Machine Fault Diagnosis through Vibration Analysis: Continuous Wavelet Transform with Complex Morlet Wavelet and Time–Frequency RGB Image Recognition via Convolutional Neural Network

Dominik Łuczak

Faculty of Automatic Control, Robotics and Electrical Engineering, Poznan University of Technology, 60-965 Poznań, Poland; dominik.luczak@put.poznan.pl

Abstract: In pursuit of advancing fault diagnosis in electromechanical systems, this research focuses on vibration analysis through innovative techniques. The study unfolds in a structured manner, beginning with an introduction that situates the research question in a broader context, emphasising the critical role of fault diagnosis. Subsequently, the methods section offers a concise summary of the primary techniques employed, highlighting the utilisation of short-time Fourier transform (STFT) and continuous wavelet transform (CWT) for extracting time–frequency components from the signal. The results section succinctly summarises the main findings of the article, showcasing the results of features extraction by CWT and subsequently utilising a convolutional neural network (CNN) for fault diagnosis. The proposed method, named CWTx6-CNN, was compared with the STFTx6-CNN method of the previous stage of the investigation. Visual insights into the time–frequency characteristics of the inertial measurement unit (IMU) data are presented for various operational classes, offering a clear representation of fault-related features. Finally, the conclusion section underscores the advantages of the suggested method, particularly the concentration of single-frequency components for enhanced fault representation. The research demonstrates commendable classification performance, highlighting the efficiency of the suggested approach in real-time scenarios of fault analysis in less than 50 ms. Calculation by CWT with a complex Morlet wavelet of six time–frequency images and combining them into a single colour image took less than 35 ms. In this study, interpretability techniques have been employed to address the imperative need for transparency in intricate neural network models, particularly in the context of the case presented. Notably, techniques such as Grad-CAM (gradient-weighted class activation mapping), occlusion, and LIME (locally interpretable model-agnostic explanation) have proven instrumental in elucidating the inner workings of the model. Through a comparative analysis of the proposed CWTx6-CNN method and the reference STFTx6-CNN method, the application of interpretability techniques, including Grad-CAM, occlusion, and LIME, has played a pivotal role in revealing the distinctive spectral representations of these methodologies.

Keywords: fault diagnosis; vibration analysis; continuous wavelet transform; complex Morlet wavelet; convolutional neural network; RGB image recognition; feature extraction; time–frequency domain; short-time Fourier transform; MQTT

1. Introduction

In contemporary settings, our surroundings, spanning from modern factories to urban landscapes and households, are increasingly populated by a plethora of electromechanical systems. These systems are not only substantial energy consumers but also possess finite lifespans. Implementing effective maintenance practices for these devices is essential to cost efficiency and environmental sustainability by mitigating the generation of electronic waste. In scientific and popular science articles, terms such as “electronic trash” or simply “trash”
are encountered, but, more prevalently, expressions such as “electronic waste” (e-waste) or WEEE (waste electrical and electronic equipment) are used. The complexity of industrial machinery, which incorporates both electrical and mechanical components, increases the challenge of maintenance. Proactive maintenance strategies not only avert production disruptions but also protect equipment from inadvertent damage.

The landscape of fault diagnosis, a key element in maintenance, becomes progressively more intricate as the volume of scientific literature grows. A search on Google Scholar under the keyword ‘fault diagnosis’ yields nearly 1.6 million articles, while narrowing the scope to “industrial machines” with the operator “AND” still results in a substantial 1.6 thousand articles. Concurrently, our contemporary milieu witnesses a surge in the capability to exchange data globally through the Internet, encompassed by terms like IoT (Internet of Things) and IIoT (industrial Internet of Things), the latter being integral to industrial interconnections. The notion of Industry 4.0, which outlines the organisation of production processes through autonomous communication between technological devices along the value chain, has evolved into the concept of Industry 5.0, which emphasises sustainability, approaches centred around humans, and the development of a resilient European industry [1].

The intersection of fault diagnosis and connectivity with IoT emerges as a thriving realm of research, evidenced by approximately 12.3 thousand articles within this multidisciplinary domain. A novel perspective is offered through the exploration of patent databases, where the International Patent Classification (IPC) has expanded its purview to include a dedicated subclass G16Y, specifically focussing on information and communication technology (ICT) tailored for the IoTs. Within this subclass, G16Y40/00 refers to IoT distinguished by its focus on information processing, with detailed classifications such as G16Y40/10 for observation and monitoring, G16Y40/20 for examination and diagnosis, and G16Y40/40 for conservation of things. The patent search engine at the Espacenet service reveals around 2.8 thousand intellectual properties under these classifications.

In the subsequent sections, an attempt is made to provide a concise examination of fault diagnosis and IoT. However, given the vastness of the literature and patent databases, coupled with the constraints of article length and the author’s time resources, certain aspects have necessarily been omitted.

The presented methodology demonstrates the effective application of convolutional neural networks (CNNs) in the recognition of multiscalograms organised into RGB images for fault diagnosis, eliminating the necessity for the prior selection of vibration axes. The innovative approach involves the recognition of six-scalogram RGB representations, leveraging the improved utilisation of continuous wavelet transform (CWT) in lieu of the short-time Fourier transform (STFT). A comparative analysis with alternative methods is summarised in Table 1.

It should be noted that CNNs have established their prowess in vision-based recognition and applications [2]. In this context, the proposed method extends the application of CNNs to the recognition of specially crafted time–frequency images, showcasing the adaptability of CNNs in fault diagnosis through the devised approach.

### Table 1. Comparison of proposed fault diagnosis methods.

<table>
<thead>
<tr>
<th>Internet of Things Linkage</th>
<th>Classes of Faults</th>
<th>Sensor Type</th>
<th>Features Extraction Technique</th>
<th>Features</th>
<th>Classifier Algorithm</th>
<th>Publication</th>
</tr>
</thead>
<tbody>
<tr>
<td>MQTT, HTTP</td>
<td>Demonstrator with fan blade imbalance (normal, fan turn off, fan fault)</td>
<td>Three-axis accelerometer and gyroscope</td>
<td>CWT with complex Morlet wavelet</td>
<td>RGB image made of six-time–frequency (time-scale) domain data</td>
<td>CNN</td>
<td>Proposed</td>
</tr>
<tr>
<td>MQTT, HTTP</td>
<td>Demonstrator with fan blades imbalance (normal, fan turn off, fan fault)</td>
<td>Three-axis accelerometer and gyroscope</td>
<td>SDFT (sliding discrete Fourier transform) or STFT at 6 axes</td>
<td>RGB image made of six spectrograms</td>
<td>CNN</td>
<td>[3]</td>
</tr>
</tbody>
</table>
The proposed CWTx6-CNN method is a fault diagnosis method that was tested with a vibration signal from more than one axis. The fault diagnosis system, illustrated in Figure 1, can be designed with different parts: a one- or multiple-axis sensor for acceleration, feature extraction method, and features and decision making classification. Additionally, the fault diagnosis system can have Internet of Things connectivity. In [4], the authors use a CNN to recognise the RGB image made by time–frequency analysis of vibrations on one axis, where colour is used to represent the magnitude of the frequency components instead of a greyscale image. In the proposed CWTx6-CNN method, a colour RGB image is created by six time–frequency images obtained by CWT with a complex Morlet wavelet.

![Figure 1. Structure overview of the data-driven fault diagnosis system.](image-url)

The suggested approach was compared with other methods to underline the increase in new knowledge by constructing RGB images from time–frequency data from a six-degrees-of-freedom inertial measurement unit (6DOF IMU) as a three-axis accelerometer and three-axis gyroscope obtained by CWT. Other methods use single-axis spectrograms calculated from a single axis. In the proposed method, all axes are used and it is shown how to combine them into one RGB image at six axes that is recognised by a CNN (CWTx6-CNN). In the previous stage of research, the RGB image was constructed by STFT on the six-axis and then recognised by a CNN (STFTx6-CNN) [3]. The benefit of changing STFT to CWT with a complex Morlet wavelet is better frequency localisation. Mechanical vibrations move in a specific direction; therefore, the three-axis sensor covers all directions. In contrast, a one-axis sensor can sense changes in a single direction, which requires preliminary knowledge of the vibration direction in the diagnosed machine for a normal operation and fault condition. The used 6DOF IMU sensor is a low-cost component. The accelerometer gives data of linear acceleration in three axes and the gyroscope gives information of angular acceleration. Therefore, both sensors cover linear and torsional displacements in the 3 axes, which can appear in the fault condition.

The fault diagnosis system, illustrated in Figure 1, operates on the principle of fault detection by monitoring changes in features over time. Employing a client–server architecture within the ICT network, this system extracts features from raw or preprocessed data to facilitate fault detection. In essence, fault detection entails recognising alterations in the...
device condition induced by one or more faults, akin to anomaly state detection, where any condition deviating from routine behaviour is identified. Moving beyond detection, the system performs fault isolation by pinpointing the specific modules of the machine affected, and fault identification quantifies the extent of the damages.

The sensor depicted in Figure 1 can take various forms, serving as a dedicated sensor designed solely for fault diagnosis or as an integral part of the system, utilised by control algorithms. Investigating electromechanical machines, rolling bearing \[5,7,8\] or power systems can employ various sensors and signals, including measurements of current \[9,10\] and voltage \[11,12\], torque \[13,14\], angular velocity/position \[15,16\], linear 3DOF acceleration/velocity/position \[3,5\], a laser Doppler vibrometer \[17\], the transmittance and reflectance of an omnidirectional antenna \[18\], strain/force \[19–22\], energy consumption \[23–26\], inner/outer temperature at specific locations \[27,28\], or outer-part temperature captured by an infrared camera \[2,29\]. The selection of sensors and signals is dependent on the frequency range and specific characteristics of the electromechanical system under examination. Possibilities include displacement \[30\], vibrations \[3,4,7,31,32\], sound \[33–35\], sound recorded with multiple microphones \[36\], or ultrasound \[37,38\]. Investigations may also include vibro-acoustic analysis \[39\], chemical analysis \[40,41\], spectral imaging for chemical analysis \[42–45\], a camera capturing images within the visible human colour spectrum \[46–49\], and even signals translated into virtual images \[10,50–53\]. This versatility in sensor types and signals enables a thorough exploration of the machine or system.

This manuscript is organised into distinct sections. In the Introduction, the research objectives and the importance of fault diagnosis in electromechanical systems are outlined. Moving on to the second section, the exploration of feature extraction in the time-scale (time–frequency) domain is initiated. Here, STFT and CWT are presented to extract frequency components from the signal. The third section focusses on the demonstrator of machine fault diagnosis, where data from each axis of the 6DOF IMU undergo transformation into the time–frequency domain CWT with a complex Morlet wavelet. This process results in a two-dimensional signal containing 65 frequencies across 96 time points for each axis. Moving to the fourth section, the results of feature extraction by CWT with a complex Morlet wavelet and subsequent fault diagnosis by the CNN are presented. This involves the transformation of the initial time-domain data into time–frequency RGB images using the process outlined. Each RGB image requires the CWT transformation of all six axes of the accelerometer and gyroscope, resulting in a total of six scalograms. The fifth and final section delves into a discussion of the proposed method, exploring the advantages and limitations of time–frequency feature extraction and comparing methods such as STFT and CWT with a complex Morlet wavelet. The author highlights a previous stage of research involving STFT with a CNN, citing satisfactory classification results. However, concerns about the time–frequency method are discussed, particularly its tendency for spectral components to appear blurred. Finally, the benefits of the proposed method are underlined.

2. Extracting Features in Time-Scale (Time–Frequency) Domain

The frequency components can be extracted from the analysed signal using fast Fourier transform (FFT). However, this analysis falls short of addressing a critical question: whether this component is a singular occurrence within the signal or if it manifests multiple times across the time domain. To delve into this question, a time–frequency analysis becomes imperative. Unfortunately, the application of short-time Fourier transform (STFT) to address this problem introduces the blurring of certain frequencies observed in FFT \[54\]. Recognising this limitation, the author sought alternative tools to achieve effective time and frequency localisation while minimising the blurring of frequency data. Continuous wavelet transform (CWT), employed with a meticulously chosen mother wavelet and optimal parameter selection, yields more satisfactory results. In particular, the use of CWT with a complex Morlet wavelet produces superior results compared to STFT, providing a clearer representation of both the time and frequency characteristics in the analysed signal \[54\].
In the previous author stage of research on vibration analysis for electric direct drive with CWT [54], the velocity signal was analysed using STFT and CWT with excitation by a linear chirp signal, without additional vibration sensors, and there was no image recognition as a decision-making system. In this stage of research, the author uses an additional sensor inertial measurement unit (IMU) with six axes, and then data are converted into a RGB image by continuous wavelet transform (CWT) and recognised by the CNN.

Short-time Fourier transform (STFT) is a technique designed to transform a one-dimensional time-based function into a bidimensional representation incorporating both time and frequency. This procedure employs a constant-duration time frame that traverses the analysed signal. In every time segment, a fast Fourier transform (FFT) is computed. This procedure is reiterated for subsequent sets of samples, with the time window shift being a configurable parameter. The time segment shift can be perceived like an overlap between consecutive time windows, allowing for a more nuanced analysis. In this study, the FFT calculation was performed subsequent to each new sample, resulting in a K-1 overlap of segment to segment, where K represents the length of the time window with a fixed number of samples. Essentially, the size of the step is adjusted to one sample.

Without any modification, the time window is configured as a rectangle. On the other hand, it is recommended to explore other well-known window shapes to mitigate spectrum leakage. In this study, a Kaiser window shape was used. The calculation of STFT follows this process:

$$F(\tau, f) = \int_{-\infty}^{+\infty} f(t) \cdot w(t - \tau) e^{-j2\pi f \cdot t} dt$$

where $\tau$ denotes a shift in time, $f$ denotes frequency, $w(t)$ is a time window of constant length, $f(t)$ the examined function, and $F$ the time–frequency result as complex numbers.

The time granularity is dependent on the size of the step, which, in this study, was set to one sample. On the contrary, the frequency granularity is computed similarly to the FFT. Consequently, the length of the constant-length time window determines the frequency granularity, expressed by the formula $f_{res} = f_s / K$, where $K$ denotes the length of the segment in samples and $f_s$ represents the frequency of sampling.

Continuous wavelet transform (CWT) is a procedure designed to transform a unidimensional time-based function into a bidimensional representation incorporating both time and scale. The primary benefit of this method lies in the ability to scale the length of the time segment. This enables the accurate selection of signal frequencies based on the window’s scale, allowing it to be tailored to match the period of the dominant signal component. Notably, varying the size of the time segment can be applied to acquire the single period of both low- and high-band frequency components. CWT is an integral transform that employs a selectable kernel function. The calculation of wavelet transform is conducted by:

$$W(a, b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} f(t) \cdot \Psi \left( \frac{t - b}{a} \right) dt$$

where $W(a, b)$ signifies the resulting coefficients of the continuous wavelet transform, where $a$ represents the scaling factor, $b$ denotes the shift factor, $\Psi$ is the conjugate counterpart of the primary wavelet function, and $f(t)$ the examined function. A detailed explanation of the influence of scale and shift can be found in [54,55]. The mother wavelet function must meets kernel conditions; therefore, in the literature, Daubechies wavelets [56], Mexican hat wavelets, complex Gaussian wavelets, complex Shannon wavelets, Morlet wavelets, and complex Morlet wavelets can be found [57]. Complex Morlet wavelets have a smooth and single dominant spectrum with the ability to select dominant frequency. On the contrary, Mexican hat wavelets have a fixed dominant (middle) frequency for the mother wavelet [55]. Complex Shannon wavelets have complicated equations and have a rectangular frequency shape with ripples for all dominant (middle) frequencies [55]. The choice was made to use the complex Morlet wavelet because of the simplicity and flexibility of the selection of
dominant frequency and its bandwidth. Other wavelet families require more studies in the field of fault diagnosis. The author in the previous stage of research used a complex Morlet wavelet in electric motor velocity analysis with good results [54] and decided to use and verify it in the proposed CWTx6-CNN at 6DOF data.

The Morlet complex wavelet function, denoted as $\Psi_M$, is defined by the expression:

$$\Psi_M(t) = \frac{1}{\sqrt{\pi f_b}} e^{\frac{-t^2}{f_b}}$$

where $f_c$ represents the dominant (centre) frequency and $f_b$ represents the variability or frequency width of the wavelet. The kernel of the Morlet complex wavelet comprises two primary components. The first component, $e^{2\pi f_c t}$, is derived from Euler’s formula and results in $\cos(2\pi f_c t) + j\sin(2\pi f_c t)$, which is denoted as the kernel of Fourier transform. The other component $e^{-t^2/f_b}$ is interpreted as shape of the envelope of the window in time.

The outcome of CWT is a function of $a$, representing the scale, and $b$, indicating the shift factor. Consequently, it is aptly termed a time-scale (scalogram) examination. However, for increased utility, transforming this analysis from a scalogram to spectrogram (pseudo-spectrogram or time–pseudo-frequency) is advantageous. This conversion is achieved through the following equation:

$$f_{\text{pseudo}} = \frac{f_{\text{middle}}}{a \cdot f_s}$$

Here, $f_{\text{middle}}$ denotes the middle wavelet frequency and $f_s$ represents the time of sampling. The wavelet function’s FFT examination may encompass multiple-frequency components, but only the dominant frequency is selected and retained, rendering it a pseudo-frequency. In particular, the middle frequency for the complex Morlet wavelet equals $f_c$.

3. Demonstrator of Machine Fault Diagnosis

The data collected from each axis of the 6DOF inertial measurement unit (IMU) were converted into a spectrogram (time–frequency) by CWT with a complex Morlet wavelet. The CWT process produced a bidimensional signal consisting of 65 frequencies at 96 time points for each axis. This process was iterated for both the accelerometer and gyroscope axes. Consequently, six time–frequency images were generated for each condition (class): idle, normal, and fault. These six spectrogram images were amalgamated into a single RGB image measuring $96 \times 130 \times 3$. The schematic of the image generation and CNN architecture is shown in Figure 2. Representative RGB images for each class are highlighted in Section 4. This comprehensive transformation and image representation offer visual insight into the time–frequency characteristics of the IMU data for each operational class. The CNN architecture is shown in Figure 2, along with the representation of the inputs. The CNN receives an RGB image as input, constituting an array of dimensions $96 \times 130 \times 3$. The training parameters are specified as follows: a maximum number of epochs set to 5, an initial learning rate of $1 \times 10^{-4}$, utilisation of the stochastic gradient descent with momentum (SGDM) optimiser, and an execution environment employing graphics processing unit (GPU) acceleration. The proposed CWTx6-CNN method shown in Figure 2 is the next stage of the investigation and improves the previously developed STFTx6-CNN method published in [3].

The demonstration illustrates the recognition of computer fan operation at one of three states: idle, normal, or fault, where the fault is induced by the addition of blue colour clip paper at a single blade of the fan. The setup for the demonstration, as shown in Figure 3, comprises a NUCLEO board equipped with an STM32F746ZG microcontroller responsible for handling the IMU-6 DOF MPU6050 sensor. Data are gathered synchronously with a consistent sampling interval of 5 ms, equivalent to a sampling frequency of 200 Hz. The buffer containing 128 samples from IMU-6 DOF is converted
into JSON (JavaScript Object Notation) format, taking the structure of `{"accelerometer": ["x":[],"y":[],"z":[]], "gyroscope": ["x":[],"y":[],"z":[]]}`. The values of samples are given in arrays `"[]"`. This collection of measurements is transmitted via MQTT (Message Queuing Telemetry Transport) by a microcontroller client to MQTT broker at a laptop as shown in Figure 4. This integrated setup allows for the monitoring and classification of the computer fan’s operational state with immediate analysis and response to faults in real-time. The proposed CWTx6-CNN method is the next stage of evaluation on the same demonstration rig shown in Figure 3 that was used in previous research published in [3].

**Figure 2.** Proposed method named CWTx6-CNN as RGB image made of six CWT scalograms and recognised by CNN with given architecture.

**Figure 3.** Demonstration for fan fault diagnosis.

**Figure 4.** MQTT communication architecture.

### 4. Results of CWT Feature Extraction with Complex Morlet Wavelet and CNN Fault Diagnosis Using CNN

For each operational class, the data initially collected in the time domain (see Figure 5) were transformed into time–frequency RGB images (see Figure 6) according to the image creation shown in Figure 2. The time domain observations, whose fragments are shown in Figure 5, are the same as those previously investigated in [3]; however, the extraction of time–frequency features by CWT with a complex Morlet wavelet is novel in the proposed CWTx6-CNN method. A single RGB image requires a CWT transformation of each axis of
the accelerometer and gyroscope, which is six scalograms in total. Two scalograms that correspond to the same axis are combined in a single channel of colour. An example RGB image of six scalograms and its red, green, and blue channels is shown in Figure 7. The consolidated dataset comprises a total of 8160 RGB images, distributed among the classes as follows: 2720 RGB images for fault, 2720 for idle, and 2720 for normal. Subsequently, the scalogram dataset was partitioned into parts for training and validation. Within the dataset of images, 80% (6528 colour images) was allocated for the training part, while the remaining 20% constituted the testing set (1632 colour images), selected randomly. Training of the CNN was carried out using the Matlab Deep Learning Toolbox, using the computational power of an NVIDIA GPU together with CUDA® (Compute Unified Device Architecture, Santa Clara, CA, USA). The accuracy of the training process is very good and is shown in Figure 8. The validation of the trained CNN demonstrated a commendable classification performance, as evidenced by the confusion matrix presented in Figure 9. This analysis affirms the CNN’s efficacy in accurately categorising the RGB images into their respective classes.

Figure 5. IMU 6DOF time domain data: accelerometer and gyroscope, where red—X axis, green—Y axis, and blue—Z axis.

Figure 6. CWT time–frequency RGB images made from 6 scalograms for each state: idle (left), normal (middle), fault (right).
Figure 7. Proposed CWT RGB image for fault condition for 6DOF IMU data. From left to right: RGB image made of 6 scalograms, red channel made of 2 scalograms, green channel made of 2 scalograms, and blue channel made of 2 scalograms.

Figure 8. Accuracy during training (left): dark blue—training smoothed; light blue—training; dotted—validation; and loss during training (right): orange—training smoothed; dotted—validation.

Figure 9. Matrix of confusion after training (train—(left); test—(right)).

5. Discussion

Time–frequency feature extraction can be carried out by STFT or CWT with a complex Morlet wavelet. In a previous stage of the research, the author used STFT with a CNN [3]. The former method [3] gives satisfactory classification results; however, time–frequency has large lickouts that can be seen as blurred spectral components, which are shown in Figure 10.
A comparison was made for the same conditions to obtain fair results. Scales for CWT analysis were selected by relationship (4), where pseudo-frequency was selected as the equal frequencies as in the reference STFTx6-CNN method. The time length was 96 samples for each of the six channels (three axes of the accelerometer and three axes of the gyroscope). The attributes of a complex Morlet wavelet (3) must be chosen by the system designer. The author selected complex Morlet wavelet parameters as follows: \( f_c = 5 \) and \( f_b = 10 \). The chosen wavelet and preferred parameters give sharp spectral time-scale (time–frequency) results for a short time window (96 samples, which is equivalent to a time of 480 ms). The proposed method has better spectral sharpness, as shown in Figure 6 compared to Figure 10, which allows one to extract frequency features more precisely compared to the STFT. To facilitate a more effective comparison, Figure 11 shows the RGB image for the fault class generated by the proposed method on the left side and the reference method on the right side. The confusion matrix (see Figure 9) of the proposed CWTx6-CNN method has the same quality as the STFTx6-CNN reference method tested on the equal demonstrator. However, the advantage of the proposed method can be noticed in the quality of the extracted frequency components. In the proposed method, the single-frequency components are concentrated, enabling a more comprehensive representation of symptoms with greater reach and clarity. This concentration contributes to a more nuanced and detailed depiction of fault-related characteristics, enhancing the diagnostic capabilities of the system.

**Figure 10.** Reference STFT made from 6 scalograms for each state: idle (left), normal (middle), fault (right).

**Figure 11.** Comparison of the RGB image (six time–frequency components) for the class fault calculated by proposed approach’s CWT with complex Morlet wavelet (left) and reference STFT method (right).
In the context of the presented case, the use of interpretability techniques has emerged as a crucial aspect, addressing the growing need for transparency in complex neural network models. Specifically, methods such as Grad-CAM (gradient-weighted class activation mapping) [58], occlusion [59], and LIME (locally interpretable model-agnostic explanation) [60] have played a pivotal role in enhancing the understanding of the model’s operation. In the comparative examination between the proposed CWTx6-CNN method and the STFTx6-CNN reference method, interpretability techniques such as Grad-CAM, occlusion, and LIME play an important role in unravelling the distinct characteristics of their spectral representations. Grad-CAM, as demonstrated in Figure 12 (left), Figure 13 (left), Figure 14 (left), Figure 15 (left), Figure 16 (left) and Figure 17 (left), provides valuable insights by highlighting the regions of interest in spectral images, offering clarity regarding the decision-making process of the neural classifier. Occlusion, another interpretability method, aids in understanding the impact of occluded areas on model predictions (see Figure 12 (middle-left), Figure 13 (middle-left), Figure 14 (middle-left), Figure 15 (middle-left), Figure 16 (middle-left) and Figure 17 (middle-left)). This technique reveals essential features and regions that influence the final results, contributing to a more transparent and interpretable model. Additionally, the incorporation of LIME further enhances the comprehensibility of the neural network’s decision-making context by generating locally faithful interpretations (see Figure 12 (middle-right), Figure 13 (middle-right), Figure 14 (middle-right), Figure 15 (middle-right), Figure 16 (middle-right) and Figure 17 (middle-right)). The interpretability of both the proposed CWTx6-CNN method and the reference STFTx6-CNN method was enhanced by presenting a unified Figure 18 that features marked frequency regions. Orange dotted lines were strategically employed to delineate and emphasise the selected regions of interest in the interpretability analysis of both the proposed CWTx6-CNN method and the reference STFTx6-CNN method. These interpretability techniques collectively confirm that the neural classifier focuses on key features extracted from spectral images, emphasising the robustness of the presented approach across various conditions. This transparency ensures that the model’s decisions are not driven by spurious correlations but are rooted in meaningful features relevant to the classification task. The benefit of the proposed CWTx6-CNN method becomes evident in the quality of extracted frequency components. The CWTx6-CNN method concentrates single-frequency components more effectively, providing a more comprehensive representation of fault-related characteristics. This concentration enhances the diagnostic capabilities of the system, offering detailed and nuanced insight into the spectral features associated with different fault classes.

Figure 12. Explaining reference STFT with CNN network predictions for class fault using: Grad-CAM (left), occlusion sensitivity (middle-left), LIME (middle-right), and input image of reference STFTx6 (right).
Figure 13. Explaining reference STFT with CNN network predictions for class idle using: Grad-CAM (left), occlusion sensitivity (middle-left), LIME (middle-right), and input image of reference STFTx6 (right).

Figure 14. Explaining reference STFT with CNN network predictions for class normal using: Grad-CAM (left), occlusion sensitivity (middle-left), LIME (middle-right), and input image of reference STFTx6 (right).

Figure 15. Explaining proposed CWT with CNN network predictions for class fault using: Grad-CAM (left), occlusion sensitivity (middle-left), LIME (middle-right), and input image of proposed CWTx6 (right).
Figure 16. Explaining proposed CWT with CNN network predictions for class idle using: Grad-CAM (left), occlusion sensitivity (middle-left), LIME (middle-right), and input image of proposed CWTx6 (right).

Figure 17. Explaining proposed CWT with CNN network predictions for class normal using: Grad-CAM (left), occlusion sensitivity (middle-left), LIME (middle-right), and input image of proposed CWTx6 (right).

The process of converting raw data, comprising $96 \times 6$ points across six axes, using continuous wavelet transform (CWT), into RGB images and subsequently saving them on an external SD card was executed within a total duration of 591 s for a set of 8160 images. The average time that encompassed both the conversion and storage phases was approximately 73 milliseconds. In a separate test, the CWTx6 calculation and RGB image creation were performed without a file-saving operation, allowing for the isolation of conversion time. This phase consumed 273 s for the same set of 8160 images, resulting in an average conversion time of 34 milliseconds for the six-axis time–frequency CWT analysis. The calculations were performed iteratively for all 8160 images, and the elapsed time was measured using MathWorks MATLAB’s built-in functions ‘tic’ and ‘toc’.

The tests were carried out on MathWorks Matlab R2023a, utilising Wavelet Toolbox version 6.3 and Deep Learning Toolbox version 14.6, on a laptop with the following specifications: Intel i7-4720HQ processor with 3.60 GHz, four hardware cores, eight threads, RAM memory of 16 GB, NVIDIA GeForce GTX 960M, and an SSD drive Samsung 850 PRO with 256 GB capacity. The computational steps involving the CWTx6 and CNN output for all 8160 RGB images were completed in 331 s, resulting in a comprehensive response time of less than 41 milliseconds for the entire method.

The data collected during these experiments are summarised and presented in Table 2. As shown in Figure 4, the fault prediction occurs after the MQTT client collects raw sensor data from the sensor and transmits an array of $128 \times 6$ samples. The collection of $128 \times 6$ samples, with a sampling time of 5 ms, took 640 ms. Consequently, the combined execution time of CWT analysis across six axes and CNN classification output should ideally be less than 640 ms for real-time feasibility. The proposed CWTx6-CNN
method, as demonstrated, accomplished this in less than 50 ms, confirming its suitability for real-time analysis.

![Figure 17](image1.png)

Figure 17. Explaining proposed CWT with CNN network predictions for class normal using: Grad-CAM (left), occlusion sensitivity (middle-left), LIME (middle-right), and input image of proposed CWTx6 (right).

![Figure 18](image2.png)

Figure 18. Interpretability of proposed CWTx6-CNN method and reference STFTx6-CNN method. Orange dotted lines—strategically and arbitrary employed regions of interest in the interpretability analysis.

<table>
<thead>
<tr>
<th>Time Measurement Condition for 8160 RGB Images</th>
<th>Total Time in Seconds for all Iterations (Ceiling Round)</th>
<th>Average Time of Single Iteration in Milliseconds (Ceiling Round)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CWTx6 image creation and classification by CNN</td>
<td>331 s</td>
<td>41 ms</td>
</tr>
<tr>
<td>CWTx6 image creation</td>
<td>273 s</td>
<td>34 ms</td>
</tr>
<tr>
<td>Classification by CNN of the same image 8160 times</td>
<td>24 s</td>
<td>3 ms</td>
</tr>
<tr>
<td>CWTx6 image creation and save to SD card for training</td>
<td>591 s</td>
<td>73 ms</td>
</tr>
</tbody>
</table>

**Funding:** This work was funded by Poznan University of Technology under Grant 0214/SBAD/0243.

**Data Availability Statement:** Data are contained within the article.

**Conflicts of Interest:** The author declares no conflicts of interest.

**References**


10. Łuczak, D.; Brock, S.; Siembab, K. Fault Detection and Localisation of a Three-Phase Inverter with Permanent Magnet Synchronous Motor Load Using a Convolutional Neural Network. Actuators 2023, 12, 125. [CrossRef]


29. Li, Z.; Zhang, Y.; Abu-Siada, A.; Chen, X.; Li, Z.; Xu, Y.; Zhang, L.; Tong, Y. Fault Diagnosis of Transformer Windings Based on Decision Tree and Fully Connected Neural Network. Energies 2021, 14, 1531. [CrossRef]


31. Meyer, A. Vibration Fault Diagnosis in Wind Turbines Based on Automated Feature Learning. Energies 2022, 15, 1514. [CrossRef]


38. Wang, W.; Xue, Y.; He, C.; Zhao, Y. Review of the Typical Damage and Damage-Detection Methods of Large Wind Turbine Blades. *Energies* 2022, 15, 5672. [CrossRef]


50. Ullah, Z.; Lodhi, B.A.; Hur, J. Detection and Identification of Demagnetization and Bearing Faults in PMSM Using Transfer Learning-Based VGG. *Energies* 2020, 13, 8383. [CrossRef]


54. Luczak, D. Mechanical Vibrations Analysis in Direct Drive Using CWT with Complex Morlet Wavelet. *Power Electron. Drives* 2023, 8, 65–73. [CrossRef]


Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.