\% \tag{5}$$

- **Receiver operating characteristic (ROC) curve** shows the true positive rate (i.e., sensitivity) against the false positive rate which is
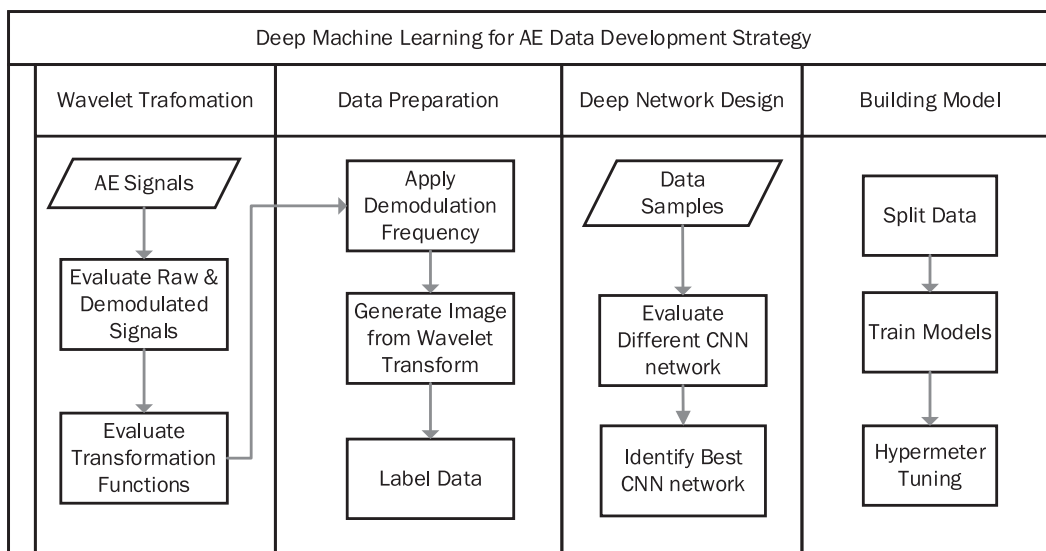

Fig. 4. Development strategy for gas turbine failure classification using convolution neural network.

the ratio of negative instances that are incorrectly classified as positive, i.e., the specificity. The ROC results show the trade-off between the detection rate (sensitivity) and the specificity. The closer the curve follows the left-hand border and then the top border of the ROC space, the more accurate the test. On the other hand, the closer the curve comes to the 45-degree diagonal of the ROC space, the less accurate the test.

- **Confusion matrix** is calculated by comparing the predictions of the model with the actual targets.
- **Area under the ROC Curve (AUC)** measures the entire two-dimensional area underneath the ROC curve. The higher the AUC, the better the performance of the model at distinguishing between the positive and negative classes.

4.1. Energy analysis

4.1.1. Turbine operation with and without load

The (uncalibrated) energy of the AE signal can be calculated by integrating the square of the amplitude over a fixed time, t, as follows:

$$E = \int_0^t v^2(t)dt \qquad (6)$$

where $v(t)$ is the amplitude of the AE waveform in volts (V); t is time in seconds (s); and E is the AE energy in $V^2.s$.

Fig. 5 shows the evolution of the AE energy plotted against FPT speed for both idling and loaded tests. For the idling test, the general trend is for an increase in energy with speed with what appears to be a peak at around 275 RPS. There is also an increase in scatter both within and between tests as speed increases.

The FPT's speed decreases when loaded, which necessitates the injection of additional fuel to align it with idling test speeds. Consequently, loaded tests are anticipated to produce more intricate AE signals that are more reflective of actual operating conditions. Fig. 5 illustrates that the energy generated in loaded tests is approximately two orders of magnitude higher than that in the idling tests. However, the most noteworthy characteristic is the more erratic variation of AE energy with speed, both within and between nominally identical tests.

4.1.2. Turbine operation with damaged blades

Fig. 6 shows the variation of the AE energy with FPT speed for the idling test with the two degrees of blade damage (D2 and D4). As can be seen, blade damage results in an increase in energy over the intact case, particularly at higher speeds, but with a very large variation between nominally identical tests. The higher degree of damage (D4), somewhat counter-intuitively, did not lead to an increase in energy over the lower degree of damage (D2), although it is not clear if this is significant given the degree of scatter between the tests.

4.2. Results of continuous wavelet transform

To prepare the AE data for use in the CNN, it is necessary to transform each raw signal into an image in the time–frequency domain. This is achieved using the red, green, and blue (RGB) color model. Fig. 7 shows two possible ways of doing this for a sample signal using the SFFT and the CWT.

The SFFT has limitations in representing an AE signal in both the time and the frequency domain due to the fixed width of the time window. In contrast, the CWT windowing function has a flexible size in both the frequency and time domains, allowing the CWT to access localized features of the signal that are not accessible by the SFFT. Furthermore, the multiresolution property of the CWT provides higher time–frequency details, especially for high-frequency events. Therefore, the CWT is a more effective at transforming the AE data into images that can be used by the CNN.

However, the computational requirements for the CWT of a raw AE sample climb steeply as the sampling frequency increases. Furthermore, the high-resolution representation of the AE may still not be suitable for direct use with a deep learning network for the following reasons. First, the localized events, such as those shown in Fig. 7 can be limited, brief and sparse. Moreover, the images are reduced to 224×224 pixels for use in ResNet-50. Therefore, only a limited number of features can be presented in each image to ensure that the network can extract meaningful patterns from the samples. The inclusion of irrelevant frequencies, such as high-frequency noise, can lead to false diagnosis results and mask other important low-frequency features. Additionally, traditional frequency transformations, such as SFFT, can generate complex sidebands, as shown in Fig. 7, which may further complicate the analysis.

To solve these issues and to improve the quality of the transformed images, demodulated frequency analysis was used before applying the CWT. The procedure for demodulated frequency analysis and the CWT is summarized in Algorithm 1. First, a low pass filter at 1 MHz is applied to avoid aliasing frequencies. Then, to work with a fixed sampling frequency and to simplify the design, the signal is decimated to focus on the low frequency range. A Hilbert transformation is used to get the
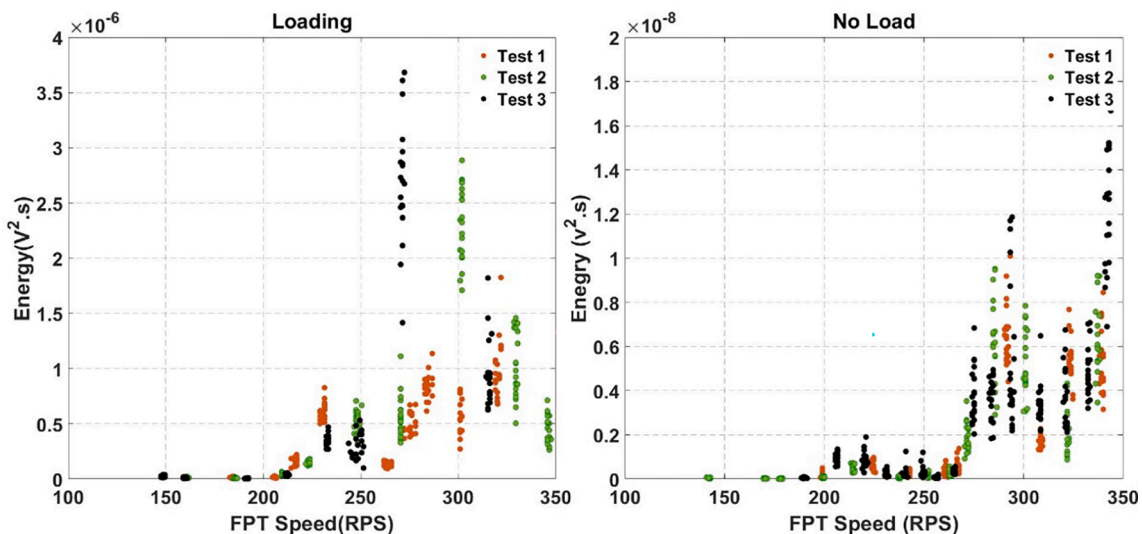

Fig. 5. AE energy vs. FPT speed for tests with and without load (three tests were conducted for every operating condition).
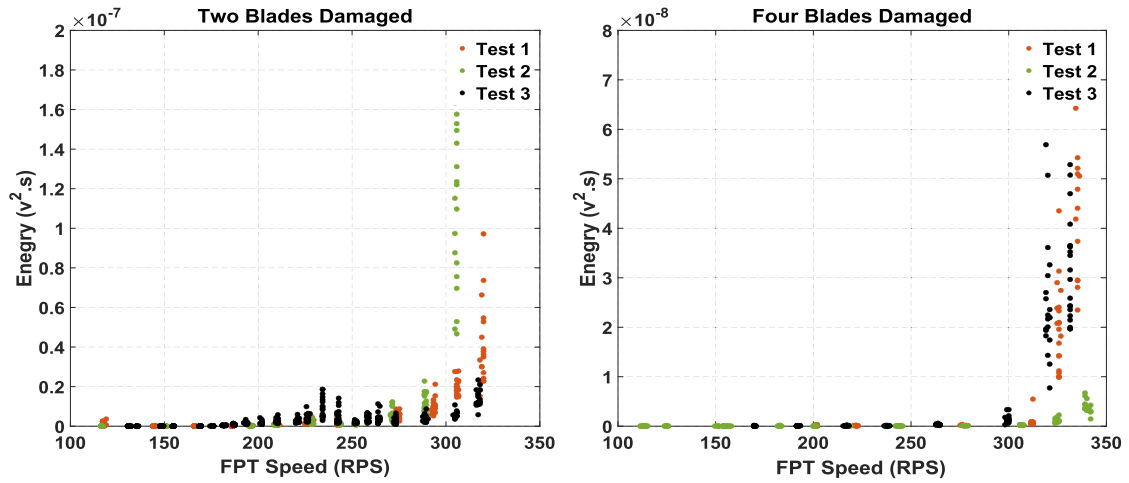
Fig. 6. AE energy vs. FPT speed for tests with damaged blades (three tests were conducted for every operating condition).
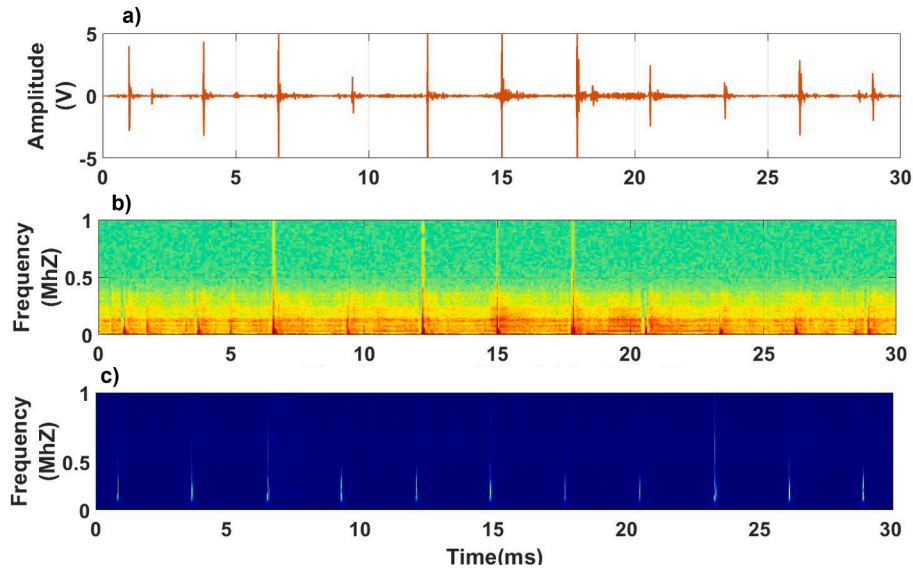

Fig. 7. Comparison between the wavelet transformation and the SFFT of a sample of raw AE: (a) raw AE signal, (b) SFFT of the raw AE signal and (c) CWT of raw AE signal.
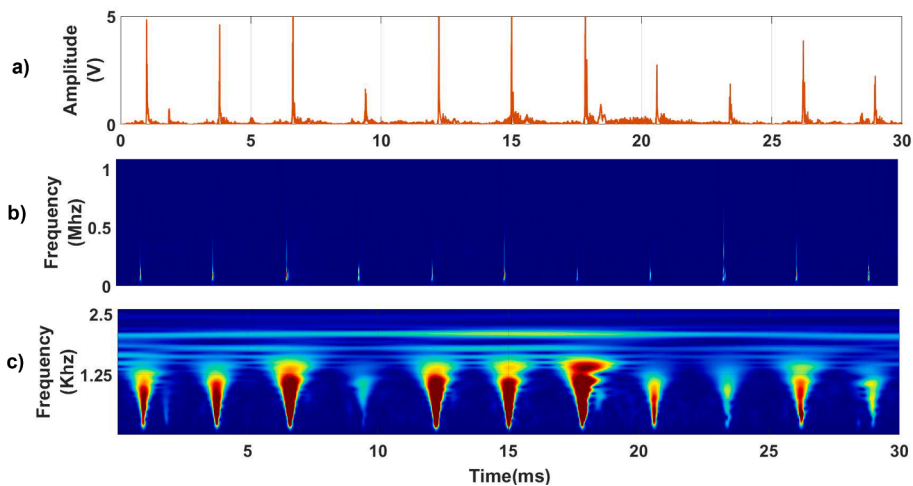

Fig. 8. CWT transformation for enveloped AE signal: (a) enveloped AE signal, (b) CWT of raw AE signal and (c) CWT of enveloped AE signal.

imaginary part of the signal and a delay filter is used to obtain the real part. The algorithm allows the extraction of an analytical signal which represents the envelope. Finally, to get the frequency component from the resulting enveloped signal the CWT is applied.
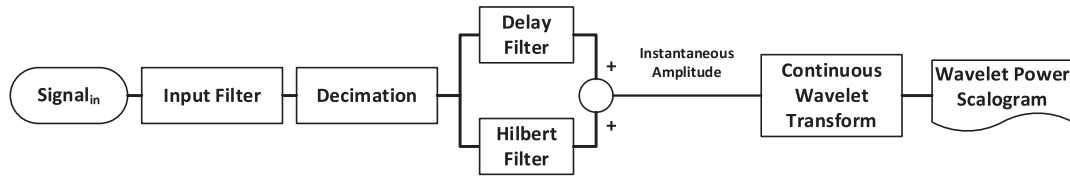
Algorithm 1: Algorithm for extracting the signal's envelope.

Fig. 8 shows the CWT for the demodulated signal where the frequency components of interest are transformed into an envelope before applying the CWT. This results in a major improvement in the computational efficiency of the CWT with improved representation of localized events. Moreover, the demodulated frequency envelope focuses on the faults and their characteristic frequencies. Thus, we can achieve improved fault and behavior characterization where interference from irrelevant high frequency components is reduced or eliminated.

4.3. Patterns in samples of data

The transformed AE signals showed different localised features in the time and frequency domains for each class of faults as can be seen in Fig. 9. The distinctive features found in each class can then be extracted by the ResNet-50. The main observations from the samples shown in Fig. 9, can be summarised as follows:

- The top row in Fig. 9 shows five common samples observed in the case of two damaged blades (D2). The domination of one or two large pulses in each sample signal is obvious as well as the spread of the response over a wide frequency range (2.5 kHz).

- The second row in Fig. 9 shows samples observed with four damaged blades (D4). Here, the number of pulses is more than those observed in the first row; the pulses are observed at lower frequencies (below 1 kHz).
- The loading test samples shown in the third row of Fig. 9 demonstrate the most complex behaviour of all types. The samples show continuity of frequency in both the lower range (below 1 kHz) and the higher range (above 1 kHz).
- The idling test samples at the bottom row of Fig. 9 show mixed behaviour between the loading test and the two damaged blades with more domination in the low-frequency range.

4.4. ResNet results

The proposed deep learning network used the images created by the CWT of the enveloped AE signals as input. This enabled efficient extraction of multiple features, thereby demonstrating a classification system capable of handling a wide range of gas turbine failures. Even if the acquired data is processed without a feature extraction step, the
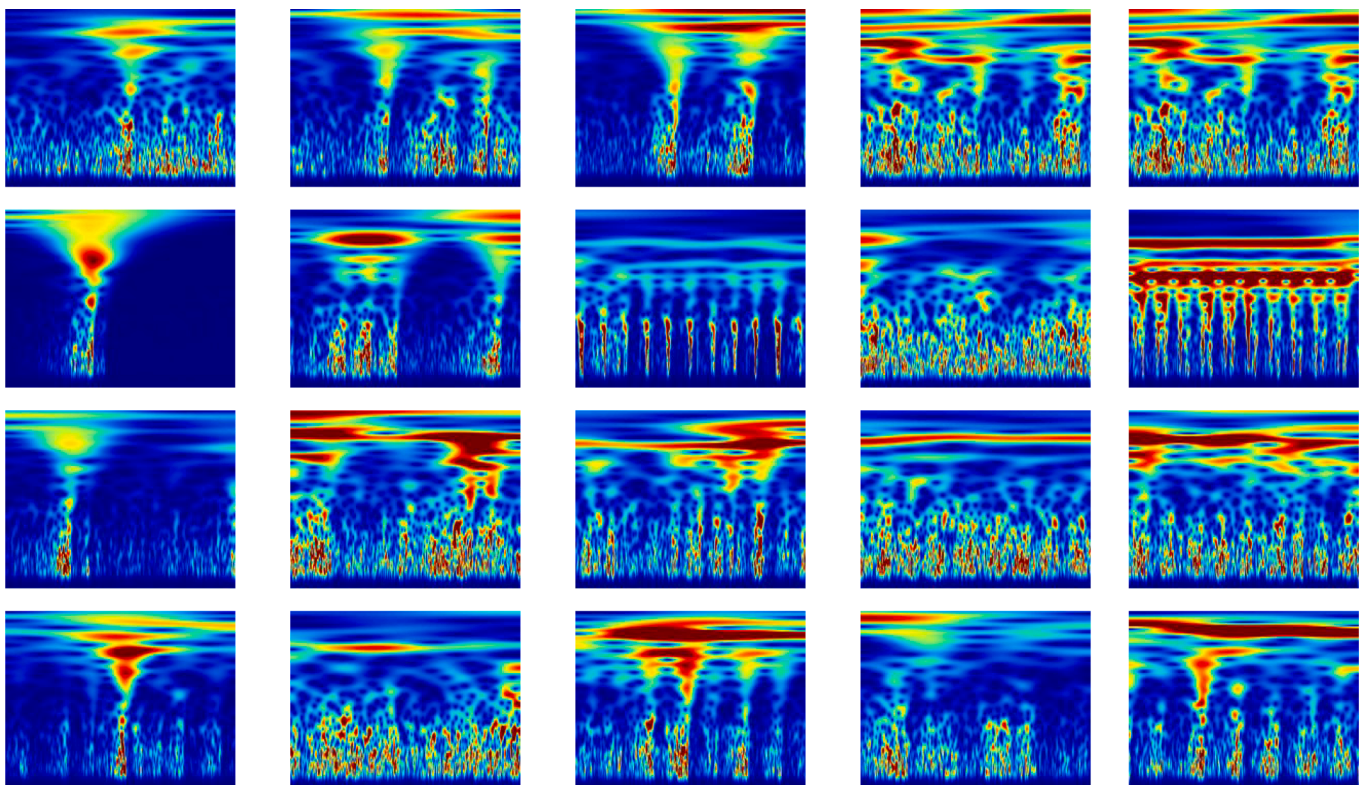

Fig. 9. Common data patterns observed in the test data (top to bottom): first row - two blade damage, second row - four blade damaged, third row - loaded running samples and bottom row - idling.

system is expected to function effectively as long as the healthy AE signal is known.

4.4.1. Implementation of the deep learning network

Several deep neural networks, including GoogleNet's ResNet-50 (Szegedy, et al., 2015), SqueezNet (Iandola et al., 2016), and AlexNet (Krizhevsky, Sutskever, & Hinton, 2012), were evaluated, and ResNet-50 exhibited the best performance on the given data. Deep neural networks have a multitude of hyperparameters that must be adjusted during the learning process, which can be challenging to optimize. However, employing pre-constructed networks such as ResNet-50 streamlines the process and limits the number of hyperparameters that must be tuned. Throughout the ResNet-50's training process, various hyperparameters were employed to optimize the network's performance, such as batch size, learning rate, adaptive moment estimation (Adam) (Kingma & Ba, 2014), and Stochastic Gradient Descent (SGD) (Rumelhart, Hinton, & Williams, 1986).

Using a large batch size may speed the computations especially when implemented on GPUs where more images can be processed at each step. Using a batch size equal to the entire dataset can increase the chance of convergence towards the global optimum. However, this may also slow the convergence and, in some cases, has the opposite effect where the solution gets stuck in a local optimum. On the other hand, using a small batch size allows faster convergence but convergence to the global optimum cannot be guaranteed. To examine this, the batch size was increased incrementally through training and the solution was observed. Batch sizes of 8 and 16 were not suitable as the accuracy of the training and the validation was poor with signs of overfitting. For example, when using a batch size of 16 images, the accuracy of the training and validation stages did not exceed 70% as demonstrated in Fig. 10(a). Batch sizes 32 and 64 showed improved convergence and good accuracy. With batch size 64, the recognition accuracy in the training and the validation was the highest and the convergence rate was the fastest without signs of overfitting.

The learning rate determines the step size at each iteration of the optimization process. A too high learning rate can cause the model to overshoot the optimal weight values, whereas a too low learning rate can lead to slow convergence. After 7,000 iterations, it was observed that a learning rate between 0.001 and 0.01 led to a faster convergence and better accuracy but only during the training stage and not the validation, which is a sign of overfitting. Fig. 10(b) shows the accuracy of the training and validation stages with a learning rate of 0.001. The network converges in 700 iterations of training, but the network

experienced overfitting as demonstrated by the gap between the accuracy of the training and that of the validation. Subsequently, increasing the learning rate to 0.1 improved the accuracy of the validation. This learning rate was adopted, and it showed no signs of overfitting as will be discussed in the next subsection.

For optimizing the network, the Adam and SGD techniques were used. The Adam technique combines the ideas of momentum optimization (Polyak, 1964) and root mean squared propagation "RMSProp" (Géron, 2019). The momentum optimization part keeps track of an exponentially decaying average of past gradients while the RMSProp keeps track of an exponentially decaying average of past squared gradients. The SGD is a backpropagation technique that computes the gradient of the network error with regard to every single model parameter. It has been reported that Adam produces slightly higher accuracy and faster convergence than SGD (Lai, Chien, Yang, & Qiang, 2019).

Deep neural networks may experience vanishing or exploding gradients during the optimization process, which can impede convergence towards the optimal solution. This occurs when the distribution of input to each layer changes as the parameters of previous layers change. To mitigate the risk of vanishing or exploding gradients during the training process, batch normalization was utilized. This technique involves zeroing the centres and normalizing each input and then scaling and shifting the results using two new vectors per layer: one for scaling and the other for shifting. This process enables the deep neural network to handle higher learning rates and to reduce the saturation effect during the learning phase.

4.4.2. Evaluating the progression of the network

The accuracy (defined in Section 4) is presented in Fig. 11(a) for the training and the validation sets. The cross-entropy loss is presented in a similar way for the validation and the training in Fig. 11(b). The cross-entropy loss is the most popular loss function to use in classification problems (Hameurlaine, et al., 2019). The two graphs demonstrate that the ResNet-50 is both stable and convergent. Throughout the training process, the accuracy and loss curves for both training and validation datasets remain close to each other, indicating the absences of overfitting. The two graphs also show that the network accuracy stays relatively steady at over 80% after 4,000 iterations. A model checkpoint function was enabled to avoid using unnecessary computations and to prevent any overfitting. The function saves and retrieves the models at the iteration point where the validation loss achieves the best accuracy. For example, Fig. 11(a) shows that the best accuracy achieved in the
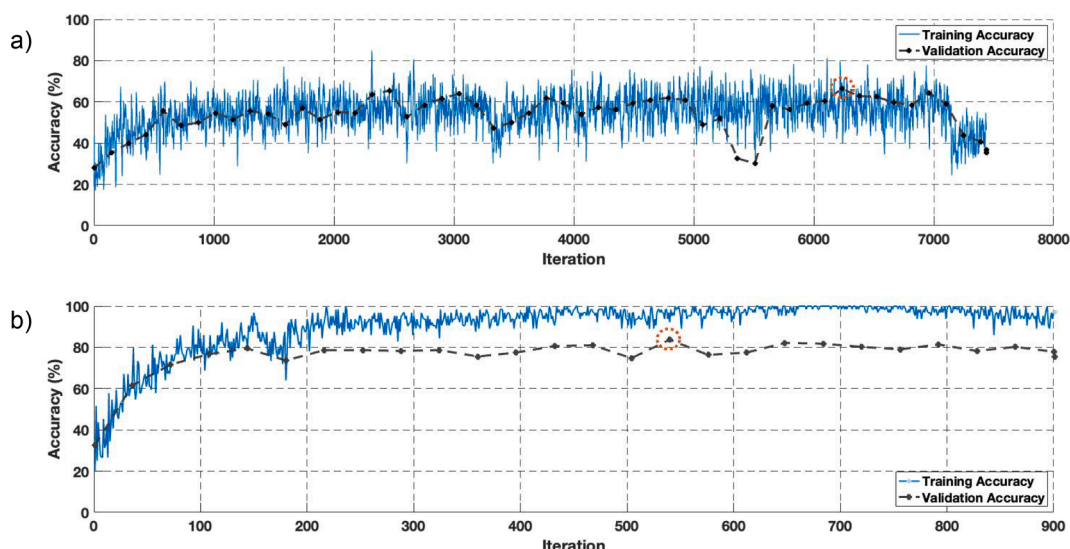

Fig. 10. Accuracy of the training with (a) low batch size and (b) low learning rate.
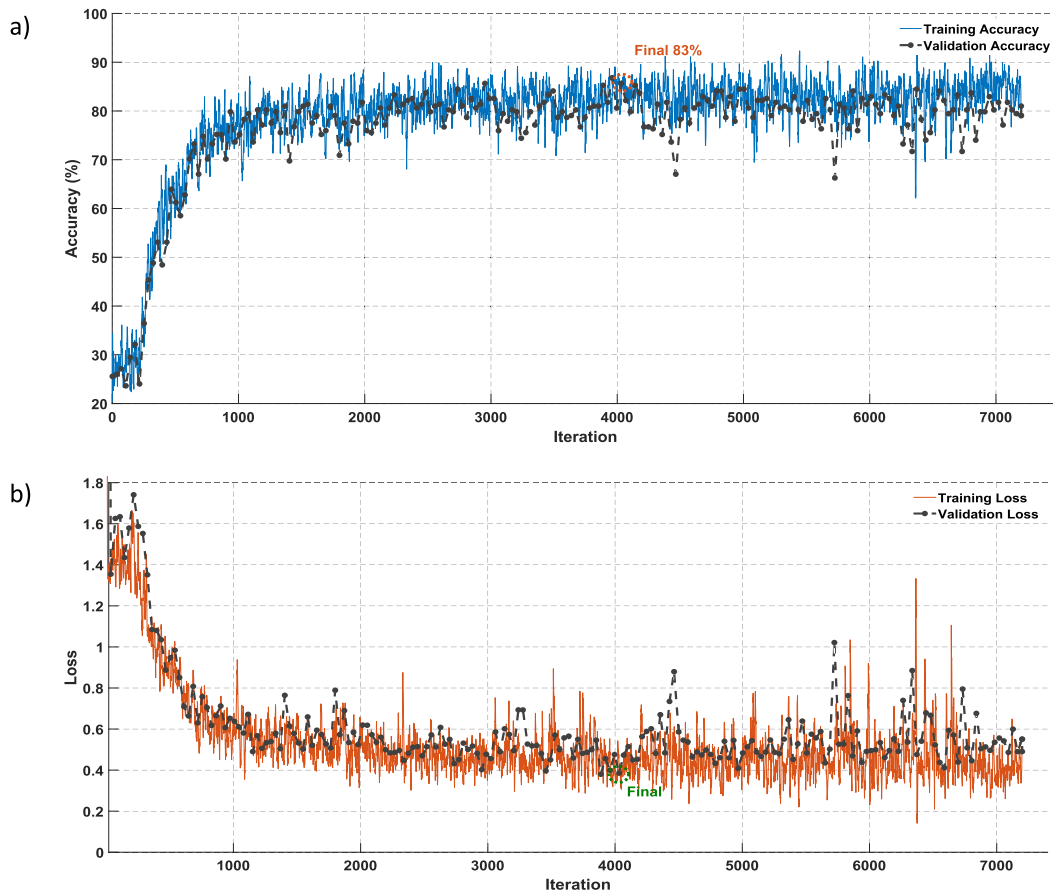
Fig. 11. ResNet-50 best performing network: a) accuracy and b) loss graphs for training and validation data.

training phase is 83% at around 4,000 iterations. After 4,000 iterations, continuing the training does not improve the results any further, and the accuracy fluctuates around 80%. Hence, the final model was fixed at 4,000 iterations which was then used for the validation.

To further investigate the performance of the trained network, the cross-validation technique was used. This technique repeatedly uses all the available data for both training and validation and can provide more robust estimates of the accuracy. This can be especially useful for deep neural networks, such as ResNet-50, which sometimes produces good accuracy for the dataset used in training, but this accuracy is not re-flected in the network performance in real-world applications. To this end, the *k*-fold cross-validation technique is used in which the data is divided into *k* equally sized sets or folds. One-fold is used as a validation dataset while the other $(k-1)$ folds are used for the training. This procedure is repeated *k*-times, each time selecting a different fold for validation and the remaining datasets for training. Consequently, *k* different estimates of the accuracy are produced from the procedure. The variance of the resulting estimate is reduced as *k* is increased, with a value of 10 being the general rule-of-thumb. The resulting array containing the 10 evaluation scores is shown in Fig. 12 where it is clear that the
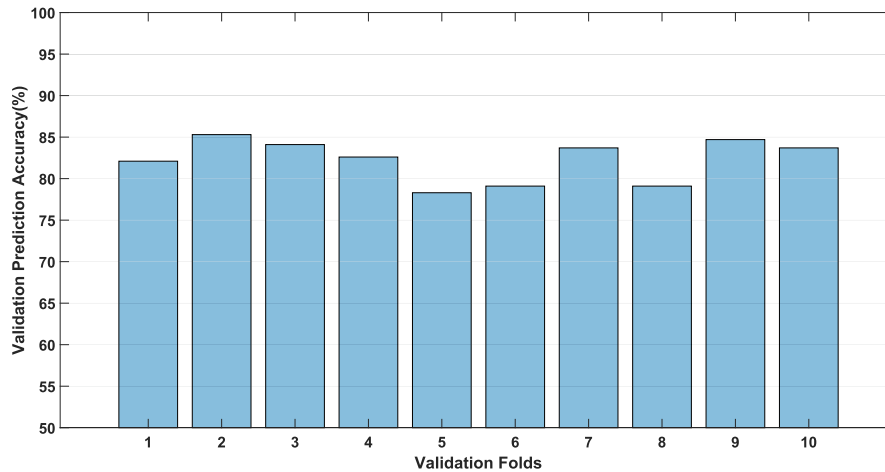

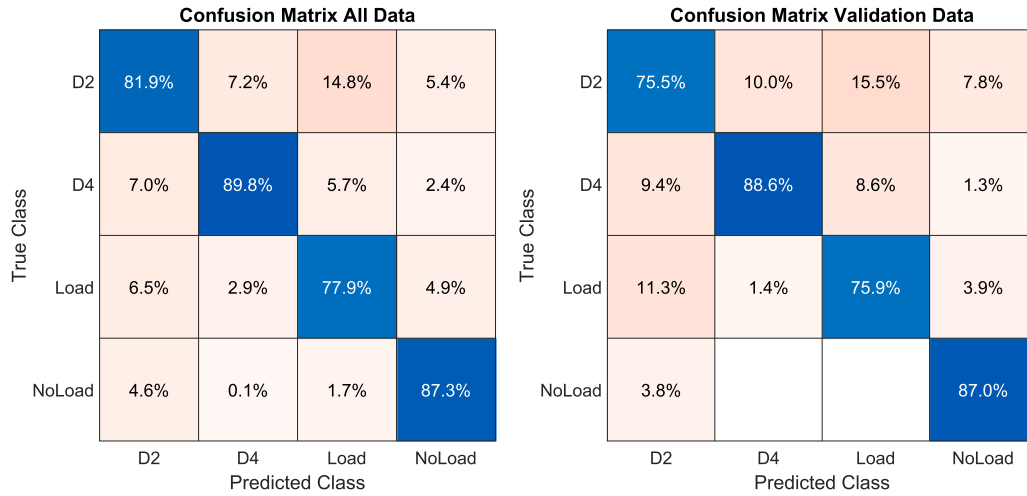**Fig. 12.** Prediction accuracy for validation data with 10 folds.

Fig. 13. ResNet 50 confusion matrix for best performing network for all and validation data (D2 represents the test with two damaged blades and D4 represents the test of four damaged blades).

variation in the accuracy is low and the performance of the best network is real and reasonable.

4.5. Classification evaluation

Fig. 13 shows the confusion matrix for all the data, as well as a separate matrix for the validation dataset. The trained model correctly predicted three classes with an accuracy greater than 81%, with the highest accuracy around 90% for D4 class. Whilst still high, the two lowest accuracy predictions are for the D2 class and loaded normal operation, and the confusion between these two classes is also relatively high. This could be attributed to the instability between tests for these two classes, as evident in Fig. 5 and Fig. 6 and to the perceived similarity between the generated CWT images as shown in the first and third rows of Fig. 9. However, resolving this issue would require more than the three relatively brief tests used in this pilot study. This may involve embedding additional statistical or temporal features in the images in order to further differentiate the four cases considered.

Another perspective on the classification accuracy is provided by the receiver-operator characteristic (ROC) curve (defined in Section 4 and shown in Fig. 13). The ROC curves for all the data plotted together and

Table 1
Performance metrics for validation data.

	Two Damaged Blades (D2)	Four Damaged Blades (D4)	Loaded	Idling
Accuracy%	82.6	82.6	82.6	82.6
AUC%	93.1	97.5	94.9	99.2
Recall%	64.5	84.9	81.5	97.1
Specificity%	88.3	81.6	82.8	77.2
Precision%	63.5	64.6	55.7	60.9
F1-measure %	64.0	73.4	66.2	74.9

then only for the validation dataset lie towards the top left corner for all classes, indicating good classification performance (see Fig. 14).

The performance metrics defined in Section 4 are summarised in Table 1 and Table 2. These include the accuracy, the area under the curve (AUC), the precision, the recall, the specificity, the true negative rate and the F1 scores. The tables demonstrate the high accuracy classification capabilities that can be achieved with the proposed algorithm.
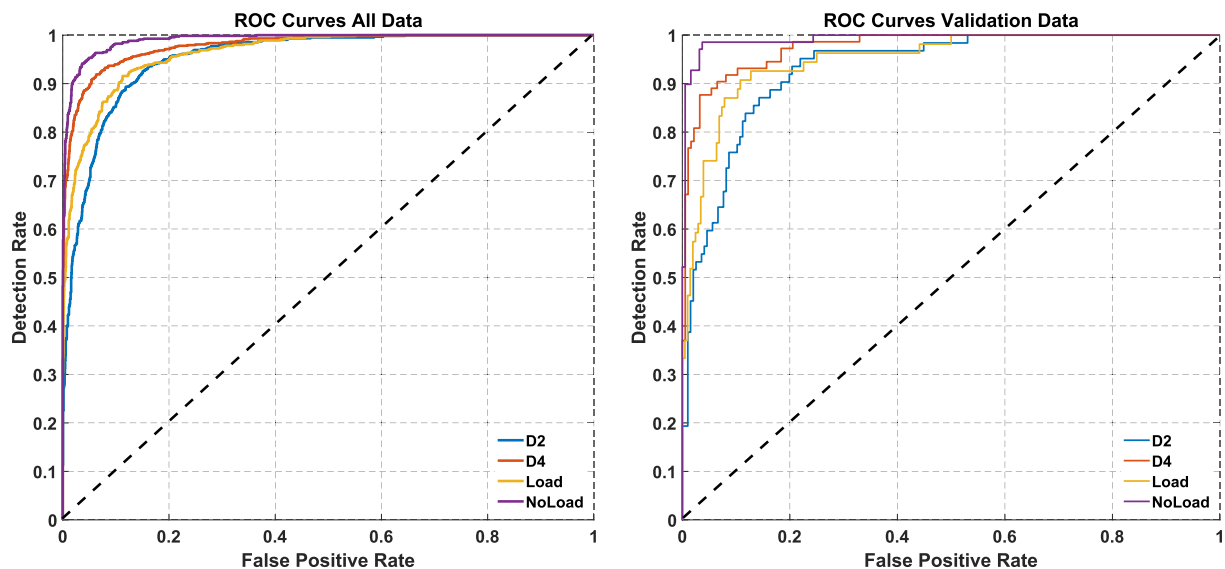

Fig. 14. ResNet-50 ROC curve for best performing network for all (left) and validation (right) datasets.

Table 2
Performance metrics for all data.

	Two Damaged Blades (D2)	Four Damaged Blades (D4)	Loaded	Idling
Accuracy%	84.7	84.7	84.7	84.7
AUC%	95.3	97.9	96.3	99.1
Recall%	71.5	87.8	83.2	94.7
Specificity%	88.9	83.5	85.1	81.1
Precision%	67.2	67.7	60.1	64.4
F1-measure %	69.3	76.5	69.7	76.7

5. Conclusion

This study presents a novel online monitoring system that utilizes acoustic emission (AE) signals to detect and classify normal and faulty operating conditions of a gas turbine. The interpretation of these signals is challenging due to two factors: (a) the fluid-induced signal can be obstructed by noise from other structure-borne AE, and (b) the signal of interest is associated with the instabilities in the combustion process which is highly nonlinear.

The proposed system overcomes these challenges by extracting a signal envelope, which is then converted into an image using the continuous wavelet transform. A convolutional neural network is developed to accurately classify four different operating conditions at a wide range of turbine speeds. Moreover, the proposed algorithm can distinguish between different cases and identify the type of fault with precision. Once the network is trained, the system can continuously monitor the running conditions with minor computational requirements and adjustments to account for modifications in the process conditions. From the perspective of a turbine operator, the presented approach requires a single AE sensor and modest computational power without any human interference which can be part of a robust continuous monitoring system.

Although this study focuses on gas turbines, the proposed method has the potential to automate complex feature extraction for any AE signal. The proposed approach offers a significant improvement in automation of fault identification and classification in comparison to other traditional methods.

Disclosure

The statements made herein are solely the responsibility of the authors, and they are not of the Qatar National Research Fund, Qatar University or Heriot-Watt University.

CRediT authorship contribution statement

M.S. Nashed: Conceptualization, Methodology, Investigation, Writing – original draft, Software, Visualization, Data curation. **J. Renno:** Conceptualization, Methodology, Validation, Formal analysis, Resources, Writing – review & editing, Visualization, Supervision, Project administration, Funding acquisition. **M.S. Mohamed:** Conceptualization, Methodology, Writing – review & editing, Project administration, Funding acquisition, Supervision. **R.L. Reuben:** Data curation, Resources, 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.

Data availability

I have shared the link to my code in the submission.

References

Bentley, D. E. (1974). Forced subratative speed dynamic action of rotating machinery. *ASME, 74*.

Bhaumik, S. K., Bhaskaran, T. A., Rangaraju, R., Venkataswamy, M. A., Parameswara, M. A., & Krishnan, R. V. (2002). Failure of Turbine rotor blisk of an aircraft engine. *Engineerin Failure Analysis, 9*, 287–301.

Bhaumik, S. K., Sujata, M., Venkataswamy, M. A., Parameswara, M.A. (2006). Failure of a low pressure turbine rotor blade of an aeroengine. *Engineerin Failure Analysis*.

Board, D. (2000). Stress wave analysis of turbine engine faults. *IEEE Aerospace Conference Proceedings, 6*, 79–93.

Caesarendra, W., Kosasih, B., Tieu, A. K., Zhu, H., Moodie, C. A. S., & Zhu, Q. (2016). Acoustic emission-based condition monitoring methods: Review and application for low speed slew bearing. *Mechanical Systems and Signal Processing, 72–73*, 134–159.

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

Chen, Z., Xia, T., Li, Y., & Pan, E. (2022). Random-Effect Models for Degradation Analysis Based on Nonlinear Tweedie Exponential-Dispersion Processes. *IEEE Transactions on Reliability, 71*, 47–62.

Cussons, T. l. (2004). P9005_8 – Cussons Technology – Two Shaft Gas Turbine manual. In.

Cyrus, B., Homji, M., Bromley, A. (2004). Gas turbine axial compressor fouling and washing. In *proceedings of the thirty-third turbomachinery symposium*.

Douglas, M., Beugné, S., Jenkins, M. D., Frances, A. K., Steel, J. A., Reuben, R. L, and Kew, P. A. (2004). Monitoring of gas turbine operating parameters using acoustic emission In *EWGAE, DGZfP-Proceedings BB 90-CD*.

Fan, Y. E. (2007). *The Condition Monitoring of Mechanical Seals Using Acoustic Emissions*. ProQuest Dissertations Publishing.

Géron, A. (2019). *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems*. O'Reilly Media.

Hall, L. D., & Mba, D. (2004). Diagnosis of continuous rotor-stator rubbing in large scale turbine units using acoustic emissions. *Ultrasonics, 41*, 765–773.

Hameurlaine, M., Moussaoui, A., & Safa, B. (2019). *Deep Learning for Medical Image Analysis*.

Haykin, S., & Lippmann, R. (1994). Neural networks, a comprehensive foundation. *International journal of neural systems, 5*, 363–364.

He, K., Zhang, X., Ren, S., & Sun, J. (2015). Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE international conference on computer vision* (pp. 1026-1034).

He, R., Li, X., Chen, G., Chen, G., & Liu, Y. (2020). Generative adversarial network-based semi-supervised learning for real-time risk warning of process industries. *Expert Systems with Applications, 150*, Article 113244.

He, Y., Li, M., Meng, Z., Chen, S., Huang, S., Hu, Y., & Zou, X. (2021). An overview of acoustic emission inspection and monitoring technology in the key components of renewable energy systems. *Mechanical Systems and Signal Processing, 148*, Article 107146.

Hou, D., Qi, H., Wang, C., & Han, D. (2022). High-speed train wheel set bearing fault diagnosis and prognostics: Fingerprint feature recognition method based on acoustic emission. *Mechanical Systems and Signal Processing, 171*, Article 108947.

Hou, J., Wicks, B. J., & Antoniou, R. A. (2002). An investigation of fatigue failures of turbine blades in a gas turbine engine by mechanical analysis. *Engineering Failure Analysis, 9*, 201–211.

Hubel, D. H., & Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *The Journal of Physiology, 195*, 215–243.

Iandola, F. N., Moskewicz, M. W., Ashraf, K., Han, S., Dally, W. J., & Keutzer, K. (2016). SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <1MB model size. *ArXiv, abs/1602.07360*.

Jiang, Y., Xia, T., Wang, D., Fang, X., & Xi, L. (2022). Spatiotemporal denoising wavelet network for infrared thermography-based machine prognostics integrating ensemble uncertainty. *Mechanical Systems and Signal Processing, 173*, Article 109014.

Jiang, Y., Xia, T., Wang, D., Zhang, K., & Xi, L. (2022). Joint adaptive transfer learning network for cross-domain fault diagnosis based on multi-layer feature fusion. *Neurocomputing, 487*, 228–242.

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

König, F., Jacobs, G., Stratmann, A., & Cornel, D. (2021). Fault detection for sliding bearings using acoustic emission signals and machine learning methods. In *IOP Conference Series: Materials Science and Engineering* (Vol. 1097, pp. 012013): IOP Publishing.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems, 25*.

Kurz, R., & Brun, K. (2001). Degradation in gas turbine systems. *Transaction of the ASME, 123*, 70–77.

Lai, C.-F., Chien, W.-C., Yang, L. T., & Qiang, W. (2019). LSTM and edge computing for big data feature recognition of industrial electrical equipment. *IEEE Transactions on Industrial Informatics, 15*, 2469–2477.

Liao, Y., Ragai, I., Huang, Z., & Kerner, S. (2021). Manufacturing process monitoring using time-frequency representation and transfer learning of deep neural networks. *Journal of Manufacturing Processes, 68*, 231–248.

Lu, Y., Wang, F., Jia, M., & Qi, Y. (2016). Centrifugal compressor fault diagnosis based on qualitative simulation and thermal parameters. *Mechanical Systems and Signal Processing, 81*, 259–273.

Mann, B. S. (1998). Particle erosion- anew concept of flow visualization and boundary layer investigation of machines at high reynolds number. *Wear, 223*, 110–118.

Mazur, Z., Luna-amirez, A., Juarez-Islas, J. A., & Campos-Amezcua, A. (2005). Failure analysis of a gas turbine blade made of Iconel 738LC alloy. *Engineerin Failure Analysis, 12*, 474–486.

Mba, D., Cooke, A., Roby, D., & Hewitt, G. (2004). Detection of shaft-seal rubbing in large-scale power generation turbines with acoustic emission. Case study. *Proceedings of the Institution of Mechanical Engineers, Part A: Journal of Power and Energy, 218*, 71–81.

Mba, D., & Hall, L. D. (2002). The transmission of acoustic emission across large scale turbine rotors. *NDT & E International, 35*, 529–539.

Metwally, M. T., & Hamed, W. A. (1995). Blade erosion in automotive gas turbine engine. *Journal of Engineering for Gas Turbine and Power, Transactions of the ASME, 117*, 213–219.

Nashed, M. S., Steel, J. A., & Reuben, R. L. (2014). The use of acoustic emission for the condition assessment of gas turbines: Acoustic emission generation from normal running. *Proceedings of the Institution of Mechanical Engineers, Part E: Journal of Process Mechanical Engineering, 228*, 286–308.

Polyak, B. T. (1964). Some methods of speeding up the convergence of iteration methods. *USSR Computational Mathematics and Mathematical Physics, 4*, 1–17.

Ranjan, N., Bhandari, S., Khan, P., Hong, Y.-S., & Kim, H. (2021). Large-scale road network congestion pattern analysis and prediction using deep convolutional Autoencoder. *Sustainability, 13*, 5108.

Ranjan, N., Bhandari, S., Zhao, H. P., Kim, H., & Khan, P. (2020). City-wide traffic congestion prediction based on CNN, LSTM and transpose CNN. *IEEE Access, 8*, 81606–81620.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature, 323*, 533–536.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., … Bernstein, M. (2015). Imagenet large scale visual recognition challenge. *International journal of computer vision, 115*, 211–252.

Sergey, I., & Christian, S. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. In (Vol. 37, pp. 448-456): PMLR.

Shahkar, S., & Khorasani, K. (2019). Gas Turbine Condition Monitoring Using Acoustic Emission Signals. *Journal of nondestructive evaluation, diagnostics and prognostics of engineering systems, 2*.

Shanbhag, V. V., Meyer, T. J. J., Caspers, L. W., & Schlanbusch, R. (2020). Condition monitoring of hydraulic cylinder seals using acoustic emissions. *International Journal of Advanced Manufacturing Technology, 109*, 1727–1739.

Shen, D., Wu, G., & Suk, H.-I. (2017). Deep learning in medical image analysis. *Annual Review of Biomedical Engineering, 19*, 221–248.

Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing & Management, 45*, 427–437.

Surucu, O., Gadsden, S. A., & Yawney, J. (2023). Condition monitoring using machine learning: A review of theory, applications, and recent advances. *Expert Systems with Applications, 221*, Article 119738.

Szegedy, C., Wei, L., Yangqing, J., Sermanet, P., Reed, S., Anguelov, D., … Rabinovich, A. (2015). Going deeper with convolutions. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 1–9).

Tong, Z., Xie, S., Liu, H., Zhang, W., Pei, C., Li, Y., … Takagi, T. (2020). An efficient electromagnetic and thermal modelling of eddy current pulsed thermography for quantitative evaluation of blade fatigue cracks in heavy-duty gas turbines. *Mechanical Systems and Signal Processing, 142*, Article 106781.

Xia, T., Zhuo, P., Xiao, L., Du, S., Wang, D., & Xi, L. (2021). Multi-stage fault diagnosis framework for rolling bearing based on OHF Elman AdaBoost-Bagging algorithm. *Neurocomputing, 433*, 237–251.

YU, J. J., Goldman, P., Bentley, D. E., Muzynska, A. (April 2002). Rotor/Seal experimental and analytical study on full annular rub. *Transactions of the ASME, 124*.

Zhou, C.a., Yang, B., Guo, K., Liu, J., Sun, J., Song, G., Zhu, S., Sun, C., & Jiang, Z. (2020). Vibration singularity analysis for milling tool condition monitoring. *International Journal of Mechanical Sciences, 166*, Article 105254.

Zhou, D., Huang, D., Hao, J., Wu, H., Chang, C., & Zhang, H. (2021). Fault diagnosis of gas turbines with thermodynamic analysis restraining the interference of boundary conditions based on STN. *International Journal of Mechanical Sciences, 191*, Article 106053.