Article:
Munko, M.J., Cuthill, F., Camacho, M.A.V. et al. (2025) An audio-based framework for anomaly detection in large-scale structural testing. Engineering Applications of Artificial Intelligence, 142. 109889. ISSN: 0952-1976

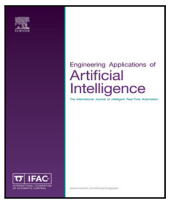
https://doi.org/10.1016/j.engappai.2024.109889

Engineering Applications of Artificial Intelligence 142 (2025) 109889

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

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



An audio-based framework for anomaly detection in large-scale structural testing

Marek J. Munko [a], Fergus Cuthill [a], Miguel A. Valdivia Camacho [a], Conchúr M. Ó Bradaigh [b], Sergio Lopez Dubon [a,*]

[a] School of Engineering, The University of Edinburgh, Sanderson Building, Robert Stevenson Road, The King's Buildings, Edinburgh, EH9 3FB, UK
[b] Faculty of Engineering, University of Sheffield, Western Bank, Sheffield, S10 2TN, UK

ARTICLE INFO

Keywords:
Anomaly detection
Fatigue testing
Artificial neural network
Convolutional autoencoder
Wavelet scattering

ABSTRACT

FastBlade is a research facility that tests large-scale composite and metal structures. To maximise its throughput by uninterrupted running of experiments, unmanned operation of the site is desired. One of its key enablers is anomaly detection, where microphones are used as a non-specific, affordable, and well-established sensing method. The dataset collected during the operation of the system consists of both normal and anomalous samples, which we need to classify. The problems associated with the dataset involve significant intraclass variability of the normal operation samples, as well as the scarcity of anomalous data, increasing the complexity of the classification problem. In this work, we evaluate the performance of several tools for time–frequency signal analysis, which are used to extract features from the original high-dimensional signal. We choose to apply the wavelet scattering transform (WST) due to its remarkable performance. Based on the findings from the literature review, we first rely on the reconstruction error of the processed WST images to detect anomalous samples. However, due to the nature of the dataset, both the convolutional autoencoder (CAE) and the principal component analysis (PCA) transform turn out to be unsuccessful. We then investigate the hidden layers of the CAE in search of features that can be used to separate normal and anomalous samples. Having identified the most suitable candidates, we discover that applying the normalised cross-correlation (NCC) to measure the similarity of the generic features generated and our dataset results in satisfactory separation. We train a number of classifiers and test the method on unseen data. The model's accuracy is 99.58%, with a recall of 100% and 92% on normal and anomalous operation samples, respectively. The model's accuracy and low latency prove the WST's suitability for robust, real-time detection of different anomaly types. Therefore, the method can be deployed in systems with limited information about the critical assets and can be easily extrapolated to other setups.

1. Introduction

Anomaly detection (AD) aims to identify patterns in a particular dataset that do not conform to the expected behaviour. By identifying periods of abnormal operation in an industrial or experimental process, relevant datasets can be analysed to examine the nature of a fault. Alternatively, AD algorithms can run in real-time and give a warning of a persistent anomaly, which might prompt a site operator to halt the process. This paper focuses on the operation of FastBlade, a research facility for testing large-scale composite and metal structures (Lopez Dubon et al., 2023d,c,b). While this paper investigates a feasible method of determining anomalies in FastBlade's operation, it is seamlessly transferable to systems in experimental and industrial settings.

1.1. FastBlade – the test facility

FastBlade is a research facility at the University of Edinburgh, located in Rosyth, Scotland. The goal of the facility is to test the structural response of large slender structures (2–14 m) under either static or fatigue loads (Lopez Dubon et al., 2023d,c,b), as shown in Fig. 1. The facility's primary focus is the examination of tidal turbine blades, which, in turn, helps exploit the potential of the tidal energy sector in the UK and worldwide. It is estimated that the UK alone has a tidal energy potential of 50 TWh/year along its coastline (Burrows et al., 2009), while a total tidal stream capacity of 101 GW is expected to be installed by 2050 (Todeschini et al., 2022). To validate the design of

* Corresponding author.
E-mail addresses: M.Munko@ed.ac.uk (M.J. Munko), Fergus.Cuthill@ed.ac.uk (F. Cuthill), MA.Valdivia@ed.ac.uk (M.A.V. Camacho), C.OBradaigh@sheffield.ac.uk (C.M.Ó. Bradaigh), Sergio.LDubon@ed.ac.uk (S.L. Dubon).

https://doi.org/10.1016/j.engappai.2024.109889
Received 13 May 2024; Received in revised form 9 December 2024; Accepted 14 December 2024

Fig. 1. The test setup at FastBlade. A tidal turbine blade is mounted on the reaction frame and supported by three actuators, which deform the blade during a test. The speckle pattern and the sensors mounted on the blade are used to monitor its structural performance.

a blade at FastBlade, the blade is loaded for many cycles, significantly exceeding the number of tide changes it would experience during its subsea deployment. By simulating the wear of a blade in such a way, its properties can be examined at an accelerated pace, reducing the risk associated with deploying tidal power technology.

1.2. Problem statement

For FastBlade and other testing facilities, the most time-efficient way of operating the facility is by running the tests continuously day and night until the target number of cycles (for fatigue tests) or target deflection time (for static tests) is reached. Therefore, allowing unmanned operation at FastBlade is crucial for ensuring its commercial success. However, operating the facility with few or no members of staff on-site is associated with multiple risks in the areas of:

- ***Safety*** - the system utilises high voltage and exerts high dynamic and static loads, increasing the fire hazard and constituting a danger for unauthorised interactions.
- ***Machinery Health*** - in addition to safety hazards, resuming operation with a failure of one of the system assets may accelerate its wear, increasing servicing costs.
- ***Test Credibility*** - running the test with changed parameters or with a data-logging fault does not meet the test criteria (crucial for certification purposes), introducing irreversible wear into the specimen.

These factors show the need for a versatile, anomaly-monitoring tool to detect errors of multiple origins, such as system imbalance, third-party interactions, or changing test parameters. Moreover, operating in (quasi-) real-time must be computationally feasible for the desired application.

1.3. Contributions and structure

The two main contributions of this work are:

- The development of a high-accuracy audio-based AD technique using a dataset with significant intraclass variations among normal samples, and the scarcity of anomalous samples.
- The evaluation of the performance of the wavelet scattering transform (WST) as a feature extraction tool in the detection of transient, non-periodic anomalies.

In this paper, we first describe the state-of-the-art solutions in the area of audio-based AD, and analyse the methodology in light of the particular features of our dataset, such as intraclass variability, and a small number of anomalous samples. We compare some of the standard time–frequency signal processing tools against the WST, whose mathematical properties should make it well-suited for the discussed application. Our evaluation involves visual comparison, as well as execution time benchmarking.

We subsequently develop an AD method based on the classification of samples using the reconstruction error, which is the most commonly found solution in the literature (Fiore et al., 2022). We then demonstrate that neither our deep-learning model nor a principal component analysis (PCA)-based model succeeds in providing sufficient separation between normal and anomalous data. As a result, we investigate the hidden layers of the deep-learning model in terms of feature maps that can be used to reduce the variability of the normal samples while providing good interclass separation. We evaluate the developed methodology using standard benchmarking tools for classification and discuss the model's strengths and shortcomings.

2. Related work

Choosing the correct parameters to analyse in a condition monitoring problem is an important decision for an engineer. The measurements of a system need to provide enough insight into the operation of an asset to make a confident decision about the state of its operation. Simultaneously, collecting too much data should be avoided as it is associated with higher data redundancy and computational challenges resulting from increased data dimensionality. Other factors to consider when choosing appropriate monitoring methods include the cost and constraints of the communication system.

Audio and vibration signals are widely used for AD in a variety of engineering systems. The records of forecast and diagnosis solutions, which often incorporate the use of deep-learning techniques, can be found in the monitoring of gearboxes (Zhao et al., 2023), bearings (Zhao et al., 2022, 2024a; Zhu et al., 2024), compressors (Mobtahej et al., 2021), and industrial machines, such as pumps, valves and fans (Fiore et al., 2022; Muller et al., 2021). High sampling frequency characterises both audio and vibration signals, and therefore, similar processing tools are utilised for classification. Considering our research goal, which is the detection of system-wide anomalies, it is decided that microphones should be used in the investigation. Since audio data is not specific to a given asset, it provides better coverage than using a fusion of other sensor types, which is highlighted as a limitation of AD-related methods (Zhao et al., 2024b). Moreover, unlike vision-based techniques, audio signals are not impacted by varying illumination levels and occlusion (de Carvalho et al., 2019).

The typical processing pipeline for classifying high-frequency signals includes three crucial stages, namely feature extraction, clustering, and classification. Feature extraction aims to decrease the redundancy of the collected information and derive the properties of the signal that are important to the classification task. Clustering transforms the data into a particular dimensional space, where the distances between samples belonging to the same class, known as intraclass distances, are small, while the interclass distances are large to allow for effective classification.

In this section, we review the feature extraction and clustering tools used by other researchers in anomaly detection problems utilising audio data. Subsequently, we review works which incorporate the wavelet scattering transform (WST) into the processing pipeline. Based on its mathematical properties (described in detail in Section 3.3), WST should provide a competitive solution to the existing feature-extraction methods, and its application to audio-based anomaly detection problems has not yet been validated.

2.1. The summary of related works

Table 1 presents the most relevant publications found in the area of AD using audio signals. The table includes the application domain and the main features of the processing pipeline, as well as comments, providing a brief summary of each application and a comparison to this study. In all presented publications, promising accuracy results can be found, proving the suitability of using microphones to solve AD tasks. It can also be noted that different processing methods have been used, incorporating various feature extraction techniques, as well as supervised and unsupervised learning techniques for clustering.

2.1.1. Feature extraction

The work presented in Ahn and Yeo (2021) is the only of the considered cases in which the audio signal recorded is not transformed into the time–frequency domain, but is solely considered in the time-domain. The authors still pre-process the signal before attempting clustering by down-sampling and combining multiple channels into a 2D image. The most common feature extraction techniques found in other works are short-time Fourier transform (STFT) and mel-spectrograms. In fact, these algorithms are very similar, as computing mel-spectrograms usually requires the computation of the STFT, which is subsequently filtered with a Mel-frequency filter bank (Ustubioglu et al., 2023). The use of mel-frequency cepstral coefficients (MFCC) is found in two publications. MFCC also requires computing the Fourier transform of the windowed signal, followed by the computation of the discrete cosine transform (DCT) on the power spectrum of the signal (Sahidullah and Saha, 2012), making it similar to the previously mentioned mel-spectrograms.

The analysis of the related works makes it evident that features relevant to AD can be obtained from an audio signal by analysing it in the time–frequency domain. Another aspect which is important for feature extraction is setting the duration of each of the samples considered. While, for example, it is set to three seconds in Mobtahej et al. (2021), and ten seconds in Fiore et al. (2022) and Muller et al. (2021), it is believed that a window of 0.5 s should be used in order to capture the transient nature of anomalies recorded at FastBlade (more detailed explanation is given Section 4.2). Moreover, since there exist other ways to analyse high-frequency signals in the time–frequency domain, such as continuous wavelet transform (CWT) and discrete wavelet transform (DWT) (Arts and van den Broek, 2022), in this work, we would like to evaluate the performance of a number of algorithms on the samples collected. Due to its interesting properties, we would also like to focus on the wavelet scattering transform (WST), whose common applications in feature extraction are described in Section 2.2.

2.1.2. Clustering

Considering the choice of correct clustering techniques, which serve as a stepping stone to classification, one of the key considerations is the ratio of normal and anomalous samples in the dataset. Since we have recorded little anomalous data at FastBlade, the techniques such as densely-connected or convolutional neural networks adopted in Kuo et al. (2022), Ahn and Yeo (2021), or McLoughlin et al. (2017), where extensive anomalous datasets are available, are not relevant to our study. It is, therefore, observed that a successful clustering technique in the study is more likely to rely on unsupervised rather than supervised learning. It is also known that large intraclass variations characterise the normal operation audio samples recorded at FastBlade. One of the implications can be the fact that conventional machine learning techniques, such as PCA and SVM applied in Mobtahej et al. (2021), might not be sufficient to achieve satisfactory separation between the varied normal and anomalous samples.

The authors in Fiore et al. (2022) present the application of audio-based AD to industrial machines. The review of the related works found in the publication states that all of the state-of-the-art AD approaches rely on the use of an autoencoder (AE) in the clustering stage (AEs are described in more detail in Section 3.5). The detection of anomalies is possible when AEs are trained on normal data and reconstruct anomalous samples with an increased reconstruction error. The recurrent neural network (RNN)-based AE presented in Park and Yun (2018) is characterised by a relatively small number of parameters and quick training time. Similarly, the authors of Lo Scudo et al. (2023) utilise both a densely-connected AE and a convolutional autoencoder (CAE) to often outperform larger state-of-the-art models in the AD tasks. However, since the authors do not mention intraclass variations explicitly, it is hypothesised that due to the small size of the models, they would not be able to learn to reconstruct different normal operation patterns with a satisfactory degree of accuracy, and would not result in a clear separation between normal and anomalous samples. A CAE is used in Oh and Yun (2018) to reconstruct samples collected during the operation of an integrated-circuit assembly machine. While the authors claim that a certain overlap between the reconstruction error of normal and anomalous samples is allowed due to the nature of the machine, it might prove to be too large for deployment at FastBlade, where long-lasting anomalies might lead to a critical system failure.

Intraclass variability within the normal operation samples can be seen in Muller et al. (2021), where the spectrograms of multiple normal operation samples are presented. The fact that the authors have only a small proportion of anomalous samples in the dataset is another similarity with our work. It is recommended in the publication that a pre-trained deep-learning feature extractor is used, such as a ResNet convolutional neural network (CNN). On the other hand, the samples considered in the work are significantly longer, and the AD accuracy, evaluated using the area under curve (AUC) metric, varies between 61.9% and 99.6%, depending on the particular asset monitored.

Given the widespread use of CAEs in the area, it is decided that the classification approach based on the analysis of the reconstruction error should be attempted first. However, as it has also been noted, the resulting separation might not always be satisfactory, particularly when differences between normal signal samples are present across a wide range of frequencies. In this case, a feature-extraction method using deep-learning-based filters can be incorporated. However, instead of experimenting with various pre-trained CNNs, the hidden layers of the CAE trained on normal operation samples will be examined.

2.2. Applications of the wavelet scattering transform

As shown in Section 2.1.1, using STFT for AD using audio data is widespread. The use of STFT attempts to address the problem of signals changing in time, which is crucial for AD problems, as many machinery faults are, in nature, nonperiodic (Pan and Sas, 1996). However, one of the shortcomings of STFT is that it has a constant resolution across all frequencies (Peng and Chu, 2004). Therefore, the application of wide windows, which is desired for analysing low-frequencies of a signal, results in poor resolution of its high-frequency components. Applying narrow windows results in the opposite effect since they impede the analysis of the low-frequency components. A way to mitigate this problem is to apply wavelet transform (WT), which incorporates filters of varying frequency. A specific embodiment of WT, which has proven to be computationally efficient and well-suited for classification problems, is wavelet scattering transform (WST) (Bruna and Mallat, 2013). The introduction of the transform and the reasons for its advantages over conventional methods are presented in Section 3.3.

The literature review has found applications of the WST in AD problems, including the classification of motor current (Toma et al., 2022), electrocardiogram (ECG) (Liu et al., 2020; Sharaf, 2023), and electroencephalogram (EEG) signals (Buriro et al., 2021; Ahmad et al., 2017). There also exist works where WST has been applied in various audio classification (Anden and Mallat, 2014) and texture discrimination tasks (Sifre and Mallat, 2013). However, no WST application in industrial audio-based AD has been found.

Table 1
The summary of the related works considered.

Application domain	Feature extraction	Clustering	Ref.	Comments
Industrial machines — pumps, valves, slide rails, and fans	Mel spectrograms	Convolutional Autoencoder (CAE); Long Short-Term Memory (LSTM) Autoencoder — both supplemented with asset IDs	Fiore et al. (2022)	The proposed method improves upon the commonly-adopted autoencoder-based AD technique by incorporating the IDs of particular assets into the training process. However, in our work, sound is recorded collectively for all system assets.
Surface-Mount Device (SMD) Assembly Machine	Short-Time Fourier Transform (STFT)	CAE	Oh and Yun (2018)	The publication presents a successful real-time classification of chosen anomalies in the machine's operation dataset. However, the level of separation between normal and anomalous samples obtained by computing the residual error might not be sufficient for our study.
SMD Assembly Machine	STFT	Two variations of Recurrent Neural Network (RNN) Encoder–Decoder; CAE	Park and Yun (2018)	The authors decide to apply STFT for feature extraction, instead of Mel-frequency cepstral coefficients (MFCC), to preserve more signal information. A successful, compact RNN Encoder–Decoder model is trained. The variation between normal operating samples is not presented. The model size might turn out to be insufficient when considerable intraclass variation is present in the dataset.
Compressor operation	STFT, MFCC, ResNet50; Spectral Centroid	Principal Component Analysis (PCA) and Support Vector Machine (SVM)	Mobtahej et al. (2021)	The study shows that the features extracted by MFCC result in better accuracy than STFT. The addition of information on spectral centroids further improves the accuracy. However, the details about the intraclass variation of the signals are not given.
Industrial machines — pumps, valves, slide rails and fans	Mel-spectrogram and a pre-trained CNN	CAE; one class SVM; kernel density estimation; Isolation Forest; Gaussian Mixture Model; Bayesian Gaussian Mixture Model	Muller et al. (2021)	The problem presented bears similarities to our research problem: the anomalous samples are scarce, and there is significant variation between samples in the normal operation class. The paper, therefore, incorporates the use of pre-trained CNNs for feature extraction. The results of the best-performing models show some variation depending on the monitored asset, with the area under curve metric varying between 61.9% and 99.6%.
Industrial machines — pumps, valves, slide rails, and fans; toy trains and cars, harmonic drives	Wavelet-denoising, Mel-spectrogram	Pre-trained convolutional neural network (CNN)	Kuo et al. (2022)	The successful implementation of AD using audio is presented. It features wavelet-based denoising and a CNN architecture for clustering and classification. However, the share of anomalous samples in the dataset exceeds what is available in our real-life application.
Car operation	Segmentation of the signal and down-sampling in the time domain, combining multiple channels into 2D images	SVM; K-means; K-nearest neighbours (KNN); CNN	Ahn and Yeo (2021)	The research shows the supremacy of the CNN over the other three techniques used to classify anomalous sounds. However, the share of the anomalous training data (more than 50% of the samples) far exceeds the proportion of the dataset considered in our research.
Various sound recordings mixed with different background noise levels.	Spectrogram image features	Deep neural network (DNN); CNN	McLoughlin et al. (2017)	The method uses an efficient feature extraction technique and successfully assigns anomalous samples to their respective class. However, in the case of supervised learning, the outcomes can be difficult to reproduce when little anomalous data is available.
Sounds from different acoustic scenes. Rare sound events.	Mel spectrograms;	Dense autoencoder (AE); Asymmetric CAE	Lo Scudo et al. (2023)	The paper presents three different clustering techniques, and the AE reconstruction error is used for classification. The authors use a dataset of manually induced anomalies, which might make the results hard to replicate when dealing with a small number of anomalous samples collected during actual system operation.

The work presented in Toma et al. (2022) uses WST in an industrial setting to perform AD in the operation of an induction motor, applying the transform to the measured current. The accuracy of the proposed framework, utilising 0th and 1st-order scattering coefficients, exceeds 99%. Moreover, Liu et al. (2020) shows how WST can be applied to classify different heart conditions based on ECG readings. The WST is used as a feature extractor in combination with different classifiers, namely Neural Network (NN), K-Nearest Neighbours (KNN), and a probabilistic neural network (PNN), which results in a minimum

disease classification accuracy of 97.2%, 98.5% and 98.6%, respectively. WST can also be used in ECG-signal analysis to classify sleep apnea (Sharaf, 2023). The cited work shows that, in conjunction with a random forest classifier (RFC), the algorithm has a minimum accuracy of 90.35%

The work presented in Buriro et al. (2021) showcases the application of WST to predict a patient's alcoholism based on EEG readings. The transform was used as a feature extractor for 256-Hz signals and resulted in 100% accuracy in classifying data from twenty healthy subjects and twenty subjects with alcoholism. Another successful implementation of WST in analysing EEG signals is evidenced in Ahmad et al. (2017), where 180 out of 197 cases of seizures have been classified correctly.

3. Method

The analysis of the AD task considered in this work reveals the complexity of the arising classification problem. On the one hand, some anomalies in the operation of FastBlade, such as asynchronous load operation, can result in signals similar to normal operation recordings. Similarly, test parameters can slowly deviate from the desired values, with the sound difference being inaudible to humans. Further, practical requirements for the AD algorithm include its low latency and computational efficiency. This section aims to introduce the most common signal processing tools described in Section 2 and introduce the WST as the optimal time–frequency filter for classification.

3.1. Short time Fourier transform

The STFT relies on using the Fast Fourier Transform (FFT) algorithm applied over a pre-defined window (Kehtarnavaz, 2008), segmenting the time-domain signal to retrieve signal information in the time–frequency domain. The application of the STFT in the continuous domain is given by (Krishnan, 2021):

$$STFT\{x(t)\} = X(\tau, w) = \int_{-\infty}^{\infty} x(t)w(t-\tau)e^{-jw}tdt \tag{1}$$

where $x(t)$ is the analysed signal and $w(n)$ is the chosen analysis window. Therefore, the discrete version of the STFT can be expressed as (Krishnan, 2021):

$$X(n, w) = \sum_{m=-\infty}^{\infty} x(m)w(n-m)e^{-jwm} \tag{2}$$

where:

$$X(n,k) = X(n,w)|_{w=\frac{2\pi}{N}k} \tag{3}$$

$$X(n,k) = \sum_{m=-\infty}^{\infty} x(m)w(n-m)e^{-j\frac{2\pi}{N}km} \tag{4}$$

where $w(n)$ is the STFT window, which is zero outside of the $[0, N-1]$ interval.

3.2. Wavelet transform

As mentioned in Section 2.2, STFT requires a trade-off between time and frequency resolution due to a fixed size of the window. This problem is mitigated by incorporating filters of varying lengths in the WT. The equation for the representation of the CWT is given by (Layer and Tomczyk, 2015):

$$CWT\{x(t)\} = W_f(a,\tau) = \int_0^{\infty} x(t)\Psi_{a,\tau}(t)dt \tag{5}$$

in which $x(t)$ is the analysed signal, a is a scaling factor, τ is a shift factor and $\Psi_{a,\tau}(t)$ is given by:

$$\Psi_{a,\tau}(t) = \frac{1}{\sqrt{a}}\psi\left(\frac{t-\tau}{a}\right) \tag{6}$$

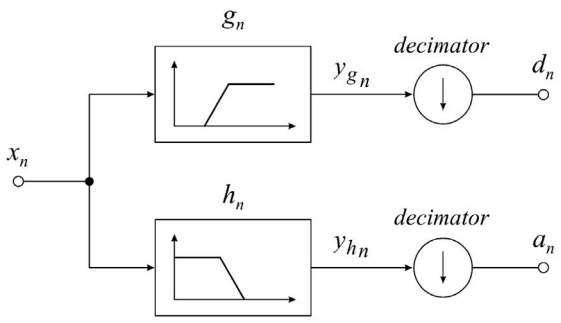


Fig. 2. The visual representation of a DWT computational block. The discrete signal input, x_n, is subject to low-pass and high-pass filtering as given in Eqs. (7) and (8), respectively, followed by decimation. This results in signals h_n and g_n as outlined in Eqs. (9) and (10). The process is cascaded m times by passing subsequent low-pass responses, $a_{(m+1)_n}$ to a_{2_n}, through the same convolution block (Layer and Tomczyk, 2015).

where ψ is the mother wavelet function. The changes in a alter the frequency of the filter, while the sweep of τ moves the kernel across the signal. However, the process is computationally expensive, and a discrete version of the process is preferred for real-time applications. The discrete wavelet transform embodiment is obtained by using finite input response (FIR) filters in a chosen digital filter bank. The outputs of the low-pass filter are given by (Layer and Tomczyk, 2015):

$$y_{h_n} = \sum_{l=0}^{L-1} x_l h_{n-l} \tag{7}$$

while the high-pass filter is represented as:

$$y_{g_n} = \sum_{l=0}^{L-1} x_l g_{n-l}, \quad n = 0, 1, \ldots, L-1. \tag{8}$$

The number of coefficients computed, y_{h_n} and y_{g_n}, is twice the number of samples of the original signal. Therefore, decimation by a factor of two is practised to yield the expressions for low-pass and high-pass filter outputs:

$$a_n = x_{2n} = \sum_{l=0}^{L-1} x_n h_{2n-l} \tag{9}$$

and

$$d_n = x_{2n} = \sum_{l=0}^{L-1} x_n g_{2n-l}. \tag{10}$$

Eqs. (9) and (10) can be used to compute the DWT coefficients for the original discrete signal x_n as presented in Fig. 2.

3.3. Wavelet scattering transform

WST is a tool used for translation-invariant filtering, which is stable to deformations, such as additive perturbations of the initial signal (Bruna and Mallat, 2013; Anden and Mallat, 2014). The wavelet scattering operation is based on the computation of the DWT of the signal, followed by averaging the coefficients over a selected window length. The resolution of the STFT, DWT and WST is schematically shown in Fig. 3.

Fig. 3 reveals the significant differences between the time–frequency analysis tools described previously. The STFT graph confirms that the use of a wide filter window at high frequencies results in poor temporal accuracy and excessive frequency resolution. As a result, STFT lacks stability as the representation of similar signals diverges for high frequencies, which is undesirable for classification problems. The DWT's time and frequency resolution changes according to the parameters of

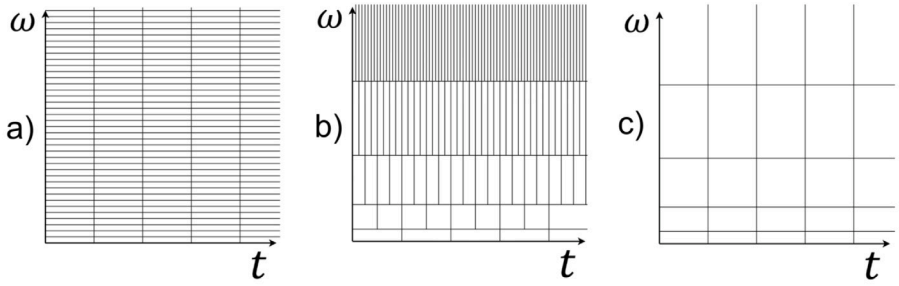


Fig. 3. The visual comparison of the underlying time–frequency resolution for (a) STFT; (b) DWT; (c) WST (Kehtarnavaz, 2008; Mallat, 2012).

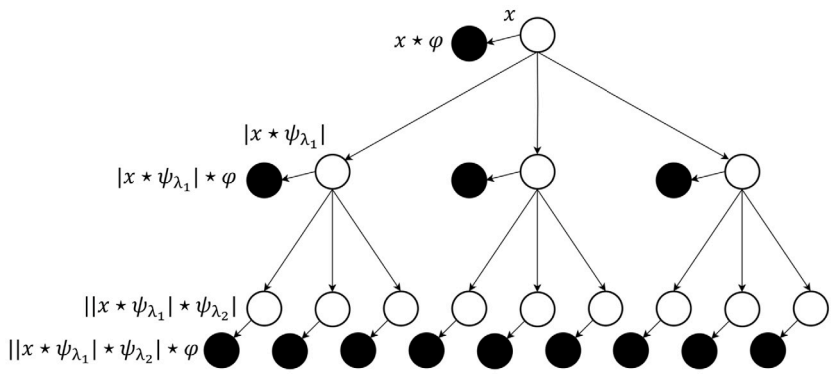


Fig. 4. The structure of a WST network, where x is the input signal, $\star\psi_{\lambda_n}$ represents the WT operation, and $\star\phi$ indicates averaging (Anden and Mallat, 2014). Black dots mark the network output, which is present at each layer.

the examined signal, solving the resolution issue of the STFT. However, due to very high temporal resolution for high frequencies, the output of the DWT is localised in time and, therefore, not translation invariant, which should also be avoided in a classification problem. The resolution representation of the WST is obtained by averaging the DWT coefficients so that the representation of the signal is less time-dependent. Therefore, the WST representation benefits from being both highly stable and invariant under translation.

However, the averaging process described results in a significant loss of information. Therefore, an iterative approach has been described in Bruna and Mallat (2013), where the averaged coefficients constitute the output of the network, while the high-resolution information is recovered by computing the WT again before the signal is averaged. This way, further features are extracted from the signal. Furthermore, the modulus operator is used before the signal is averaged to prevent coefficients of periodic signals from vanishing (Sifre and Mallat, 2013). This way, an invariant scattering convolution network is built, as demonstrated in Fig. 4.

While the process presented in Fig. 4 can be repeated for a growing number of layers, it must be noted that the signal's energy dissipates as it propagates down the network, and the significance of the coefficients decreases. The model's authors also observe that the averaging operator resembles kernels, such as pooling commonly used in CNNs. In contrast, the modulus operator is non-linear, similar to activation functions found in various deep neural networks (Li and Bonner, 2022).

Due to WST's applicability to classification problems and the WST network's documented success in feature extraction problems, the model will be incorporated into the processing pipeline.

3.4. Classification framework

The framework followed in this work is presented using the basic signal processing stages outlined in Fig. 5. The first block represents the high-dimensional data coming from a sensor, which, in the case of this work, is a digital microphone. Using a kernel reduces data dimensionality, which has the benefit of decreased data redundancy and improved computational efficiency. By means of unsupervised learning, data samples can be clustered together based on the similarity of extracted features, simultaneously reducing the variability among a single class (Mallat, 2012). In this work's context, this stage's purpose is to group normal and abnormal operation samples into separate sets, irrespective of the differences between various fault types. Applying supervised learning, which determines a condition for assigning incoming data to specific clusters, results in the final sample classification.

3.5. Autoencoders

Autoencoders are machine learning (ML) models whose distinguishable feature is the ability to learn from unlabelled data. An AE consists of two networks, namely an encoder and a decoder, whose goal is to produce an output identical to its input. The encoder takes an input which can be multi-dimensional and compresses it into a latent code, which is then reconstructed by the decoder. The decoder structure is usually a mirror reflection of the encoder (Lopez Pinaya et al., 2020).

While AEs can consist of solely fully connected layers, they can also incorporate convolutions to increase their efficiency in working with image-like data (Lopez Pinaya et al., 2020). Due to a smaller number of parameters than a conventional AE, Convolutinonal auteoncoders (CAEs) also benefit from a reduced learning time (Chen et al., 2018). A sample structure of a CAE is presented in Fig. 6.

3.5.1. Sample classification based on the reconstruction error

AEs can classify anomalies based on the latent code (Lopez Dubon et al., 2023a) or based on the reconstruction error (Torabi et al., 2023; Givnan et al., 2022). By training the model on normal operation data, it learns the patterns which are predominant in normal operation and minimises the reconstruction error of the output. However, if anomalous data is passed, it is expected to contain features which the model will not be able to reconstruct with equally good accuracy.

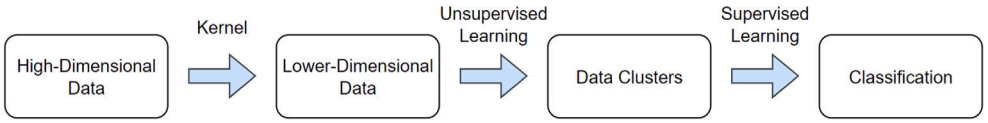


Fig. 5. The steps in the signal processing framework for classification. The high-dimensional data undergoes filtering via a kernel, which feeds into an unsupervised learning model. Supervised learning constitutes the basis for classifying the resulting clusters of data.

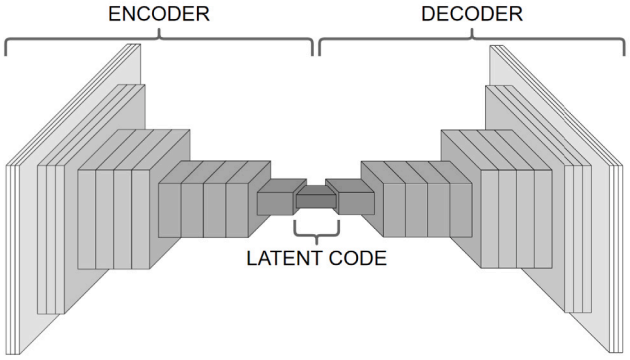


Fig. 6. The structure of a simple CAE, where the shape of the cuboids represents the changing dimensions of data as it propagates through the network. The encoder, latent code and decoder regions of the CAE are marked.

Therefore, each sample can be classified by setting a threshold in the reconstruction error in the following manner (Chen et al., 2018):

$$c_{CAE}(x_i) \rightarrow \begin{cases} \text{normal}, & \epsilon_i < \theta \\ \text{anomalous}, & \epsilon_i \geq \theta \end{cases} \tag{11}$$

where x_i is a sample considered, c_{CAE} denotes the encoding and decoding process of the CAE, ϵ_i is the corresponding reconstruction error and θ is the set threshold. Setting the reconstruction error threshold corresponds to the supervised learning process presented in Fig. 5.

3.5.2. Sample classification based on extracted feature maps

In cases when the sample classification based on the reconstruction error does not bring satisfactory results, e.g. due to significant intraclass variability in the normal dataset, the intermediate feature maps of the CAE can be investigated. Since the CAE is trained to compress the input image during encoding, the encoder will be equipped with filters that extract the most relevant features of the input signal. The investigation of these feature maps should identify particular signal properties that are common to all normal samples and cannot be observed among anomalous samples. Therefore, the feature maps present at a given layer of an encoder can be averaged for all training samples, creating a dataset of generic features typical of all normal operation samples. Similarity factors can be computed between these features and both, normal and anomalous data. When sufficient separation between the similarity values is reached through a combination of a particular feature map and the type of computed error, the framework can be used to classify new samples.

3.6. Processing pipeline

In this work, we first verify if a CAE and a PCA transform trained on normal operation data can successfully classify abnormal operation samples. We subsequently present the solution which utilises chosen feature maps from the deep layers of the CAE. The associated processing pipeline consists of three stages: dataset creation; training; and testing. These are presented in Figs. 7–9 respectively.

4. Experimental results

4.1. Instrumentation and data collection

The data was collected during a fatigue test of a composite tidal turbine blade. The blade is 5.25 m long and weighs 1588.59 kg. Its natural frequency is approximately 18 Hz, and the NACA 63-4XX aerofoil defines its cross-section (McLoughlin et al., 2023). During the test, the blade was actuated by a series of three actuators, and the exerted loads followed a sinusoidal waveform. The loading frequency was 0.75 Hz, while the actuators' minimum and maximum loads were 20 kN and 60 kN, respectively.

The audio was recorded for a total of 5 h 30 min. The audio capture device used was a micro-electromechanical system (MEMS) microphone, IMP34DT05, by STMicroelectronics (STMicroelectronics, 2021). It was purchased on a coupon board (STEVAL-MIC003V1) and connected through a development board (X-NUCLEO-CCA02M1) to a laptop to log audio using Audacity software (Audacity, 2021). It ran in the single-channel mode, and its operating frequency was 48.0 kHz. The microphone was mounted in the test hall approximately five metres away from the reaction frame, with the specimen and the pumping machinery (electric motors, pumps and oil hoses) located in the pit below. The experimental setup is presented in Fig. 10. The dataset used to train, validate, and test the model developed in this work can be downloaded from https://doi.org/10.5281/zenodo.14298279 Munko et al. (2024).

4.2. Dataset parameters

The audio data was recorded and stored at a resolution of 48.0 kHz. The abnormal operation periods were first determined by analysing the process parameters recorded during the test, including actuator load, pump pressure and motor speed. The example of a fault detected in the operation of the loading actuators is presented in Fig. 11. The figure shows the case of the system loads acting out of phase, which is the immediate cause of the increased load amplitude. The uncontrolled behaviour of the system might not only affect the reliability of the test but might also lead to the system losing stability.

Equally, the period when the system is winding up and down simulates power loss, which would be abnormal if it occurred during a test. Any third-party interactions resulting from, e.g. periodic noise in the hall made by the operation of an elevating work platform, are also classified as undesirable during unmanned operation. Once an abnormal operation was identified, the corresponding audio track was extracted and labelled as anomalous.

All data in the dataset was divided into 0.5-second-long samples, each overlapping by 0.1 s with the previous sample. The interval length was chosen based on the duration of a transient fault identified in the asynchronous operation of the system loads. However, we also considered the duration of a single sample to be long enough not to capture solely normal behaviour in the case of periodic noises. It was decided that the samples should overlap to avoid omitting any unusual behaviour occurring at the sample's boundary and to avoid any potential information loss due to the cone of influence (CoI) effect (Chen et al., 2023). The list of identified anomalies and the resulting number of samples obtained for each one of them is presented in Table 2.

The microphone has a 16-bit output, which, for signed integers, can be decoded to into the effective range of values between -2^{15} and 2^{15}.

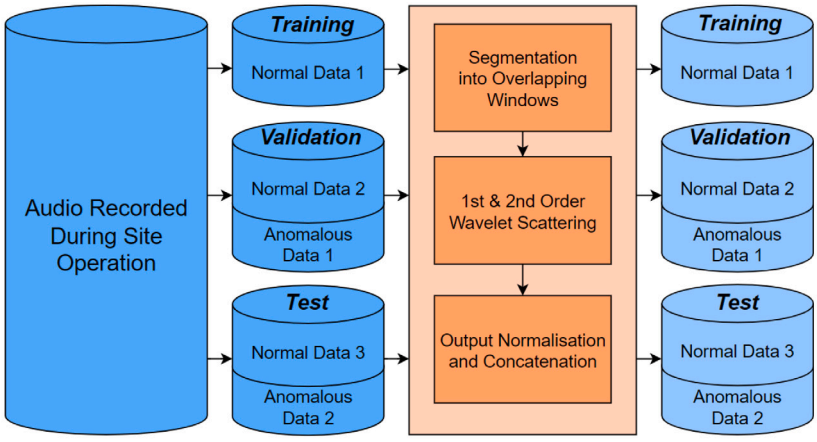


Fig. 7. The dataset creation stage.

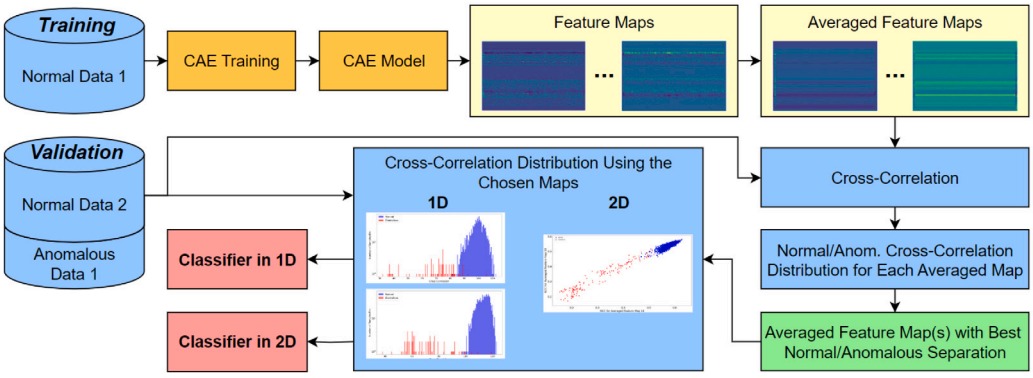


Fig. 8. The training pipeline followed in the work.

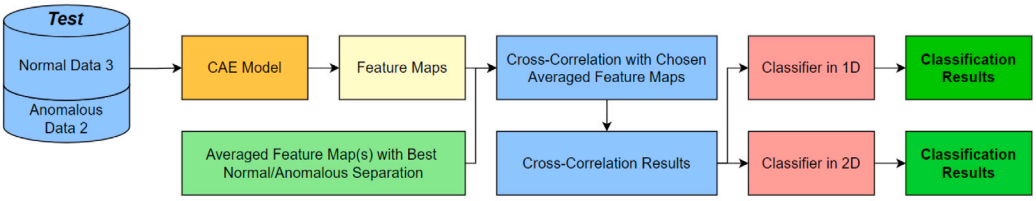


Fig. 9. The testing pipeline followed in the work.

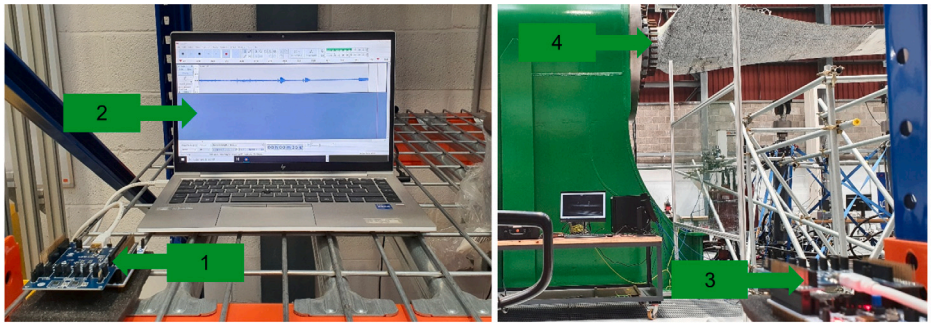


Fig. 10. Left: The connection between the microphone and the logging laptop in the FastBlade test hall. Right: The sensing setup's location relative to the facility's reaction frame. 1, 3: Microphone on the development board; 2: Laptop running Audacity (Audacity, 2021); 4: The specimen mounted on the reaction frame.

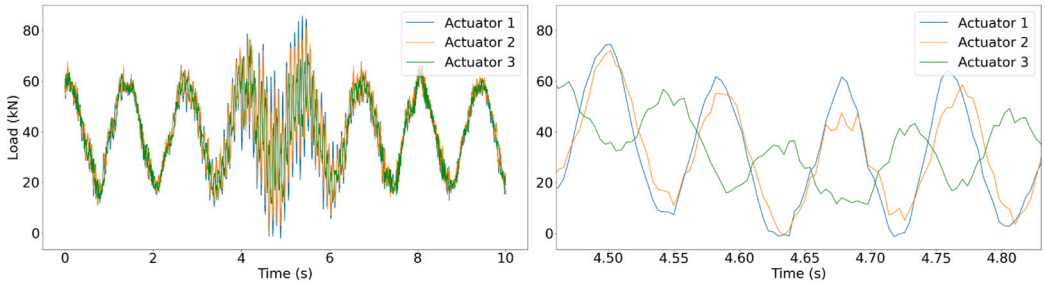


Fig. 11. Left: The load traces for each actuator used in the test. The transient anomaly occurring in the test is characterised by the increased frequency and loading amplitude relative to normal system operation. Right: The magnification of loading traces for a shorter time interval. The loss of synchronism becomes apparent on the new time scale, with the third actuator acting out of phase.

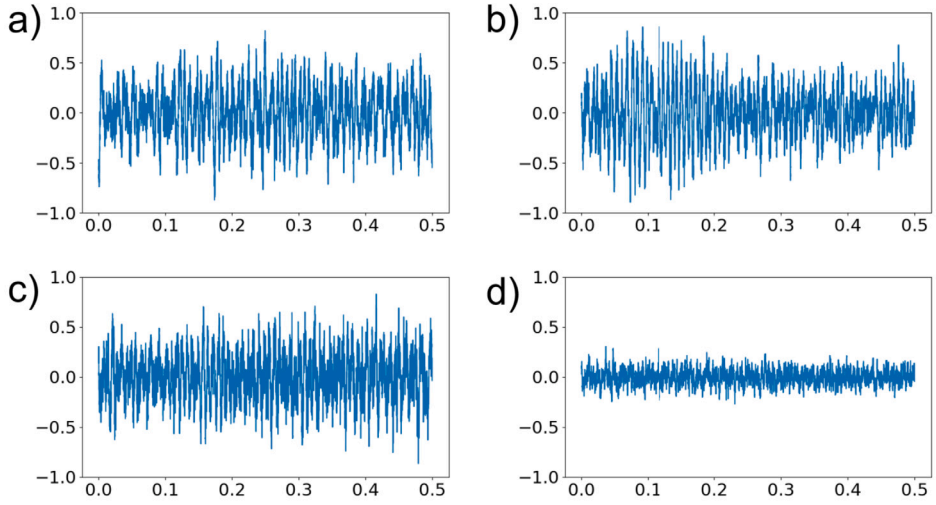


Fig. 12. Normalised audio data of the facility operation recorded for (a) normal operation, (b) random noise coming from a 3rd party interaction, (c) asynchronous operation of the actuators, (d) system operating at reduced power (during the start-up process).

Table 2
Anomalous and normal operation samples constituting the dataset used in the research.

Type	Duration (s)	No. of samples
Asynchronous actuator operation 1	0.95	2
Asynchronous actuator operation 2	1.56	3
System winding down	26.92	67
System winding up	5.55	13
Random noise 1	18.41	45
Random noise 2	7.74	19
Random noise 3	1.29	2
Periodic noise	3.58	8
Combined anomalous dataset 1	40.00	80
Combined anomalous dataset 2	39.50	79
Normal operation dataset 1	540.00	1347
Normal operation dataset 2	540.00	1347
Normal operation dataset 3	540.00	1347

Therefore, the raw microphone data was divided by 2^{15}, which results in a normalised waveform between −1 and 1, a common representation for audio signals. No further pre-processing was performed on the signal before applying the time–frequency analysis tools. This way, reliance on the inherent noise-reduction properties of the algorithms ensures that no meaningful signal features are lost. The signal was first examined in the time domain. Selected audio samples are presented in Fig. 12. The investigation of time-domain signals highlights the need for a specific feature extraction technique to differentiate between most normal and anomalous samples.

4.3. Signal processing

In this section, the time–frequency analysis tools introduced in Section 3, namely the WST, CWT, DWT, and STFT, will be evaluated on the collected samples. The algorithms are applied to two samples collected during three distinct periods of normal operation, and two anomalous samples. The comparison of all extracted features is presented in Fig. 13.

The values in each signal presented in Fig. 13 are normalised between zero and one. Since the performance of different feature extraction methods greatly depends on the particular choice of their tunable parameters, it is difficult to compare them directly. Therefore, the parameters of the presented images are chosen to be as similar to one another as possible. Where possible, the frequency range covered extends to 24,000 Hz, which is determined by the Nyquist frequency (McLean et al., 2005) as the maximum examinable frequency for a signal with a sampling rate of 48,000 Hz.

The WST is computed using the *Kymatio* Python library (Andreux et al., 2018). The J and Q parameters, corresponding to the maximum log scale of the transform (equal to 2^J) and the number of wavelets per octave respectively, need to be determined. The use of J and Q parameters, such that $J, Q \in \{6, 8, 10, 12, 14, 16\}$, is examined. The number of wavelets per octave for the 2nd-order scattering is fixed at $Q = 1$. For a sample consisting of 24,000 data points, the choice of $J = 6$ and $Q = 16$ results in the most significant number of output parameters. These J and Q values are used in the research to maximise the amount of information retrieved from the original signal. The Kyamtio's default wavelet is used, which is the Morlet wavelet.

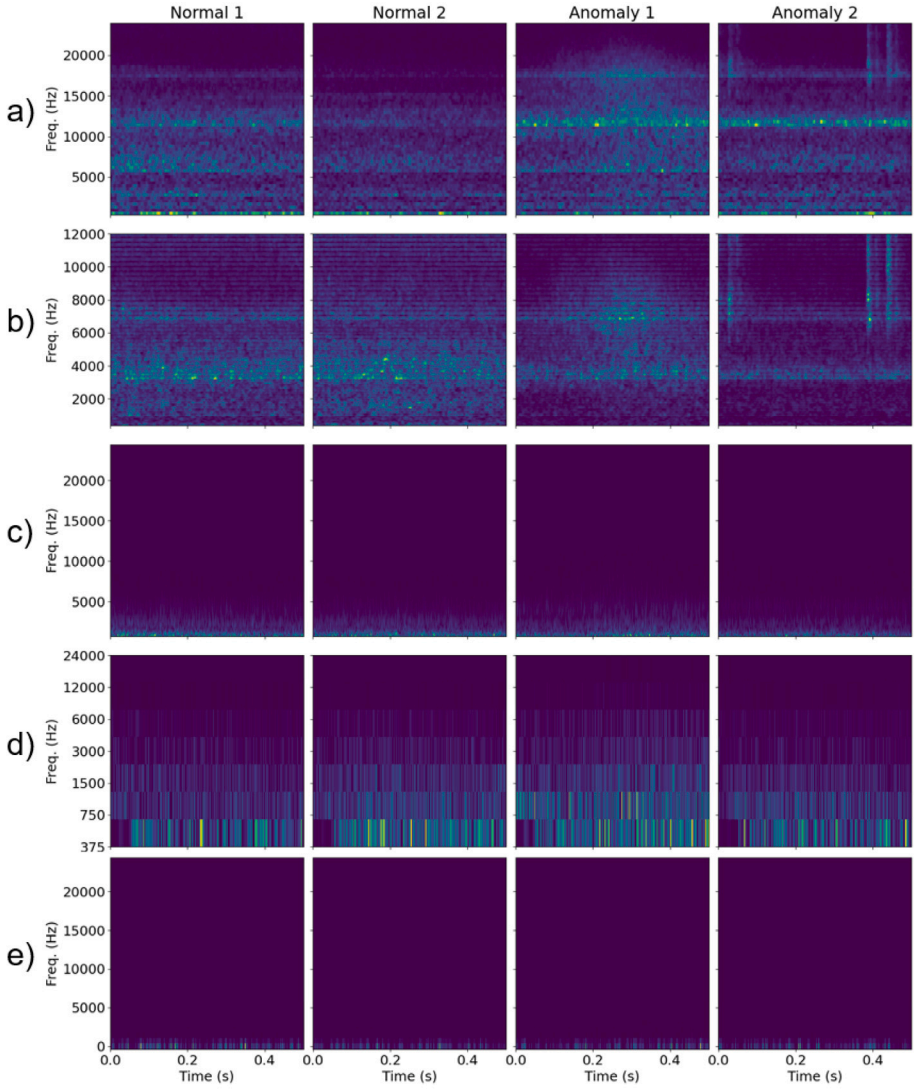


Fig. 13. The comparison of the time–frequency analysis tools discussed in the work, applied to two normal and two anomalous samples collected at FastBlade. The outputs are computed with: (a) 1st-Order WST; (b) 2nd-Order WST; (c) CWT; (d) DWT; (e) STFT. To allow for direct comparison, the transforms analyse the signal in the range of up to 24,000 Hz, which is half of the sampling frequency. The only exception is the 2nd-order WST, which investigates lower frequencies than the 1st-order transform. Each of the y-axis labels of the DWT graph corresponds to a discrete frequency band investigated.

Therefore, the CWT also utilises the Morlet wavelet, and frequencies down to 750 Hz are examined, which is derived from the signal's sampling frequency, 48,000 Hz, and the scale of the transform equal to 2^6 (i.e., 2^J). The filter chosen from the DWT's filter bank is *db20*, one of the Daubechies family's wavelets, matching the Morlet wavelet's characteristics. To match the examined frequency range with the STFT, the length of each segment required to compute the STFT is set to 64, since $\frac{48{,}000\ \text{Hz}}{64} = 750$ Hz, which is the lowest examined frequency.

The comparison of the outputs of the time–frequency tools presented in Fig. 13 suggests that the differences between normal and anomalous samples are best observed in the WST outputs. The area with increased energy distribution in the middle of *Anomaly 1*, and the sharp vertical edges visible in *Anomaly 2* make the samples different from the normal operation data. However, the normal samples also differ. Relative to *Normal 1*, the energy distribution for *Normal 2* is dimmer in the 1st-order scattering output, while there is also more activity across all frequencies in the 2nd-order output. While the DWT output for *Anomaly 1* might suggest some non-conforming patterns, no change is visible for *Anomaly 2*. With the chosen parameter settings, the outputs for the CWT and the STFT show no significant patterns. Further investigation leads to discovering patterns in the CWT and STFT outputs as presented in Fig. 14, derived through radical capping of maximum values displayed to 0.01. However, changing the range of values displayed requires prior knowledge of the signal and increases noise levels, and is therefore not advisable.

Another aspect of great importance for our case study is the processing time associated with the techniques, as the application is expected to run in real-time. Therefore, the execution time of each of the techniques is benchmarked by averaging it over 1000 computations. The results presented in Table 3 are obtained on 11th Gen Intel(R) Core(TM) i7-1185G7, with four physical cores, unless otherwise stated.

The benchmarking results show that the WST is the most computationally expensive of the processing tools considered. The discrepancy between the 1st-order, 2nd-order, and concatenated outputs is small and likely results from random variations observable on the CPU. However, the WST operation can also be easily parallelised, and the Kymatio library used provides GPU support. Therefore, by benchmarking the WST on the GPU available at FastBlade, on which the developed application is to be deployed, the execution time decreased drastically, mitigating the impact of the higher computational burden. Therefore, the WST output can be used in further study. Moreover, the 1st-order and 2nd-order outputs are concatenated to ensure that all necessary

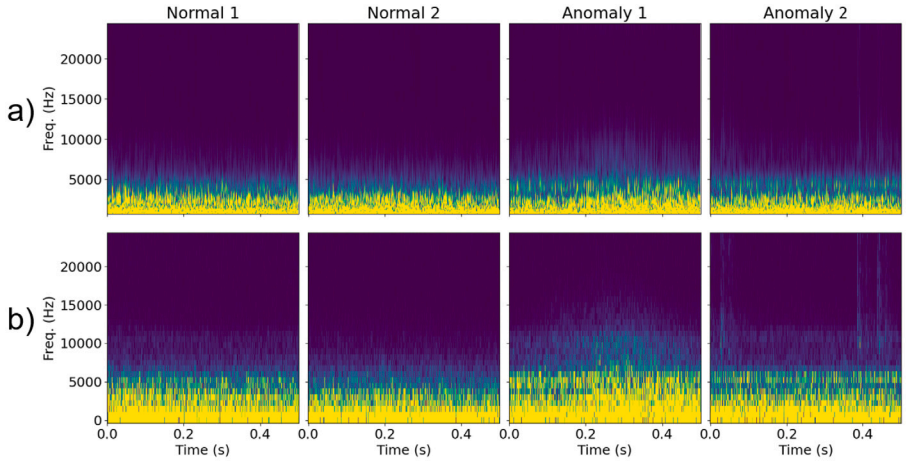


Fig. 14. The comparison of the time–frequency plots for (a) CWT; (b) STFT, with the effective value range set between 0.00 and 0.01.

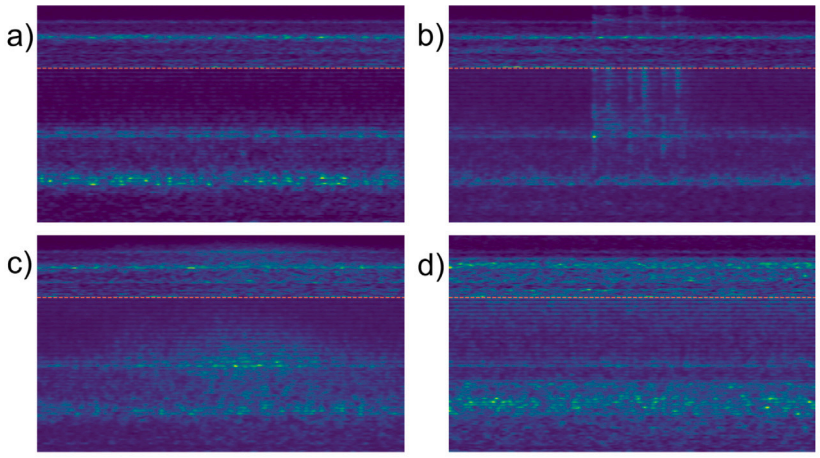


Fig. 15. Concatenated and normalised output for 1st and 2nd-order wavelet scattering; (a) normal operation, (b) random noise coming from a 3rd party interaction, (c) asynchronous operation of the actuators, (d) system operating at reduced power (during the start-up process). The red lines mark the edge where 1st and 2nd-order scattering coefficients are concatenated.

Table 3
Execution time for the feature extraction tools considered.

Processing tool	Execution time (s)
1st Order WST	0.5780
2nd Order WST	0.6654
Concatenated 1st and 2nd Order WST	0.6214
	0.0902[a]
CWT	0.3492
DWT	0.0038
STFT	0.0006

[a] Run on the graphics processing unit (GPU), model: NVIDIA RTX A6000.

Table 4
List of layers in the autoencoder trained. The number of downsampling layers matches the number of upsampling layers. The input and output layers adjust the shape of the image.

Layer/combination of layers	Quantity
ZeroPadding2D	1
Conv2D, MaxPoooling2D	5
Conv2D	1
Flatten, Reshape	1
Conv2DTranspose, UpSampling2D	5
Conv2DTranspose	1
Cropping2D	1

features are extracted from the audio signal. The 1st and 2nd-order outputs are first normalised, and then concatenated as presented in Fig. 15.

4.4. Classification based on reconstruction error

4.4.1. Convolutional autoencoder implementation

A CAE is used to learn the features of normal operation data and subsequently classify samples as either normal or anomalous. The network architecture is presented in Table 4. The *ZeroPadding2D* and *Cropping2D* layers are used to ensure that the dimensions of the input and output of the model match. The number of layers of the encoder network matches the number of layers of the decoder. The latent code of the network consists of twenty-four dimensions. The detailed structure of the model is presented in Table 7 in Appendix.

The model is trained based on the mean absolute error (MAE) metric, which computes and averages the absolute difference between all corresponding pixels in the input and output images. To enhance the understanding of the model's operation, chosen feature maps are extracted for four representative samples, and presented in Fig. 16. While the outputs extracted from the first convolutional layer present features easily identifiable by a human, such as sharp horizontal and vertical lines, the features in the outputs of the third convolutional layer become harder to interpret. The presentation of the latent code shows

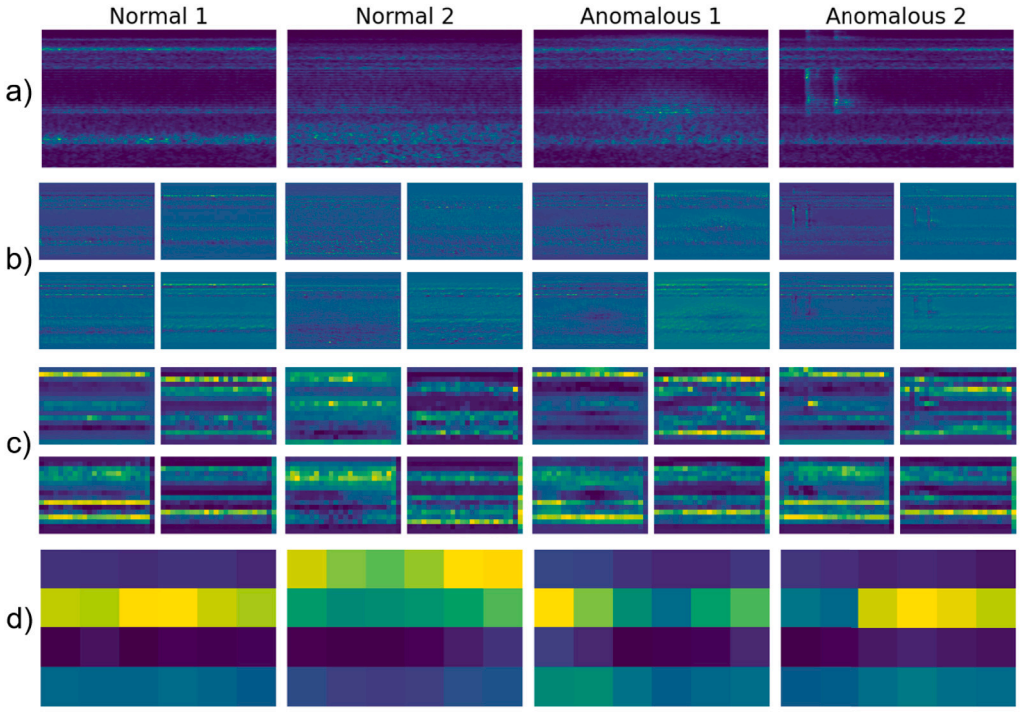


Fig. 16. Four CAE inputs representing two normal and two anomalous samples, together with their feature maps extracted from the model. (a) Original inputs; (b) Feature maps output by the first convolutional layer; (c) Feature maps output by the third convolutional layer; (d) Signal representation in the latent code.

the extent to which the encoder compresses the signal by learning to represent the most important features in a greatly reduced dimensional space.

4.4.2. Principle component analysis implementation

To compare the performance of a deep-learning technique with the implementation of a conventional dimensionality-reduction method, alongside the training of the CAE, PCA is used to compress and reconstruct the WST data. PCA is an operation used to reduce the dimensionality of the original data while maintaining much of its original variance. A reverse operation also exists in which the low-dimensional signal can be used to reconstruct the approximation of the original signal (Vaish and Kumar, 2014). In this work, the PCA is obtained with twenty-four principal components, matching the number of dimensions in the CAE's latent code, and with 1347 dimensions, which matches the number of samples in the normal training dataset, resulting in the transform preserving a significant amount of the original variance. This method also has a much shorter training time than the CAE.

4.4.3. Achieved accuracy

The model is validated on both normal and anomalous data. The normal operation dataset is distinct from the dataset on which the model was trained, and all anomalous samples used in this investigation are used to validate the model's performance. The reconstructions using both PCA transforms, as well as the CAE, are presented in Fig. 17 for the previously shown normal and anomalous samples.

The success of the reconstruction method can be investigated by evaluating the level of separation attained between the reconstruction error computed for normal and anomalous datasets. Four of the commonly used error types are used to compute the error between the input and the reconstructed samples, namely: mean absolute error, mean squared error (MSE), peak signal-to-noise ratio (PSNR), and normalised cross-correlation (NCC). They are given by the following set of equations, where x_i is the input sample, $C(x_i)$ is the reconstructed sample, and n is the number of pixels in an image:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|x_i - C(x_i)| \tag{12}$$

as found in (Hodson, 2022);

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(x_i - C(x_i))^2 \tag{13}$$

as found in (Hodson et al., 2021);

$$PSNR = 10 \cdot \log\left(\frac{MAX_i^2}{MSE}\right) \tag{14}$$

where MAX_i is the signal's maximum value (Horé and Ziou, 2013); and

$$NCC = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(C(x_i) - \bar{C})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(C(x_i) - \bar{C})^2}} \tag{15}$$

where $\bar{x}$ and $\bar{C}$ are the mean values of the original and reconstructed signal (Yoo and Han, 2009).

The distribution of the reconstruction error for normal and anomalous samples is presented as histograms, shown in Figs. 18–20 for the 24-dimensional PCA, 1347-dimensional PCA, and the CAE respectively. Each of the plotted datasets is distributed over 200 bins. The visual inspection of the figures makes it evident that the separation between normal and anomalous data is insufficient to allow for effective classification, as the two distributions overlap significantly for each of the computed error types. This can be explained either by the fact that each model can also reconstruct anomalous samples with high fidelity or by the intraclass variations among normal operation samples, as noted previously. The investigation of the NCC results for all three cases shows interesting behaviour of the distribution for normal samples, where two distinct peaks are visible, suggesting that there are indeed samples within the normal dataset whose reconstruction is much closer to the original images relative to the anomalous samples. Therefore, the latter of the given explanations is more likely to be true; namely that the complexity of the patterns in some of the normal operation samples makes the models fail to reconstruct them, resulting in increased reconstruction error. This can be further supported by the example given in Fig. 17(b), where none of the models reconstructs the normal operation sample with sufficient detail. Therefore, this part of the study can be

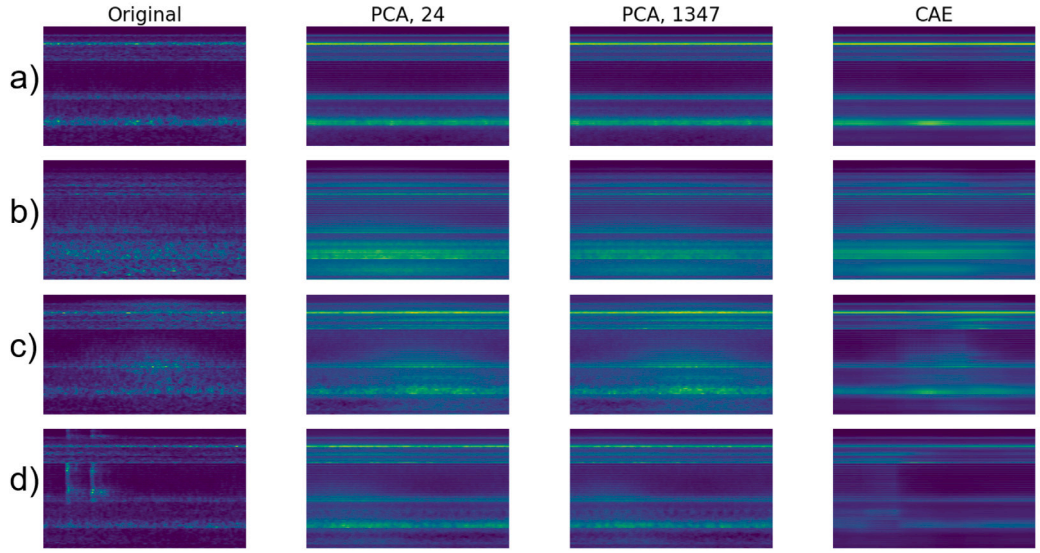


Fig. 17. The visualisation of the original input data and the associated reconstruction using two PCA transforms and a CAE for (a) Normal Operation Sample 1; (b) Normal Operation Sample 2; (c) Anomalous Operation Sample 1; (d) Anomalous Operation Sample 2.

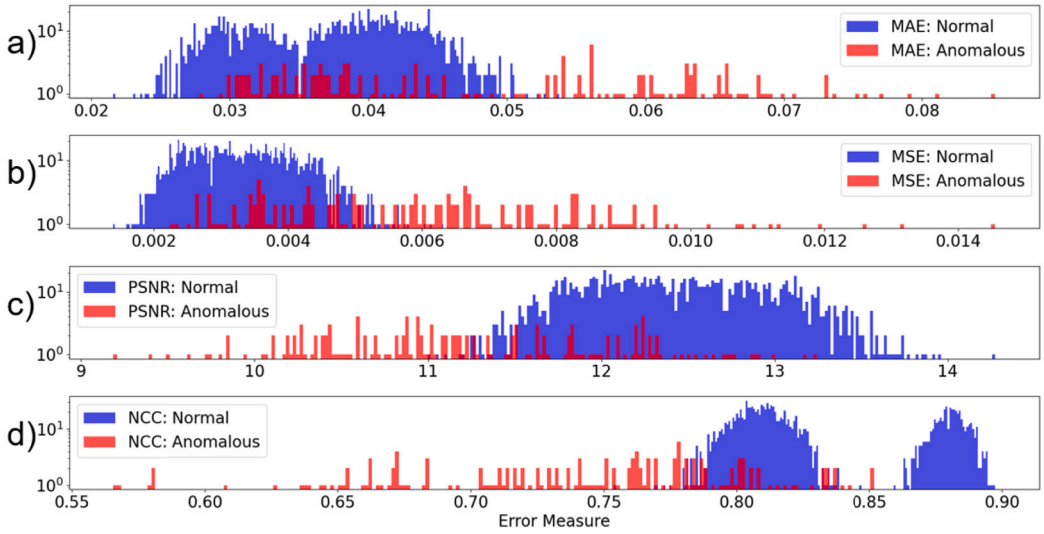


Fig. 18. The error distribution for normal and anomalous samples reconstructed with 24-dimensional PCA. Presented error types are: (a) MAE; (b) MSE; (c) PSNR; (d) NCC. The y-axis is in logarithmic scale to aid visibility.

concluded by stating that the methods based on the reconstruction error have failed to provide an AD method with satisfactory performance. Possible improvements to the method will be discussed in Section 5.

4.5. Feature map-based anomaly detection

4.5.1. Anomaly detection model development

Since the method based on the reconstruction of the entire sample computed with the WST did not result in a satisfactory separation between normal and anomalous data, the feature maps of the CAE's hidden layers are investigated. The rationale for this investigation is the search for features that are common to all normal operation samples but cannot be observed in the case of anomalous operation. The hidden layers of the CAE's encoder extract features representative of the input signal, even if they cannot be interpreted by a human. The examples of the outputs of the first and third convolutional layers are presented in Fig. 16(b) and (c) respectively. Since one convolutional layer might not be sufficient to extract the underlying patterns in the data, we investigate the feature maps which are the output of the second convolutional layer. The second layer has 64 feature maps, which is twice as many as the third convolutional layer.

We compute all 64 feature maps for each of the samples in the training dataset. Subsequently, looking for generic features common to all normal operation data, we compute the average for each feature map across all samples. As a result, we obtain 64 averaged feature maps, which we expect to contain universal features.

To verify if the features derived are indeed common to normal operation samples, and do not exist in the anomalous data simultaneously, we compute the MSE, PSNR, and NCC errors between the averaged maps and the validation dataset. The validation dataset consists of another batch of normal operation data and half of the anomalous samples, with the other half used for testing the model. As in the method applied to the PCA transform and the CAE, we use histograms to determine the separation between normal and anomalous data distributions. To determine how well the data is separated for each feature map and each error metric, we quantify the overlap between the histograms by counting the number of samples common to both distributions. The results of this process are presented in Fig. 21, where

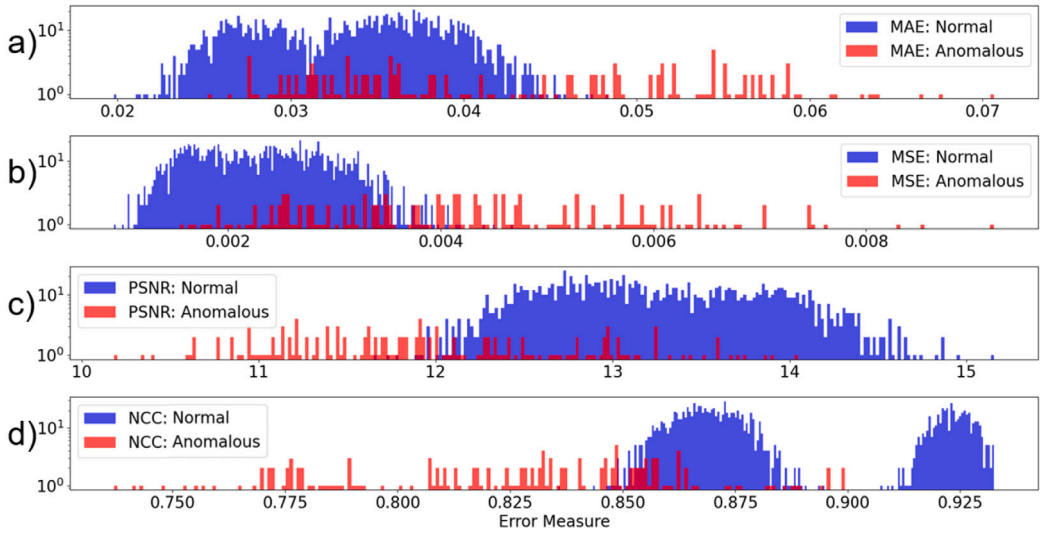


Fig. 19. The error distribution for normal and anomalous samples reconstructed with 1347-dimensional PCA. Presented error types are: (a) MAE; (b) MSE; (c) PSNR; (d) NCC. The y-axis is in logarithmic scale to aid visibility.

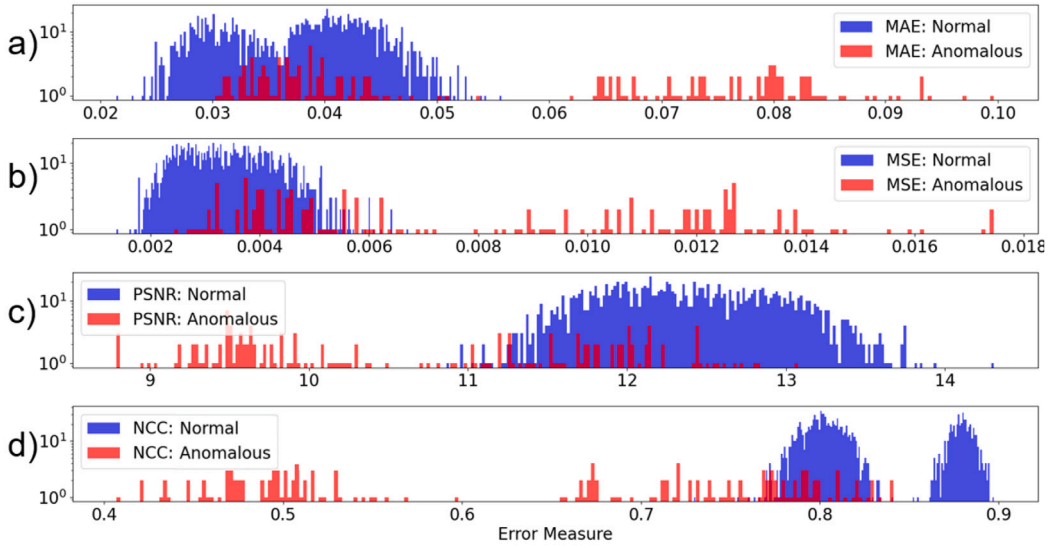


Fig. 20. The error distribution for normal and anomalous samples reconstructed with the CAE. Presented error types are: (a) MAE; (b) MSE; (c) PSNR; (d) NCC. The y-axis is in logarithmic scale to aid visibility.

each of the 64 feature maps has three associated overlap scores. The smallest overlap, and therefore the greatest separation between the normal and anomalous samples, is obtained when the averaged feature maps with IDs 18 and 19 are used together with the NCC.

The associated histograms for which the smallest overlap scores are derived are presented in Figs. 22 and 23 for averaged feature maps number 18 and 19 respectively. The separation visible in the graphs is better than the one presented in Figs. 18, 19, or 20 computed using the error for complete reconstructions. The overlap between the distributions will be further discussed in Section 5, where the reasons for these occurrences will be analysed.

The averaged feature maps discussed are plotted in Fig. 24, alongside the corresponding feature maps obtained when the CAE inference is run on two normal samples, characterised by significant intraclass variations, and an anomalous sample. The average feature maps, presented in the third column from the left, clearly show that they extract horizontal patterns from the data. According to our findings, these patterns have a higher NCC score for the two normal samples presented than for the anomalous sample, which results in a poorer correlation. Substantial noise can be observed in the anomalous sample, which occurs to a lesser extent in the normal operation samples.

Having achieved a seemingly satisfactory separation between normal and anomalous data, the same dataset is used to train classifiers, whose performance will be validated on the test dataset. Numerous classifiers are trained, and their performance will be evaluated using relevant metrics. The chosen models, namely k-nearest neighbours (KNN), logistic regression (LR), support vector machine (SVM), and decision tree (DT), are commonly applied in classification tasks. Two KNN variants are trained, with k equal to three and five, and two SVM variants are also trained, one with a linear kernel and the other with a radial basis function (RBF) kernel. They are implemented using the *Scikit-Learn* library in Python (Pedregosa et al., 2011) and trained on:

- one-dimensional data computed for the averaged feature map with ID 18 (resulting in better separation than the feature map with ID 19);
- two-dimensional data computed for the averaged feature maps with IDs 18 and 19.

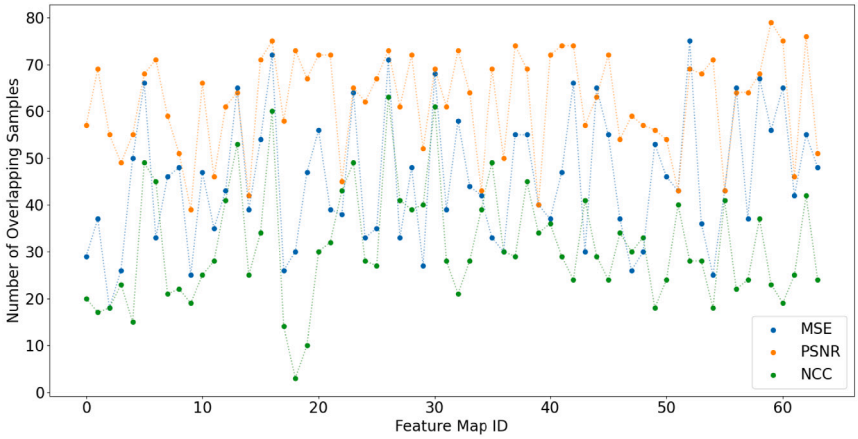


Fig. 21. The number of overlapping samples between the normal and anomalous distributions computed for each of the 64 averaged feature maps derived from the CAE. The error metrics used are MSE, PSNR, and NCC.

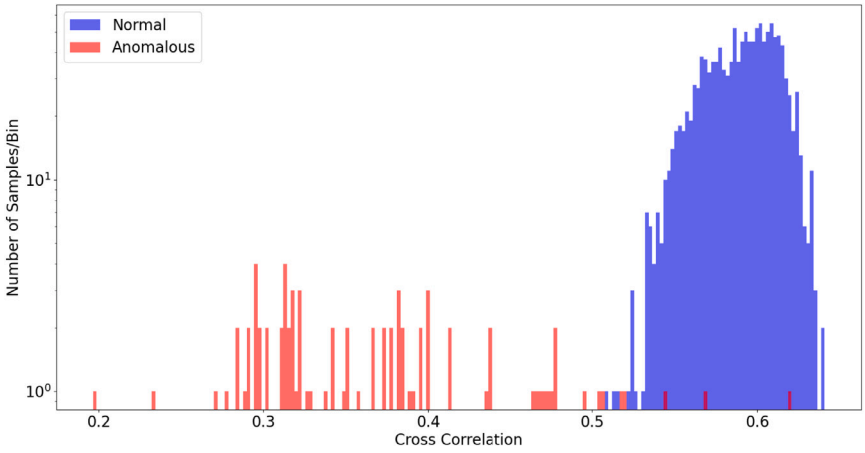


Fig. 22. The distribution of the NCC values computed between the averaged feature map with ID 18 and the samples in the validation dataset. Both distributions are plotted for 200 bins.

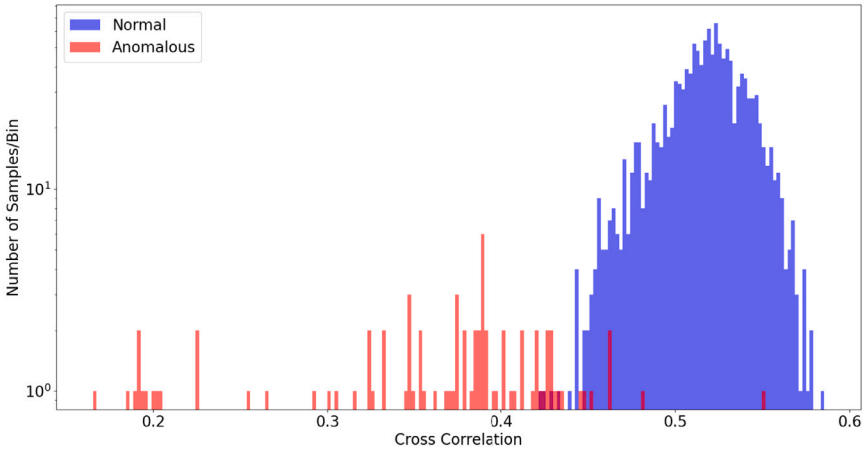


Fig. 23. The distribution of the NCC values computed between the averaged feature map with ID 19 and samples in the validation dataset. Both distributions are plotted for 200 bins.

4.5.2. Testing the anomaly detection model

The test dataset used to evaluate the performance of the developed classifier consists of another batch of normal operation data and the other half of the anomalous samples considered in this work. The error distributions computed for the one-dimensional and the two-dimensional cases are presented in Figs. 25 and 26 respectively. In both cases, the test data distribution results in a slightly lower NCC score than the data on which the classifiers were trained. However,

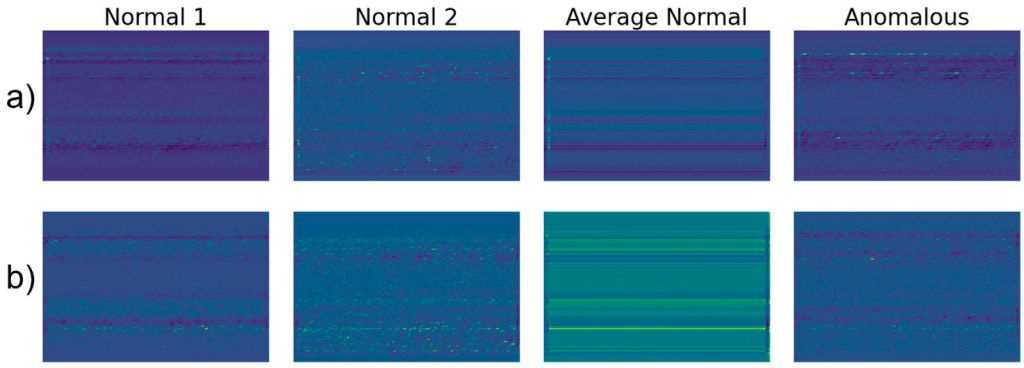


Fig. 24. Feature maps for two normal operation samples, averaged feature maps, and feature maps for an anomalous sample. Feature map ID is (a) 18; (b) 19.

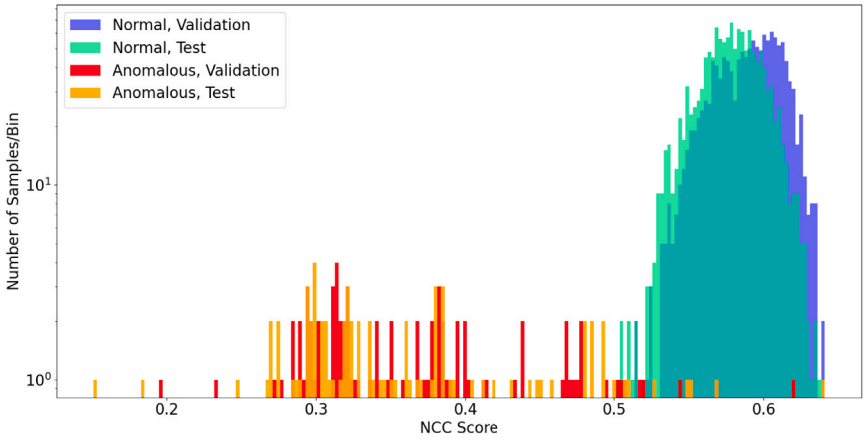


Fig. 25. The distribution of NCC values for validation and test data, when the feature map with ID 18 is considered.

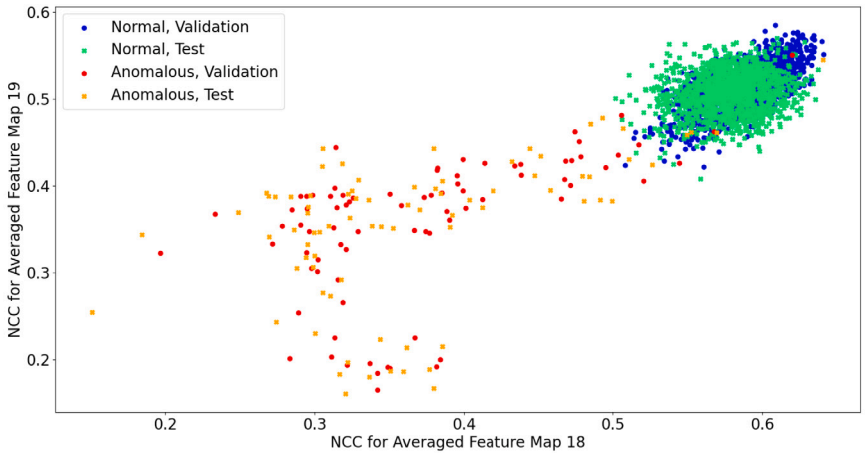


Fig. 26. The distribution of NCC values for validation and test data, when the feature maps with IDs 18 and 19 are considered.

the separation between normal and anomalous data is certainly still evident.

The performance of each classifier is quantified using overall accuracy, which is the ratio of correct predictions to the total number of predictions, as well as:

- precision, given by $\frac{TP}{TP+FP}$, where TP and FP are the number of true and false positives respectively;
- recall, given by $\frac{TP}{TP+FN}$, where FN is the number of false negatives;
- 1 score, which is evaluated as $2 \times \frac{precision \times recall}{precision+recall}$.

While the classifier's accuracy provides a single-value measure of the model's performance, examining the other scores is particularly important when the number of samples in the classes is unbalanced. The results of running the classifiers on the test dataset are presented in Tables 5 and 6 for one-dimensional and two-dimensional data, respectively.

It is discovered that the trained classifiers achieve similar performance for both one-dimensional and two-dimensional data. The accuracy gain when moving into two-dimensional data is either moderate or as in the case of both SVM classifiers, no gain is observed. While the accuracy of all presented classifiers might suggest their suitability

Table 5
The evaluation of the classifiers for one-dimensional data. In all cases, the number of samples, known as support, is 1347 for normal samples, and 79 for anomalous samples.

Classifier	Accuracy	Class	Precision	Recall	F1 Score
KNN, k = 3	0.9923	Normal	1.00	1.00	1.00
		Anomalous	0.94	0.92	0.93
KNN, k = 5	0.9944	Normal	0.99	1.00	1.00
		Anomalous	0.99	0.91	0.95
LR	0.9516	Normal	0.95	1.00	0.98
		Anomalous	1.00	0.13	0.22
SVM, linear	0.9867	Normal	0.99	1.00	0.99
		Anomalous	1.00	0.76	0.86
SVM, RBF	0.9951	Normal	0.99	1.00	1.00
		Anomalous	1.00	0.91	0.95
DT	0.9867	Normal	1.00	0.99	0.99
		Anomalous	0.85	0.92	0.88

for the application, it is impacted by the fact that normal samples greatly outnumber the anomalous ones in the test dataset. Examining the recall and the F1 score shows the performance of the LR is certainly unacceptable due to a high number of false negatives in the anomalous class. The presence of false negatives in the anomalous class means that the model is more likely to classify anomalous samples as normal operation samples, which is unacceptable in the actual operation of the facility. The best-performing classifier for one-dimensional data is the SVM with the RBF kernel, and the best-performing model for two-dimensional data is the KNN with $k = 5$.

The choice between using one-dimensional or two-dimensional data will depend on the benchmarked computational time, as the application is intended to detect anomalies in the system's operation in real-time. The processing required entails running inference on the CAE to produce the necessary feature maps, computing its NCC with the averaged feature map (two feature maps in the case of the two-dimensional data), and running the prediction on the classier. The functionality is benchmarked on the computer available at FastBlade, with the NVIDIA RTX A6000 GPU. The classifier used for benchmarking both cases is the SVM with a linear kernel. The execution time average for 1000 iterations is 0.1041 s and 0.1053 s for the one-dimensional and the two-dimensional case respectively. The computational increase is deemed to be negligible and implies that the main processing burden remains the inference of the CAE. It is therefore decided that two-dimensional data can be used in the final processing pipeline, although it is noted that the performance increase resulting from it is marginal.

4.5.3. Misclassification analysis

In this subsection, we aim to identify the reasons for the misclassification of samples. The learning outcomes can be used to evaluate the quality of the dataset and steer the future direction of research. Figs. 27(a) and 28(a) show the correctly-classified normal and anomalous samples respectively, while samples in (b) in both figures have been misclassified.

The normal operation sample, which is classified incorrectly, certainly has energy distributed across a wider spectrum of frequencies than the other sample presented, which is also visible in the feature map. However, it must be stressed that some of the classifiers have succeeded in assigning the correct class to it. Therefore, this example can be used to stress the significant intraclass variability observed for normal operation samples. Considering the anomalous case, which is one of the few failures of the classifiers, the feature map and the WST output in Fig. 28(a) both exhibit a wider high-energy band in the bottom part of the spectrum, which likely leads to a lower NCC metric when compared to the averaged feature map. On the other hand, the samples in (b), are characterised by a dimmer, yet equally-wide high-energy band in the same portion of the spectrum. This might point to the shortcoming of the method, where a truly anomalous sample was not classified correctly. On the other hand, the anomaly label associated with this sample is *periodic noise*, which means there is a chance that in between anomalous sounds, a normal sample was recorded.

5. Discussion

The study of the related work has shown that microphones are widely used sensors to tackle AD problems in an industrial setting. It has been implemented in this work primarily due to its non-specificity, low cost, and successful implementation in various AD problems which are documented in the studied literature. The use of the microphone has helped successfully detect anomalies of multiple natures, which would be difficult to replicate using a different sensor. While actuators' asynchronous operation could potentially be seen by looking at load or pressure traces, other anomalies occurring in the test hall would probably need to rely on an image-based method. Using cameras would undoubtedly be more costly and computationally demanding.

The theory behind the development of the WST network indicated that it could be a superior classification solution to other commonly used feature extraction tools, like STFT or DWT. Its processing time is significantly reduced due to the possible parallelisation of its computation on the GPU. The execution time is important for a real-time application and is measured to be approximately 0.09 s. Therefore, since the time needed to extract the feature map, compute the NCC, and run the classifier is estimated at 0.11 s, the total time of running inference for a 0.5-second-long sample is approximately 0.2 s. This means the method meets the time limit, set to 0.4 s, as the samples overlap by 0.1 s. However, deploying the method on a CPU would be too time-consuming. This problem could be mitigated by resorting to a more computationally feasible feature extraction tool, even if it results in poorer performance.

In the method developed, NCC can be computed with one or two of the averaged feature maps for every incoming sample. The most important metric presented is the recall for anomalous samples, which determines the number of anomalous samples classified as normal. The best performance is observed when two-dimensional data is used together with a KNN with k set to 5, reaching an accuracy of 99.6% and a recall of 92%. Considering the size of the dataset, this means that 6 out of 79 anomalous samples were wrongly classified as normal. For reference, the SVM with the RBF kernel deployed on one-dimensional data misclassified 7 out of 79 samples. However, since their computational times are very similar, the approach that uses two dimensions can be used.

Considering the classification attempt based solely on the reconstruction error of the input sample, increasing the size of the CAE, could be a solution to its unsatisfactory performance. By having a bigger number of trainable parameters, the model could learn to reconstruct even the more complex patterns of normal operation samples. This approach could be further investigated due to the ease of collecting normal operation data. The major limitations of this method would be the CAE's inference time, which has a hard limit set by the sample duration, and the learning time, which would need to remain feasible given that a test at FastBlade can last a few days.

The greatest limitation of our model is that contrary to classification based on the reconstruction error, it no longer relies solely on unsupervised learning. The anomalous data labelled as *validation* had to be used to determine which feature maps resulted in the greatest separation between normal and anomalous samples. Moreover, they were used to train the classifier, which was then evaluated on the unseen normal and anomalous data labelled as *test*. In light of the scarcity of anomalous samples, data augmentation techniques can be used to create them artificially.

Table 6
The evaluation of the classifiers for two-dimensional data. In all cases, the number of samples, known as support, is 1347 for normal samples, and 79 for anomalous samples.

Classifier	Accuracy	Class	Precision	Recall	F1 Score
KNN, k = 3	0.9951	Normal	1.00	1.00	1.00
		Anomalous	0.99	0.92	0.95
KNN, k = 5	0.9958	Normal	1.00	1.00	1.00
		Anomalous	1.00	0.92	0.96
LR	0.9691	Normal	0.97	1.00	0.98
		Anomalous	1.00	0.44	0.61
SVM, linear	0.9867	Normal	0.99	1.00	0.99
		Anomalous	1.00	0.76	0.86
SVM, RBF	0.9951	Normal	0.99	1.00	1.00
		Anomalous	1.00	0.91	0.95
DT	0.9895	Normal	1.00	0.99	0.99
		Anomalous	0.89	0.92	0.91

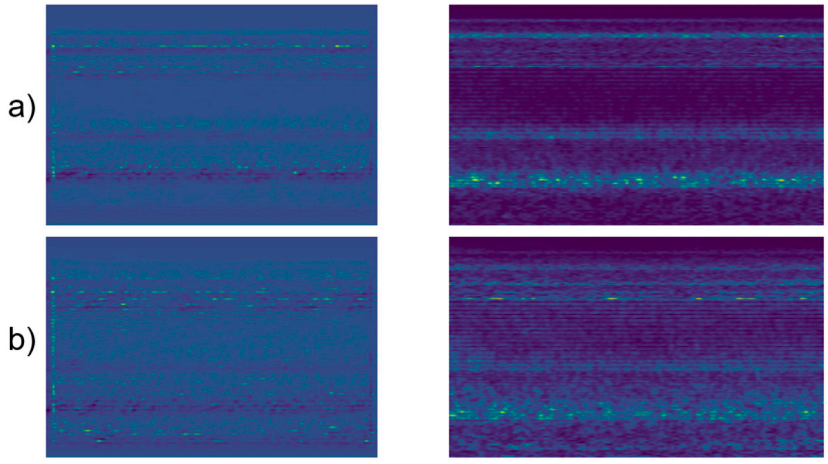


Fig. 27. (a) Correctly classified feature map (left) and the original WST output (right) for a normal operation sample; (b) Incorrectly classified feature map (left) and the original WST output (right) for a normal operation sample.

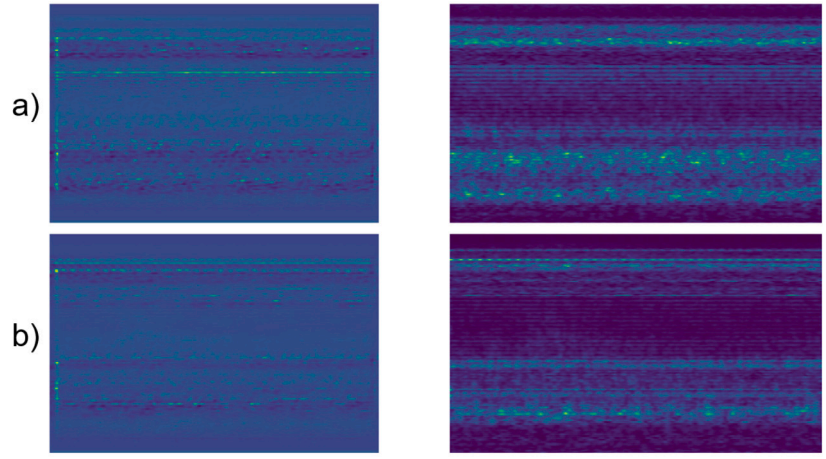


Fig. 28. (a) Correctly classified feature map (left) and the original WST output (right) for an anomalous operation sample; (b) Incorrectly classified feature map (left) and the original WST output (right) for an anomalous operation sample.

6. Conclusions and further work

The work shows a successful implementation of the WST network in an audio-based AD problem. It takes advantage of the computational efficiency and stability of the DWT, as well as the invariance of the suggested WST network in the feature extraction process, allowing it to run in quasi-real-time. A CAE trained on normal samples in an unsupervised way fails to reconstruct the more complex of the normal samples to a satisfactory degree, caused by significant intraclass variability among normal operation samples. Therefore, the reconstruction error, which is widely used in existing works to identify anomalous operation samples,

Table 7
The detailed structure of the CAE architecture proposed.

Layer ID	Layer	Number	Size/Kernel size	Strides	Padding/Interpolation	Pool size
1	ZeroPadding2D	Padding = [17, 18], [4, 5]				
2	Conv2D	128	[7, 7]	[1, 1]	Same	–
3	MaxPooling2D	–	–	[2, 2]	Same	[2, 2]
4	Conv2D	64	[5, 5]	[1, 1]	Same	–
5	MaxPooling2D	–	–	[2, 2]	Same	[2, 2]
6	Conv2D	64	[3, 3]	[2, 2]	Same	–
7	MaxPooling2D	–	–	[2, 2]	Same	[2, 2]
8	Conv2D	32	[2, 2]	[1, 1]	Same	–
9	MaxPooling2D	–	–	[2, 2]	Same	[2, 2]
10	Conv2D	16	[2, 2]	[1, 1]	Same	–
11	MaxPooling2D	–	–	[2, 2]	Same	[2, 2]
12	Conv2D	1	[2, 2]	[1, 1]	Same	–
13	Conv2DTranspose	16	[2, 2]	[1, 1]	Same	–
14	UpSampling2D	–	[2, 2]	–	Nearest	–
15	Conv2DTranspose	32	[2, 2]	[1, 1]	Same	–
16	UpSampling2D	–	[2, 2]	–	Nearest	–
17	Conv2DTranspose	64	[2, 2]	[1, 1]	Same	–
18	UpSampling2D	–	[2, 2]	–	Nearest	–
19	Conv2DTranspose	64	[3, 3]	[2, 2]	Same	–
20	UpSampling2D	–	[2, 2]	–	Nearest	–
21	Conv2DTranspose	128	[5, 5]	[1, 1]	Same	–
22	UpSampling2D	–	[2, 2]	–	Nearest	–
23	Conv2DTranspose	1	[7, 7]	[1, 1]	Same	–
24	Cropping2D	Cropping = [17, 18], [4, 5]				

cannot be applied to our use case. Further, reconstructing WST outputs using PCA does not bring satisfactory results.

The final processing pipeline takes advantage of the feature maps extracted by the CAE's hidden layers to identify features that could be used to separate normal and anomalous samples. NCC turns out to be the best-suited error metric for this task. Each incoming sample can have its NCC computed with one or two of the identified maps. Therefore, the classifier can be trained on either one-dimensional or two-dimensional data. Using two maps to train the classifier results in marginally better accuracy with virtually no extra processing time required. The study shows that an SVM with an RBF kernel is the best classifier considering one-dimensional data, and a KNN with k set to 5 performs the best on one-dimensional data. The suggested data-processing pipeline has proven to detect faults of different natures. This is crucial for enabling the uncrewed operation of FastBlade, the experimental site considered in this work. Compared to using one-dimensional data, the use of two-dimensional data has resulted in one more correct anomalous sample classification.

The solution is characterised by low system specificity, as it circumvents the need for detailed knowledge of system assets and their failure modes, making the method easy to extrapolate to other systems in both experimental and industrial domains. Therefore, we have developed a model with satisfactory performance despite the variability of the normal operation samples, and in light of restricted anomalous operation data. Moreover, we demonstrated that WST can be successfully implemented for related AD problems. The major limitations of the method are the need for an anomalous dataset, which is required to train the classifier and identify the most suitable feature maps, as well as the WST's processing time in case GPU support is unavailable. The particular benefits of the method include short development time, which is largely independent of the complexity of the system, and low cost of the AD hardware.

CRediT authorship contribution statement

Marek J. Munko: Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Fergus Cuthill:** Writing – original draft, Validation, Supervision, Resources, Data curation. **Miguel A. Valdivia Camacho:** Writing – original draft, Formal analysis, Data curation. **Conchúr M. Ó Bradaigh:** Writing – original draft, Supervision. **Sergio Lopez Dubon:** Writing – original draft, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization.

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.

Acknowledgements

Marek Jan Munko: The doctoral research project has been funded by Babcock International and The Data Lab, UK. *Miguel Angel Valdivia Camacho*: This research was supported by the Wind & Marine Energy Systems & Structures CDT, UK under the Engineering and Physical Sciences Research Council grant agreement EP/S023801/1.

Sergio Lopez Dubon: This project has received funding from the European Union's Horizon 2020 research and innovation programme, UK under the Marie Skłodowska-Curie grant agreement No 801215 and the University of Edinburgh Data-Driven Innovation programme, UK, part of the Edinburgh and South East Scotland City Region Deal.

All authors: The authors also wish to thank the Supergen ORE Hub for funding received through the Flexible Fund Award, UK FF2020-1063.

Appendix

See Table 7.

Data availability

Data will be made available on request.

References

Ahmad, M.Z., Kamboh, A.M., Saleem, S., Khan, A.A., 2017. Mallats scattering transform based anomaly sensing for detection of seizures in scalp eeg. IEEE Access 5, 16919–16929.

Ahn, H., Yeo, I., 2021. Deep learning based approach to anomaly detection techniques for large acoustic data in machine operation. Sensors 21 (16).

Anden, J., Mallat, S., 2014. Deep scattering spectrum. IEEE Trans. Signal Process. 62 (16), 4114–4128.

Andreux, M., Angles, T., Exarchakis, G., Leonarduzzi, R., Rochette, G., Thiry, L., Zarka, J., Mallat, Stephane, J, Belilovsky, E., Bruna, J., Lostanlen, V., Chaudhary, M., Hirn, M.J., Oyallon, E., Zhang, S., Cella, C., Eickenberg, M., 2018. Kymatio: Scattering transforms in python.

Arts, L.P.A., van den Broek, E.L., 2022. The fast continuous wavelet transformation (fcwt) for real-time, high-quality, noise-resistant time–frequency analysis. Nat. Comput. Sci. 2, 47–58.

Audacity, 2021. Audacity(r) software is copyright (c) 1999-2021 audacity team. the name audacity is a registered trademark.

Bruna, J., Mallat, S., 2013. Invariant scattering convolution networks. IEEE Trans. Pattern Anal. Mach. Intell. 35 (8), 1872–1886.

Buriro, A.B., Ahmed, B., Baloch, G., Ahmed, J., Shoorangiz, R., Weddell, S.J., Jones, R.D., 2021. Classification of alcoholic eeg signals using wavelet scattering transform-based features. Comput. Biol. Med. 139.

Burrows, R., Yates, N.C., Hedges, T.S., Li, M., Zhou, J.G., Chen, D.Y., Walkington, I.A., Wolf, J., Holt, J., Proctor, R., 2009. Tidal energy potential in uk waters. Proc. Inst. Civ. Eng. - Marit. Eng. 162 (4), 155–164.

Chen, X., Gupta, R.S., Gupta, L., 2023. Exploiting the cone of influence for improving the performance of wavelet transform-based models for erp/eeg classification. Brain Sci. 13 (1).

Chen, Z., Yeo, C.K., Lee, B.S., Lau, C.T., 2018. Autoencoder-based network anomaly detection. In: 2018 Wireless Telecommunications Symposium. WTS, IEEE, pp. 1–5.

de Carvalho, G.H.F., Thomaz, L.A., da Silva, A.F., da Silva, E.A.B., Netto, S.L., 2019. Anomaly detection with a moving camera using multiscale video analysis. Multidimens. Syst. Signal Process. 30, 311–342.

Fiore, E.D., Ferraro, A., Galli, A., Moscato, V., Sperlì, G., 2022. An anomalous sound detection methodology for predictive maintenance. Expert Syst. Appl. 209, 118324.

Givnan, S., Chalmers, C., Fergus, P., Ortega-Martorell, S., Whalley, T., 2022. Anomaly detection using autoencoder reconstruction upon industrial motors. Sensors 22 (9), 3166.

Hodson, T.O., 2022. Root-mean-square error (rmse) or mean absolute error (mae) when to use them or not.

Hodson, T.O., Over, T.M., Foks, S.S., 2021. Mean squared error, deconstructed. J. Adv. Modelling Earth Syst. 13.

Horé, A., Ziou, D., 2013. Is there a relationship between peak-signal-to-noise ratio and structural similarity index measure? IET Image Process. 7, 12–24.

Kehtarnavaz, N., 2008. Chapter 7 - frequency domain processing. In: Kehtarnavaz, N. (Ed.), Digital Signal Processing System Design (Second Edition), second edition ed. Academic Press, Burlington, pp. 175–196.

Krishnan, S., 2021. Advanced analysis of biomedical signals. In: Krishnan, S. (Ed.), Biomedical Signal Analysis for Connected Healthcare. Academic Press, pp. 157–222.

Kuo, J.Y., Hsu, C.Y., Wang, P.F., Lin, H.C., Nie, Z.G., 2022. Constructing condition monitoring model of harmonic drive. Appl. Sci. (Switzerland) 12 (19).

Layer, E., Tomczyk, K., 2015. Signal Transforms in Dynamic Measurements. Springer International Publishing.

Li, S.P., Bonner, M., 2022. An interpretable alternative to convolutional neural networks: the scattering transform. J. Vis. 22 (14), 3762.

Liu, Z., Yao, G., Zhang, Q., Zhang, J., Zeng, X., 2020. Wavelet scattering transform for ecg beat classification. Comput. Math. Methods Med..

Lo Scudo, F., Ritacco, E., Caroprese, L., Manco, G., 2023. Audio-based anomaly detection on edge devices via self-supervision and spectral analysis. J. Intell. Inf. Syst..

Lopez Dubon, S., Valdivia Camacho, M., Lam, R., Sellar, B., 2023a. A machine learning approach for tidal flow classification. SSRN.

Lopez Dubon, S., Vogel, C., Cava, D., Cuthill, F., Mccarthy, E., Bradaigh, C.O., 2023b. Fastblade: A technological facility for fullscale tidal fatigue testing.

Lopez Dubon, S., Vogel, C., Cava, D., Cuthill, F., McCarthy, E., Bradaigh, C.O., 2023c. Multi-actuator full-scale fatigue test of a tidal blade. In: Proceedings of the European Wave and Tidal Energy Conference. vol. 15.

Lopez Dubon, S., Vogel, C.R., Garcia Cava, D., Cuthill, F., McCarthy, E.D., O Bradaigh, C.M., 2023d. Fastblade: A technological facility for full-scale tidal blade fatigue testing. SSRN.

Lopez Pinaya, W.H., Vieira, S., Garcia-Dias, R., Mechelli, A., 2020. Autoencoders. In: Mechelli, A., Vieira, S. (Eds.), Machine Learning. Academic Press, pp. 193–208.

Mallat, S., 2012. Scattering Invariant Deep Networks for Classification. Institute for Pure Applied Mathematics (IPAM), UCLA.

McLean, R., Alsop, S., Fleming, J., 2005. Nyquist—overcoming the limitations. J. Sound Vib. 280, 1–20.

McLoughlin, J., Munko, M.J., Camacho, M.A., Cuthill, F., Lopez Dubon, S., 2023. Use of digital image correlation and machine learning for the optimal strain placement in a full scale composite tidal turbine blade. In: Proceedings of the 19th International Conference on Condition Monitoring and Asset Management. Northampton.

McLoughlin, I., Zhang, H., Xie, Z., Song, Y., Xiao, W., Phan, H., 2017. Continuous robust sound event classification using time-frequency features and deep learning. PLoS One 12, e0182309.

Mobtahej, P., Zhang, X., Hamidi, M., Zhang, J., 2021. Deep learning-based anomaly detection for compressors using audio data. In: Proceedings - Annual Reliability and Maintainability Symposium. Vol. 2021-May, Institute of Electrical and Electronics Engineers Inc.

Muller, R., Ritz, F., Illium, S., Linnhoff-Popien, C., 2021. Acoustic anomaly detection for machine sounds based on image transfer learning. In: ICAART 2021 - Proceedings of the 13th International Conference on Agents and Artificial Intelligence. volume 2, SciTePress, pp. 49–56.

Munko, M., Lopez Dubon, S., Cuthill, F., 2024. Normal and anomalous audio data processed with the wavelet scattering transform, collected during the operation of fastblade, a site for regenerative fatigue testing.

Oh, D.Y., Yun, I.D., 2018. Residual error based anomaly detection using auto-encoder in smd machine sound. Sensors 18, 1308.

Pan, M.-C., Sas, P., 1996. Transient analysis on machinery condition monitoring. In: Proceedings of Third International Conference on Signal Processing (ICSP 96). Vol. 2, pp. 1723–1726.

Park, Y., Yun, I.D., 2018. Fast adaptive rnn encoder–decoder for anomaly detection in smd assembly machine. Sensors 18, 3573.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., 2011. Scikit-learn: Machine learning in Python. J. Mach. Learn. Res. 12, 2825–2830.

Peng, Z.K., Chu, F.L., 2004. Application of the wavelet transform in machine condition monitoring and fault diagnostics: A review with bibliography.

Sahidullah, M., Saha, G., 2012. Design, analysis and experimental evaluation of block based transformation in mfcc computation for speaker recognition. Speech Commun. 54, 543–565.

Sharaf, A.I., 2023. Sleep apnea detection using wavelet scattering transformation and random forest classifier. Entropy 25 (3).

Sifre, L., Mallat, S., 2013. Rotation, scaling and deformation invariant scattering for texture discrimination. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. pp. 1233–1240.

STMicroelectronics, 2021. Mems audio sensor omnidirectional digital microphone for industrial applications, imp34dt05 datasheet (revision 4).

Todeschini, G., Coles, D., Lewis, M., Popov, I., Angeloudis, A., Fairley, I., Johnson, F., Williams, A.J., Robins, P., Masters, I., 2022. Medium-term variability of the uks combined tidal energy resource for a net-zero carbon grid. Energy 238.

Toma, R.N., Gao, Y., Piltan, F., Im, K., Shon, D., Yoon, T.H., Yoo, D.S., Kim, J.M., 2022. Classification framework of the bearing faults of an induction motor using wavelet scattering transform-based features. Sensors 22 (22).

Torabi, H., Mirtaheri, S.L., Greco, S., 2023. Practical autoencoder based anomaly detection by using vector reconstruction error. Cybersecurity 6 (1), 1.

Ustubioglu, A., Ustubioglu, B., Ulutas, G., 2023. Mel spectrogram-based audio forgery detection using cnn. Signal Image Video Process. 17, 2211–2219.

Vaish, A., Kumar, M., 2014. A new image compression technique using principal component analysis and huffman coding. In: 2014 International Conference on Parallel, Distributed and Grid Computing. pp. 301–305.

Yoo, J.-C., Han, T.H., 2009. Fast normalized cross-correlation. Circuits Systems Signal Process. 28, 819–843.

Zhao, X., Hu, Y., Liu, J., Yao, J., Deng, W., Hu, J., Zhao, Z., Yan, X., 2024a. A novel intelligent multicross domain fault diagnosis of servo motor-bearing system based on domain generalized graph convolution autoencoder. Struct. Health Monit..

Zhao, X., Yao, J., Deng, W., Ding, P., Ding, Y., Jia, M., Liu, Z., 2023. Intelligent fault diagnosis of gearbox under variable working conditions with adaptive intraclass and interclass convolutional neural network. IEEE Trans. Neural Netw. Learn. Syst. 34, 6339–6353.

Zhao, X., Yao, J., Deng, W., Jia, M., Liu, Z., 2022. Normalized conditional variational auto-encoder with adaptive focal loss for imbalanced fault diagnosis of bearing-rotor system. Mech. Syst. Signal Process. 170, 108826.

Zhao, X., Zhu, X., Liu, J., Hu, Y., Gao, T., Zhao, L., Yao, J., Liu, Z., 2024b. Model-assisted multi-source fusion hypergraph convolutional neural networks for intelligent few-shot fault diagnosis to electro-hydrostatic actuator. Inf. Fusion 104, 102186.

Zhu, X., Zhao, X., Yao, J., Deng, W., Shao, H., Liu, Z., 2024. Adaptive multiscale convolution manifold embedding networks for intelligent fault diagnosis of servo motor-cylindrical rolling bearing under variable working conditions. IEEE/ASME Trans. Mechatronics 29, 2230–2240.