Anomaly Detection Framework for Rotary Equipment Using Continuous Wavelet Transform and U-Net Autoencoders

Mohamed Zamil Kanjirathingal Rafeek[1], Ulrich Schäfer[2]

[1,2] *Ostbayerische Technische Hochschule Amberg-Weiden*
Department of Electrical Engineering, Media, and Computer Science, D-92224 Amberg, Germany
m.kanjirathingal-rafeek@oth-aw.de, u.schaefer@oth-aw.de

ABSTRACT

Recent advances in data-driven methods, particularly deep learning, have transformed predictive maintenance for rotary machinery. These methods enable intelligent, sensor-based condition monitoring from unlabeled operational data, even under rare-fault conditions. This study proposes an unsupervised anomaly detection framework for rotary equipment that utilizes continuous wavelet transform (CWT) to transform unlabeled, multichannel vibration signals into stacked time-frequency scalograms using complex Morlet wavelet. These scalograms are then processed by an enhanced U-Net deep convolutional autoencoder (CWT-U-Net CAE), which learns features of healthy operational conditions and detects anomalies by identifying significant deviations in reconstruction error. Coupled with its edge-compatibility, the framework enables scalable real-time condition monitoring in industrial environments. A custom test bench with an induction motor was used to obtain realistic vibrational signatures under normal operating conditions, assessing the effectiveness of the proposed approach.

1. INTRODUCTION

Rotary equipment, particularly induction motors, is responsible for driving a wide range of industrial assets. Ensuring reliable operation of rotary equipment is a key strategy in Industry 4.0. Recent advances integrate sensor-based condition monitoring with data-driven machine learning to improve efficiency, predict failures, and optimize maintenance schedules (Benhanifia, Cheikh, Oliveira, Valente, & Lima, 2025). Manufacturing environments operate predominantly under normal conditions, and faulty/anomalous conditions rarely occur. This can cause class imbalances when relying on supervised learning approaches and has motivated the adoption of unsupervised anomaly detection methods that leverage abundant healthy operational data (Pang, Shen, Cao, & van den Hengel, 2021).

Recent developments in MEMS accelerometers have further advanced condition monitoring through improved vibration analysis. These sensors offer the necessary bandwidth and noise performance for cost-effective monitoring (Jain, Patel, & Raj, 2020). These advances address previous limitations where high-quality vibration data acquisition required expensive piezoelectric systems that hindered scalable deployment.

However, rotary vibration signals are inherently non-stationary, and this characteristic limits the effectiveness of conventional frequency-domain analysis. Time-frequency methods, such as continuous wavelet transform scalograms, preserve both temporal dynamics and spectral content, thus capturing transient fault signatures (Bernitsas & Kourkoutos-Ardavanis, 2021). The image-like structure of scalogram representations further enables the application of convolutional architectures, which originally is designed for visual pattern recognition tasks.

Building on this, autoencoder networks, when combined with convolutional layers, have demonstrated effectiveness for unsupervised anomaly detection using visual features. By learning to reconstruct normal patterns, the models flag deviations during inference, making them well suited for analyzing scalogram representations of vibration signals. However, traditional CNN autoencoders, while effective, often struggle to capture fine-grained spatial details due to information loss in pooling operations. U-Net architectures address this limitation through skip connections that preserve spatial resolution throughout the encoding-decoding process (Yedurkar et al., 2023). This capability proves particularly important when analyzing scalogram representations where localized time-frequency anomalies must be accurately reconstructed.

In industries, edge computing improves continuous vibration monitoring by processing large volumes of high-frequency data, which are impractical to stream to the cloud in real time. By shifting the processing pipeline closer to the machine, edge devices allow complex models such as U-Net-based autoencoders to be deployed directly at the source of data. This architecture supports rapid reconstruction error analysis and anomaly detection without dependency on remote servers, ensuring both resilience to network interruptions and compliance with industrial demands for real-time decision support (Kanungo, 2025). The synergy of MEMS sensing, time-frequency analysis, and edge-optimized deep learning facilitates the practical deployment of unsupervised anomaly detection in constrained industrial environments.

This study explores a unified framework that combines advanced signal representations with deep learning architectures. It addresses challenges in predictive maintenance through unsupervised learning, while dealing with the non-stationary nature of vibration signals and the practical constraints of deploying models on edge hardware. Data, the non-stationary nature of vibration signals, and the practical constraints of deploying models on edge hardware. The following section reviews previous work on time-frequency analysis, deep autoencoder-based anomaly detection, and edge-enabled.

2. RELATED WORK

Rotary equipment produces inherently non-stationary vibration signals, making time-frequency analysis crucial for fault detection and health monitoring. The continuous wavelet transform (CWT) offers a more effective trade-off between temporal and spectral resolution, allowing subtle anomalies in bearings and motor systems to be detected (Guo, Yang, Gao, & Zhang, 2018; Zhang & Chen, 2023; Chou & Wang, 2025). By representing vibration signals as two-dimensional scalograms, CWT facilitates the application of deep learning architectures for enhanced analysis.

However, most studies treat these scalograms as static images, overlooking temporal dependencies across consecutive signal windows. To address this limitation, hybrid models such as CNN-LSTM and residual CNNs have been developed to incorporate temporal correlations, leading to improved robustness under variable operating conditions (Yang & Li, 2021; Tang & Han, 2024; Zheng & Xu, 2023). Attention mechanisms have also been introduced to highlight weak fault signatures, demonstrating improved anomaly localization in industrial data sets (Park & Kim, 2023).

Beyond vibration-only analysis, researchers have explored multi-sensor fusion approaches, particularly combining vibration with stator current or power signals, to improve diagnostic accuracy (Li & Zhou, 2023; Huang & Sun, 2025). Although these strategies report improved performance on benchmark datasets, they often rely on high-precision sensors and controlled experimental setups, which limits their practicality in cost-constrained industrial environments. Studies employing MEMS accelerometers demonstrate their potential for low-cost condition monitoring, although challenges remain in maintaining signal quality and noise robustness (Chen & Wu, 2024; Ventricci & Marino, 2024).

In parallel to methodological advances, there is growing interest in the deployment of anomaly detection models at the edge. Lightweight CNN and CAE architectures implemented on embedded or FPGA-based platforms have demonstrated feasibility for real-time fault detection (Malviya & Singh, 2022; Chou & Wang, 2025). However, most evaluations are based on curated datasets, which do not fully capture the variability and unpredictability of real-world factory environments, such as variable load conditions, temperature fluctuations, and concurrent machine operations.

Previous research has advanced CWT-based feature extraction, hybrid deep learning architectures, sensor fusion, and edge-compatible deployment. However, a unified framework that combines these advancements with cost-effective MEMS-based sensing, CWT-derived scalogram representations, and unsupervised CAE-driven anomaly detection for real-time edge deployment remains underexplored. This study addresses this gap by proposing an end-to-end methodology and presents the following contributions.

- A cost-effective predictive maintenance framework for anomaly detection utilizing low-cost accelerometers for condition monitoring
- Implementation of wavelet transform with MEMS sensor, making the data acquisition and processing pipeline lighter while integrating multi-channel sensor fusion
- An enhanced CWT-U-Net CAE architecture is introduced for anomaly detection using scaleograms
- A practically validated approach that addresses real-world deployment challenges at the Edge

The paper is organized as follows. Section 3 provides an overview of the anomaly detection framework, including its key components and theoretical foundations. Section 4 then describes the experimental test setup and data generation process. The experimental results and framework evaluation are presented in Section 5. Finally, Section 6 summarizes the conclusions and outlines directions for future work.

3. PROPOSED ANOMALY DETECTION FRAMEWORK

3.1. Framework Overview

The framework, Figure 1, illustrates the proposed pipeline for condition monitoring and anomaly detection in rotary equipment. The data acquisition stage records multichannel vibrational data as input reflecting the operating condition of the machine using MEMS accelerometers. These raw 1D time series signals undergo pre-processing and are then segmented into discrete windows for feature extraction. Applying a Continuous Wavelet Transform (CWT) to each window produces

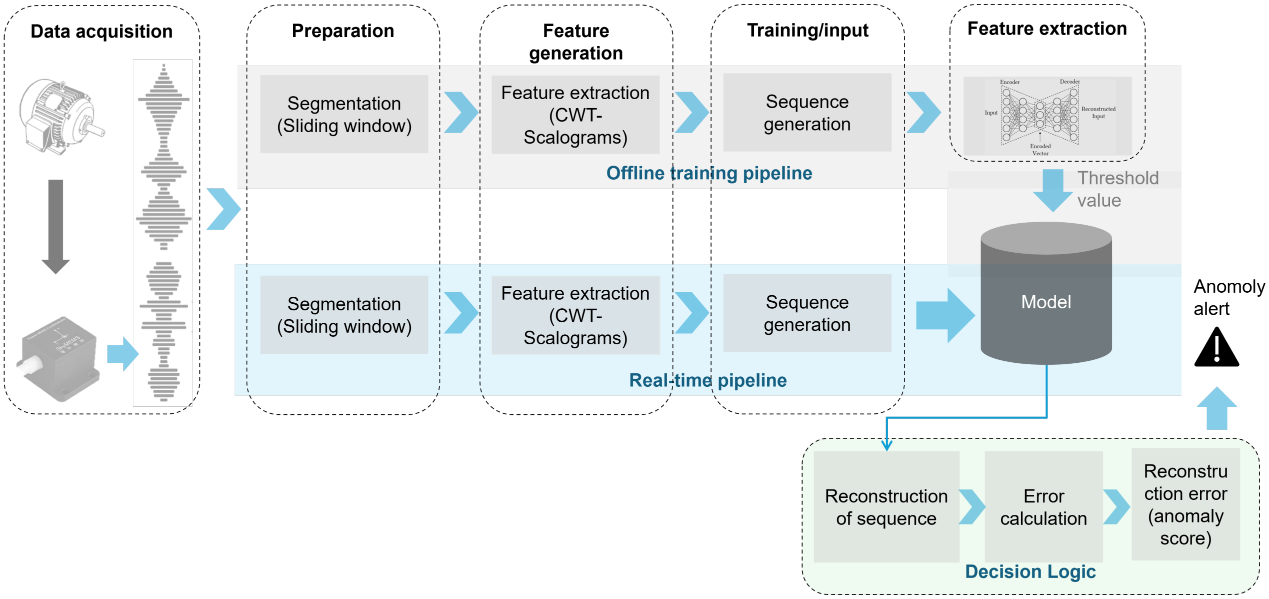

Figure 1. The proposed framework architecture consists of three main sections. The training pipelines encode raw data into an image representation using CWT and train a CNN autoencoder. The real-time pipelines assess the machine's health using the generated sequences for condition monitoring. Finally, the decision logic block computes the anomaly metric of the transformed segments, providing real-time, unsupervised anomaly detection.

a 2D image-like format, known as scalograms, which are then encoded into a multi-dimensional array suitable for deep learning input. These stacked multichannel scalograms input sequences preserve both temporal and spectral information across axes.

During the offline training pipeline, a U-Net 2D Convolutional Neural Network (CNN)-based autoencoder was trained on the generated sequences. This CWT-CAE learns to extract hierarchical features and reconstruct the scalogram with minimal loss, effectively capturing the underlying patterns for anomaly detection. The framework then evaluates an anomaly segment by calculating the reconstruction error and comparing it against the threshold value established from the training data. Finally, for the real-time in-pipeline deployment, the model is implemented on a Raspberry Pi as an Edge device for on-site inference.

3.2. Data Acquisition

A tri-axial MEMS accelerometer is adopted for vibration sensing in this study due to its compact form factor, cost efficiency, and suitability for edge deployment. Data acquisition focuses on utilizing a multi-channel signal configuration from tri-axial accelerometers. This setup provides improved sensitivity by offering spatial diversity in vibration measurements along different axes, enabling the detection of directional fault signatures and enhancing overall diagnostic accuracy. As recommended in the work done by (Guo et al., 2018), changes in rotating speed and load can significantly affect CWT calculations; therefore, vibration signals are collected with a stable rotating speed to ensure accurate CWT representations.

3.3. Preprocessing

The preprocessing pipeline begins with the removal of the DC component from the vibration signal, thereby eliminating static bias and gravity-related offsets associated with the MEMS sensor positioning. This step is crucial for vibration analysis, as the presence of DC components can introduce artifacts in the frequency-domain representation and compromise the accuracy of the continuous wavelet transform (CWT) analysis. The continuous 1-D accelerometer signals are segmented into fixed-length windows (W_l) using a sliding-window approach. This method includes an optional overlap (e.g., 50%), which serves to increase the number of training samples and improve the likelihood of capturing short-lived transients. Each window produces an array of shape (N, C), where N is the number of samples per channel and C is the number of channels.

In theory, the window length (W_l) is determined primarily by the timescales of anomalous phenomena, which are closely linked to the machine's rotational dynamics. Since fault signatures such as bearing defects, gear meshing, imbalance, and misalignment occur at frequencies tied to shaft speed (RPM), W_l is typically chosen to span one or more characteristic cycles. In practice, it is set as a function of the rotational period, balancing temporal resolution and computational cost.

The sampling frequency f_s likewise affects preprocessing:

higher f_s offers finer temporal resolution, but increases computational load for downstream continuous wavelet transform (CWT) analysis. For this study, a sampling frequency $f_s = 3.2$ kHz was chosen to capture harmonics up to 1.6 kHz, which balances the need for detail with the feasibility of edge deployment. For constrained devices, lower rates (e.g., 1 kHz) reduce memory and computation requirements but at the expense of spectral detail.

Thus, parameter selection requires balancing diagnostic fidelity with deployment feasibility. While high-resolution DAQ systems allow very high sampling rates, this work emphasizes the practicality of MEMS accelerometers at moderate rates, which limit cost and computational demand while still providing sufficient fidelity for wavelet-based feature extraction.

3.4. Continuous Wavelet Transformation

Vibration signals from rotating machines are typically non-stationary, with frequency content that changes over time due to transients and evolving spectral patterns (Yan, Gao, & Chen, 2014). Continuous wavelet transform (CWT) provides an adaptive framework for time-frequency analysis. CWT addresses the fixed-resolution issue of Short Time Fourier Transform (STFT) by decomposing a signal into wavelets, which are small, oscillatory functions localized in both time and frequency. The foundation of the wavelet transform is to scale and translate the mother wavelet $\psi(t)$, to analyze different frequencies and localized points in time within a signal, and generate the family of 'daughter' wavelets. The continuous wavelet transform of a signal $f(t)$ is mathematically defined as:

$$W_f(a,b) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{\infty} f(t)\, \psi^* \left(\frac{t-b}{a} \right) \, dt$$

Here, $a > 0$ is the scale parameter (inversely related to frequency), $b \in R$ is the translation parameter (shifting the wavelet in time), and ψ^* is the complex conjugate of the mother wavelet. The normalization term ensures constant energy across scales.

The output of the transform, $W_f(a,b)$, is visualized as a scalogram. Varying b and a visualizes the energy distribution of the vibration signal, thus resulting in a 2D plot, which is typically visualized as a heatmap representing the intensity of the signal's energy. With time on the horizontal axis and scale on the vertical axis, and color (or intensity) representing the magnitude of the wavelet complex wavelet coefficient at that specific time and scale. Unlike fixed-window methods (STFT), scalograms with their variable window lengths adapt according to the frequency content, providing optimal time-frequency trade-offs. An important preprocessing step is required before the final generation of the time-frequency image. Three common techniques are applied in this framework for preprocessing the scalogram:

1. **Logarithmic Transformation** (to compress dynamic range):
$$X_{\text{log}} = \log(1+X) \tag{1}$$
Logarithmic scaling compresses the wide dynamic range of scalograms, enhancing visibility of weak features while preventing high-intensity regions from dominating. The offset ensures stability near zero.
2. **Min-Max Normalization** (to scale into $[0,1]$ for image representation):
$$X_{\text{norm}} = \frac{X - X_{\text{min}}}{X_{\text{max}} - X_{\text{min}}} \tag{2}$$
Normalization maps the data to a fixed intensity interval, ensuring consistent contrast across windows.
3. **Resizing** (to reduce computational cost):
$$X_{\text{resized}} = \text{Resize}(X_{\text{norm}}, H, W) \tag{3}$$
To lower memory and processing demands, normalized images are resized to fixed dimensions (H, W). This preserves the essential structure of the signal representation.

For this predictive maintenance framework, the complex Morlet wavelet was selected for its ability to capture both amplitude and phase information in vibration signals (Yan et al., 2014). A key novelty of this framework is the per-axis stacking of scalograms into a 3D representation, enabling simultaneous visualization of tri-axial accelerometer signals. The preprocessed scalograms are structured as three-dimensional tensors ($H \times W \times C$), where each channel represents a signal axis.

3.5. U-Net CNN Autoencoder (CAE)

Although standard CNN autoencoders serve as effective feature extractors, they often struggle to accurately reconstruct fine-grained spatial details due to information loss during repeated downsampling and pooling operations. To address this limitation, this framework adopts a U-Net architecture, which has demonstrated superior performance in tasks requiring precise localization and reconstruction in anomaly detection (Yedurkar et al., 2023). Their U-shaped architecture with skip connections bridging encoder-decoder layers preserves high-resolution spatial features throughout the network, enabling the accurate reconstruction of textures and edges that are critical for identifying subtle, localized anomalies. This allows the decoder to recover high-frequency details and texture information that would otherwise be lost in the bottleneck layer, making it particularly suitable for analyzing industrial scalograms. Originally developed for biomedical imaging with limited data, the U-Net is particularly well-suited for prognostics applications where scarce labeled normal samples are available, as it minimizes smoothing artifacts and enhances sensitivity to deviations.

The proposed framework is illustrated in Figure 2. It enhances a standard U-Net with residual learning and regularization for improved performance and training stability. The architecture is as follows.

- The encoder path follows a convolutional design, consisting of a series of repeated downsampling blocks, each followed by a max-pooling operation, with LeakyReLU activation (ReLU), batch normalization, and spatial dropout (rate=0.3) for regularization. Deeper layers utilize residual blocks (He, Zhang, Ren, & Sun, 2016), each containing two convolutional layers with batch normalization. An identity shortcut connection bypasses these layers, projected via a 1x1 convolution when necessary to resolve dimensional mismatches. This design mitigates the vanishing gradient problem, enabling a more effective and deeper architecture.
- At its base lies the bottleneck. The encoded features are distilled at the base of the network through two successive residual blocks, forming a rich latent representation of the input.
- The decoder reconstructs the scalogram from this latent space using a sequence of transposed convolution layers for upsampling. Critically, feature maps from the encoder are concatenated to the corresponding decoder layers via skip connections at each step, preserving the fine-grained spatial information lost during downsampling. The final layer is a 1x1 convolution with a sigmoid activation function, producing a reconstructed output with pixel values normalized to [0, 1].

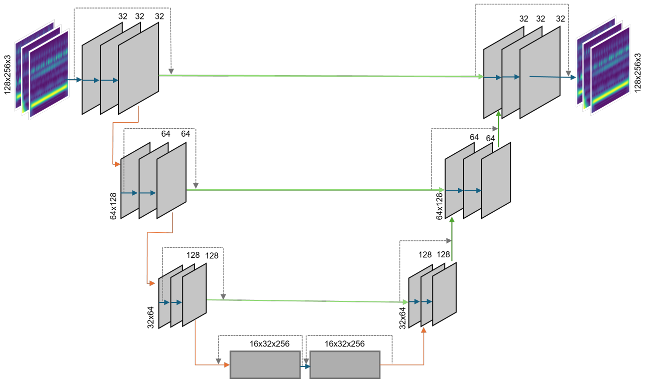

Figure 2. Proposed U-Net Architecture for anomaly detection framework

The U-Net's skip connections preserve fine spatial details across the network, while the integrated residual blocks enable stable training of a deeper network and enhance feature learning.

4. EXPERIMENTAL VALIDATION

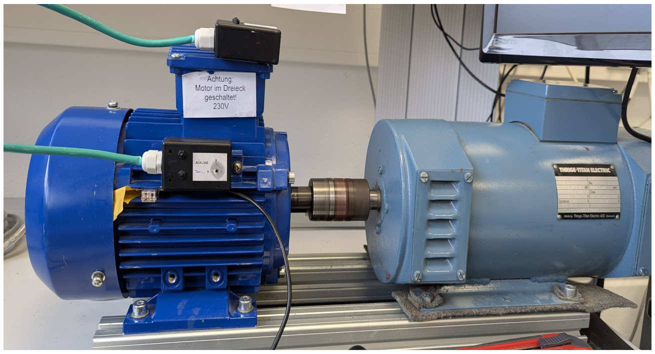

Figure 3. Experimental setup of Induction motor

Data Acquisition: To validate the proposed anomaly detection framework, a test bench was set up consisting of a three-phase induction motor mechanically coupled to a variable DC motor acting as a controllable load, as shown in Figure 3. The induction motor was run at a nominal speed of 1440 rpm under 75% load, as standard operating conditions. While vibration data were acquired using a tri-axial digital MEMS accelerometer (ADXL345, Analog Devices) under both healthy and induced fault scenarios.

Two triaxial accelerometers were screw-mounted onto the motor casing in vertical and horizontal orientations, providing redundancy in the measurement. The accelerometers' axes were aligned with the motor's axial (x), horizontal radial (y), and vertical radial (z) directions. Of the two accelerometers recorded, only the accelerometer with the highest signal-to-noise ratio was used to generate the dataset for model training. Each sensor connects to a custom data acquisition (DAQ) device based on an STM microcontroller via an individual data channel, interfaced with a Raspberry Pi 5 for storage. For this experimental setup, the signals were recorded at the highest possible sampling rate supported by the accelerometer at 3200 Hz with a ±2 g range and 13-bit resolution, satisfying the Nyquist criterion for the sensor's 1600 Hz bandwidth. A total of thirty minutes of known healthy operation of the motor was recorded, which served as the input dataset for training the autoencoder.

To assess real-time capability, different anomalies were introduced by varying mechanical load (abrupt torque changes, irregular speed fluctuations, and transient spikes) to the system, simulating realistic operational disturbances. Furthermore, the test bench facilitates the introduction of controlled faults, including rotor imbalance, rotor foot damage, misalignment, damaged coupling, and bearing defects. Because such scenarios are difficult to simulate during otherwise normal operation, a series of datasets was created by concatenating normal operating data with faulty scenarios, with the number of test datasets corresponding to the number of simulated fault conditions, discussed in the evaluation.

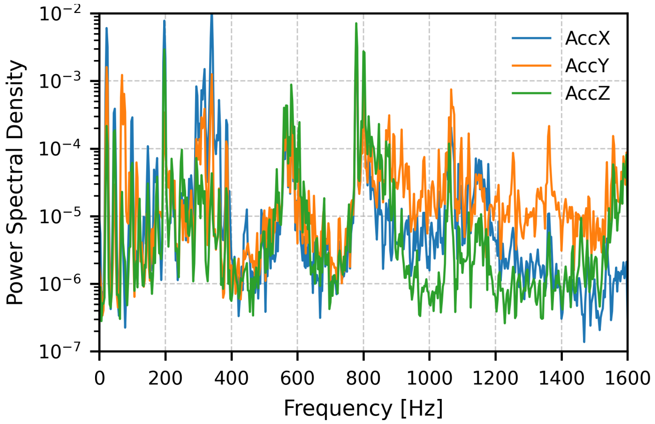

Figure 4. Power spectral density of accelerometer signals in the x, y, and z axes, obtained using Welch's method.

The raw recorded vibration signals undergo an initial preprocessing by removing the DC component. Power spectral density analysis (Figure 4) reveals that the energy is broadly distributed up to 1600 Hz bandwidth, with notable peaks occurring near 600–800 Hz for the signals. This indicates the presence of dominant vibration modes in this frequency range. These pronounced peaks are often linked to structural resonances or harmonics from rotating machinery, are valuable for tracking resonance behavior and detecting early mechanical faults.

Data Preparation: The conditioned vibration signals from each axis were simultaneously segmented in temporal order into fixed-length window length (W_l = 1 second), aligned across channels using the sliding window technique. At the sampling rate of 3200 Hz, each fixed window contains 3200 samples per channel, resulting in a two-dimensional array of shape (3200,3). This segmentation converts the continuous signals into arrays, where rows represent temporal samples and columns correspond to the x, y, and z channels, suitable for time-frequency transformation and deep learning.

Thereafter, each window was transformed into a scalogram by applying the continuous wavelet transform (CWT) using complex Morlet (cmor) as the mother wavelet. This multi-channel signal-to-scalogram conversion method is shown in Figure 5. CWT computation was performed using 128 logarithmically spaced scales, spanning the sensor's usable frequency range (10–1600 Hz), producing a coefficient matrix of size 128 × 3200 per channel. A log1p transformation was first applied to the absolute values of the complex coefficients before they were normalized to $[0, 1]$ (min–max) using global statistics computed from the healthy training dataset. This approach eliminates amplitude differences across measurement sessions while preserving relative spectral patterns relevant to anomaly detection. Normalization was performed separately for each axis prior to fusion.

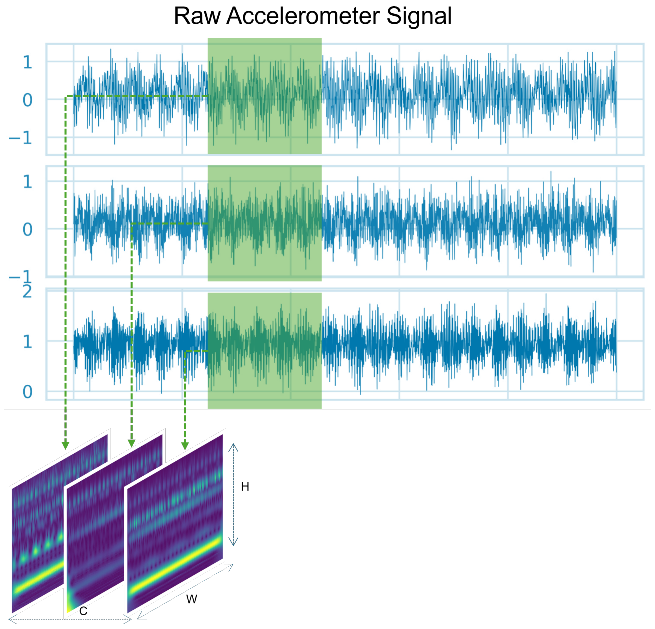

Figure 5. Data preparation pipeline from raw tri-axial vibration to multi-channel CWT scalograms.

For computational efficiency, the normalized scalograms are resized to a fixed resolution of 128×256 through bilinear interpolation, reducing input size while preserving spectral continuity across adjacent scales and time bins. This step standardizes input dimensions across all samples for CNN input, while reducing memory requirements, and facilitates batch training. Downsampling compresses the temporal axis, and the chosen resolution was found sufficient to retain transient features relevant to anomaly detection in preliminary experiments.

Subsequently, the three normalized scalograms (corresponding to the x, y, and z axes) were concatenated along the channel dimension, forming a unified tensor of shape $(128, 256, 3)$ per window. This leverages cross-axis relationships, which mimic RGB image representations and allow convolutional neural networks to exploit cross-axis correlations that may not be apparent when analyzing each axis independently. The framework's integration of directional information via channel fusion enhances the network's capacity to detect complex vibration signatures associated with machine anomalies.

4.1. Model Architecture and Training

The model was implemented according to the U-Net-based architecture with residual blocks detailed in Section 3.5. The specific layer configuration and hyperparameters are summarized in Table 1. The network was trained to minimize the mean squared error (MSE) between the input and reconstructed scalograms, using the Adam optimizer with a learning rate of 0.001 and gradient clipping at a norm of 0.5 to ensure stable training. The model was regularized during training with spatial dropout (rate=0.3) and L2 weight decay ($\lambda = 1 \times 10^{-4}$).

Encoder	
Conv2D Block 1	32 filters, 3 × 3, LeakyReLU, BN
MaxPool2D	2 × 2 stride
ResBlock 2	64 filters, 3 × 3, BN, Dropout
MaxPool2D	2 × 2 stride
ResBlock 3	128 filters, 3 × 3, BN, Dropout
MaxPool2D	2 × 2 stride
Bottleneck	
ResBlock ×2	256 filters, 3 × 3, BN, Dropout
Decoder	
Conv2DTranspose	128 filters, 2 × 2, stride 2
Skip + ResBlock	128 filters, skip from Encoder
Conv2DTranspose	64 filters, 2 × 2, stride 2
Skip + ResBlock	64 filters, skip from Encoder
Conv2DTranspose	32 filters, 2 × 2, stride 2
Skip + ResBlock	32 filters, skip from Encoder

Table 1. U-Net autoencoder architecture

The training dataset was divided into 1860 windows, of which 70% (1302 segments) were used for training and 30% (558 segments) for validation. Training was conducted with a batch size of 32. To prevent overfitting, early stopping with a patience of 10 epochs (restoring the best weights) and a learning rate reduction by a factor of 0.5 after 5 epochs of validation loss plateau were applied. This configuration ensured stable convergence and good generalization.The testing dataset, which remained unseen during training and validation, comprised 1507 samples in total, including 386 normal windows and 1121 fault windows randomly selected from six distinct fault conditions. To ensure fair evaluation, faulty samples were evenly distributed across these fault types, enabling consistent assessment of the model's anomaly detection capability across diverse failure modes.

5. EXPERIMENTAL RESULTS AND ANALYSIS

In this section, we present the results organized into two subsections: one discussing the anomaly detection performance on the induction machine test rig, and the other evaluating the efficacy of edge deployment with the anomaly detection framework.

5.1. Induction machine anomaly detection performance

The anomaly detection capability of the proposed U-Net autoencoder-based framework was assessed based on its reconstruction error on unseen test data. Consequently, the Mean Squared Error (MSE) between the original and reconstructed scalogram windows was used as a robust anomaly score, as it quantifies the deviation from the learned healthy state. Higher MSE values, specifically those exceeding a predefined threshold, directly indicated deviations from the learned healthy state, signaling a potential anomalous state.

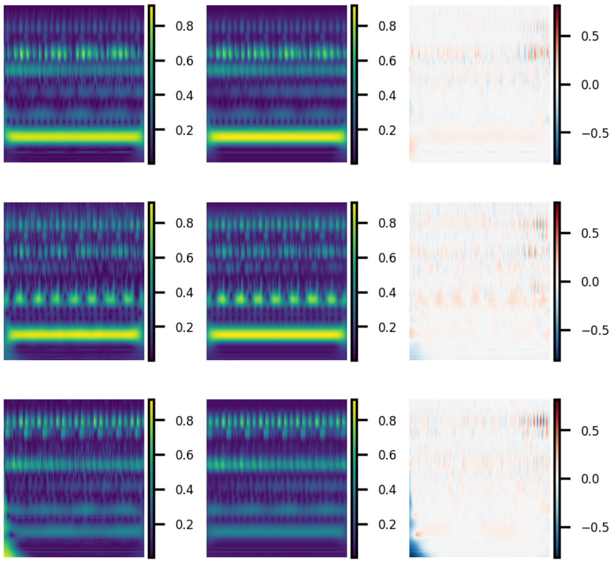

Figure 6. Scalogram comparison for normal operation: original (left), reconstructed (middle), and difference (right) for x, y, z axes

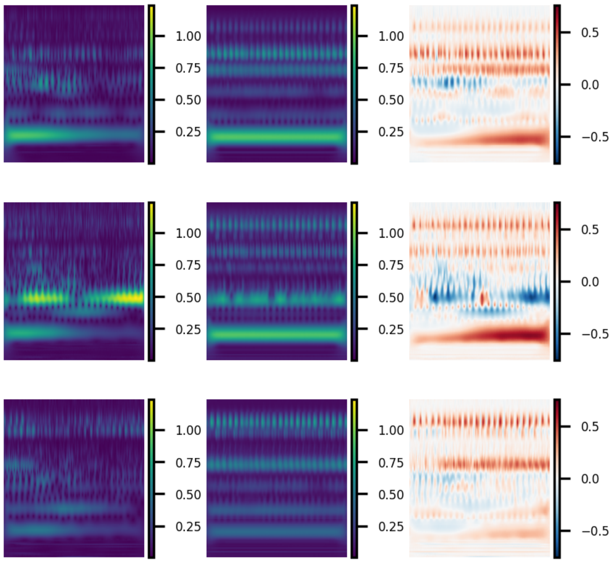

Figure 7. Scalogram comparison for an anomalous segment: original (left), reconstructed (middle), and difference (right) for x, y, z axes

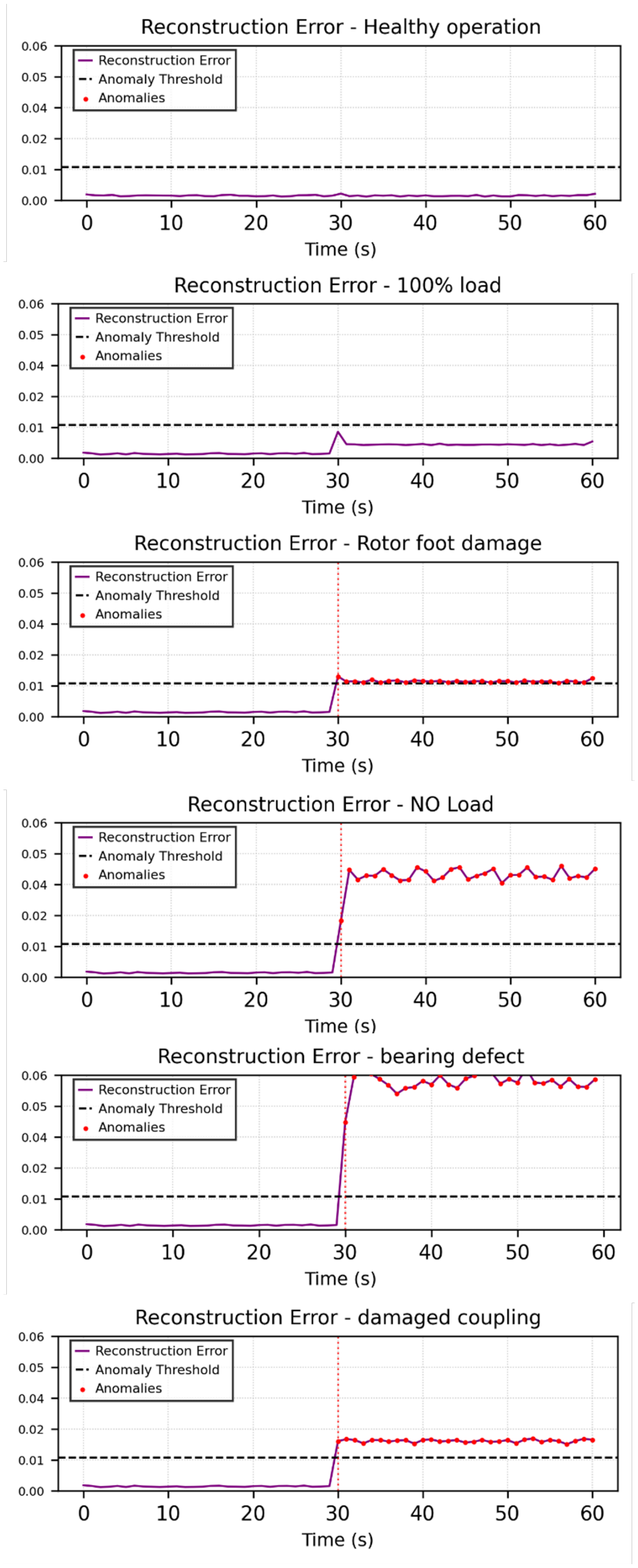

Figure 8. Reconstruction error (MSE) over time under normal operation and different induced faults

Post-training, the model achieved a mean reconstruction error of 0.001725 on the training set and 0.001832 on the validation set. The narrow gap between these values indicates minimal overfitting and confirms the model's ability to capture normal operational patterns, establishing a reliable baseline for unsupervised anomaly detection.

A visual comparison in Figure 6 illustrates the reconstruction quality, where the output scalograms closely match the original inputs across all three accelerometer axes within a representative time window. This close correspondence demonstrates that the model successfully preserves essential time-frequency features and captures the fine-grained structure of healthy operation. In contrast, Figure 7 highlights the discrepancy between original and reconstructed scalograms when anomalous inputs are presented. Here, the difference between the two is explicitly shown; in anomalous cases, this difference becomes more pronounced, making the deviations clearly visible. The threshold for anomaly detection was set using the mean plus three standard deviations ($\mu + 3\sigma$) of reconstruction errors derived from a validation subset of healthy data.

To further evaluate robustness, Figure 8 illustrates, different fault scenarios concatenated with 30 seconds of healthy data, which plots the anomaly score (MSE) over time. During normal operation, the score remained below threshold. In these sequences, the reconstruction error displays a sudden and sustained spike precisely at the onset of faults, showing that the model can reliably distinguish faulty conditions without supervised labels. The reconstruction of healthy segments serves as a stable baseline, while the absence of false alarms in the 100% load scenario, previously seen during training demonstrates that the model does not misclassify familiar operating states as anomalies.

Models	**P/%**	**R/%**	**F1/%**	**ROC AUC/%**
U-Net-CAE (CWT)	98.42	100.00	99.20	100.00
2D-CAE (CWT)	98.67	92.95	95.73	94.17
1D-CAE (Raw)	97.56	90.00	92.77	95.85

Table 2. Comparison of experimental results

To contextualize these findings, we performed a comparative analysis against two common unsupervised baselines: a standard 1D Convolutional Autoencoder (1D-CAE) operating on the raw time-series data, and a standard 2D-CAE (without U-Net skip connections) operating on the same CWT scalograms. The experimental results, as shown in Table 2, compare the performance of the anomaly detection models. The U-Net–CAE (CWT) demonstrates clear superiority, achieving perfect recall (100%) and ROC AUC (100%), while also attaining the highest F1-score (99.20%). This highlights the advantage of its architecture, where skip connections effectively preserve high-resolution features from the CWT, which is critical for reconstructing the complex patterns of healthy operation and detecting subtle deviations.

Taken together, these results validate the effectiveness of the proposed anomaly detection framework. The low reconstruction error on healthy data confirms its ability to learn the complex representation of normal operation, while the consistent error escalation and clear visual separability under faults underline its sensitivity to abnormal patterns. These findings position the model as a strong candidate for industrial anomaly detection, offering both accuracy and robustness in real-world settings.

5.2. Real-Time Edge Deployment and Monitoring System

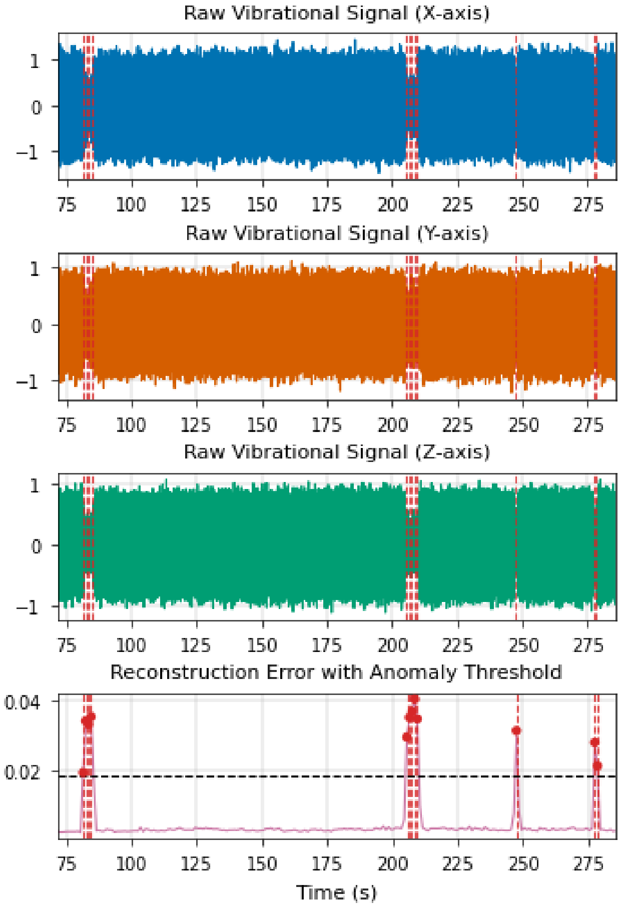

Figure 9. Raw vibration signal vs. reconstruction error for test data with induced anomalies. Spikes correspond to detected anomalies.

For real-time deployment on an edge device, the number of scales in the Continuous Wavelet Transform (CWT) was reduced from 128 to 64, significantly lowering the computational load during feature extraction and model parameters. Within the framework, the CWT remains the most resource-intensive step. Its implementation was carried out on a Raspberry Pi 5, equipped with a high-speed SSD to support rapid data transfer from the acquisition system and the execution of the CWT algorithm.

The deployment architecture also integrated a lightweight web interface built with Flask[1], hosted directly on the Raspberry Pi. This interface enabled real-time visualization of diagnostic metrics, including reconstruction error and anomaly alerts allowing operators to monitor system health remotely. Benchmarking showed that the revised CWT computation across 64 scales required an average of 705 milliseconds per cycle, almost halving the 1395 milliseconds observed with the original 128-scale configuration. Subsequent preprocessing steps, such as normalization and resizing to match the autoencoder's input, added only 2.3 milliseconds per instance. The inference stage, involving a compact 31 MB autoencoder and threshold-based anomaly scoring, averaged 177 milliseconds per cycle and ran effectively on the Pi's CPU without GPU support. Taken together, the end-to-end pipeline from raw acquisition to anomaly classification operated with an average latency of 900 milliseconds, comfortably within the sub-second response times demanded in industrial edge applications. During this end-to-end inference cycle, the average CPU load on the Raspberry Pi 5 remained at approximately 27-40%, and the entire process consumed 776-900 MB of system RAM. These metrics confirm the feasibility of deploying the framework on such constrained devices without overwhelming the system, leaving resources for parallel tasks.

The system's ability to separate normal from abnormal operation was assessed through semi-supervised testing. To this end, anomalies were deliberately introduced in the form of short-duration load spikes, gradual torque shifts, and other fault states. Figure 9 shows the reconstruction error profile over one such test sequence, with pronounced spikes aligning closely with the onset of induced anomalies. The dataset contained six mechanically induced load spikes, during which the reconstruction error exceeded the anomaly threshold by 30–45%. These events were consistently detected within 2 seconds of their occurrence, confirming the framework's responsiveness. The performance gains can be attributed to the multichannel design, which processes vibration and other sensor signals simultaneously to improve sensitivity across fault modes.

6. CONCLUSION AND FUTURE WORK

This study presented an anomaly detection framework tailored for rotary equipment, with a focus on induction motors. By combining continuous wavelet transform (CWT) with a residual U-Net autoencoder, the approach leverages scalogram representations of tri-axial vibration signals to capture both temporal and spatial features across machine axes. The

[1] `https://palletsprojects.com/projects/flask/`

reconstruction error provided a robust anomaly score, allowing subtle deviations from normal operation to be identified without the need for supervised labels.

Experimental validation confirmed the framework's ability to detect a range of induced fault scenarios with high sensitivity while retaining robustness under known operating conditions. The reduction of CWT scales from 128 to 64, coupled with an optimized autoencoder, enabled real-time deployment on a Raspberry Pi 5 with sub-second inference latency, demonstrating its suitability for industrial edge applications.

Despite these promising results, we acknowledge several limitations inherent in this study. The experimental validation was conducted exclusively on a single induction motor test bench under controlled laboratory conditions. The vibration characteristics can differ significantly across various types of rotary equipment, and the model's generalization to other machines has not yet been verified. Furthermore, the experiments did not cover the full range of complexities found in real industrial environments, such as high environmental noise or highly dynamic operating conditions. The unavailability of a suitable, publicly available benchmark dataset for our specific multi-axis MEMS sensor configuration also constrained our ability to perform a broader comparative analysis against other published methods.

Future work will be directed at addressing these limitations. A primary goal is to validate the framework's robustness by applying it to different equipment types and, where possible, under the variable speed and load conditions common in real-world scenarios. Additionally, to improve adaptability in diverse industrial settings, we will explore more dynamic anomaly thresholding techniques (e.g., adaptive, statistics-based thresholds) as an alternative to the current static $\mu + 3\sigma$ rule, aiming to reduce false alarms and enhance sensitivity.

ACKNOWLEDGMENT

The work described in this paper has been carried out in the context of the project Industry 4.0 ACE, funded under contract KK5538701 FG3 by the German Federal Ministry for Economic Affairs and Climate Action. We thank aia automations institut GmbH and MSF-Vathauer Antriebstechnik GmbH & Co KG for kind collaboration and support.

REFERENCES

Benhanifia, A., Cheikh, Z. B., Oliveira, P. M., Valente, A., & Lima, J. (2025). Systematic review of predictive maintenance practices in the manufacturing sector. *Intelligent Systems with Applications*, 200501.

Bernitsas, E., & Kourkoutos-Ardavanis, N. (2021). The emerging role of scalogram-based convolutional neural networks in epileptic seizure detection. *Brain Sciences*, *11*(11), 1424. doi: 10.3390/brainsci11111424

Chen, J., & Wu, H. (2024). Convolutional autoencoder-based anomaly detection using MEMS vibration sensors. *IEEE Sensors Journal*, *24*(6), 10245–10256.

Chou, H., & Wang, Y. (2025). YOLO-based fault detection using time–frequency representations of vibration signals. In *Proceedings of the IEEE international conference on prognostics and health management (phm).*

Guo, S., Yang, T., Gao, W., & Zhang, C. (2018). A novel fault diagnosis method for rotating machinery based on a convolutional neural network. *Sensors*, *18*(5), 1429.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)* (pp. 770–778).

Huang, W., & Sun, L. (2025). Engineering-driven fault diagnosis using vibration and current signal fusion. *Mechanical Systems and Signal Processing*, *200*, 110945.

Jain, M., Patel, V., & Raj, T. (2020). Vibration monitoring of CNC machinery using mems sensors. *Journal of Vibroengineering*, *22*(4), 899–910. doi: 10.21595/jve.2020.21125

Kanungo, P. (2025). Edge computing in healthcare: Real-time patient monitoring systems. *World Journal of Advanced Engineering Technology and Sciences*, *15*(1), 001–009. doi: 10.30574/wjaets.2025.15.1.0168

Li, C., & Zhou, Y. (2023). Multi-sensor fusion with autoencoder models for industrial anomaly detection. *IEEE Transactions on Industrial Informatics*, *19*(4), 5640–5651.

Malviya, A., & Singh, P. (2022). Edge computing for predictive maintenance: FPGA-based anomaly detection with lightweight autoencoders. In *Proceedings of the IEEE international conference on edge computing* (pp. 45–52).

Pang, G., Shen, C., Cao, L., & van den Hengel, A. (2021). Deep learning for anomaly detection: A review. *ACM Computing Surveys*, *54*(2), 1–38. doi: 10.1145/3439950

Park, J., & Kim, D. (2023). Multi-head attention networks for weak fault diagnosis in industrial motors. *Neural Computing and Applications*, *35*, 15187–15199.

Tang, Y., & Han, J. (2024). Anomaly detection in rotating machinery using residual CNNs with temporal modeling. *Mechanical Systems and Signal Processing*, *190*, 110289.

Ventricci, A., & Marino, S. (2024). Motor fault classification using low-cost MEMS accelerometers and deep learning. *Sensors*, *24*(3), 865.

Yan, R., Gao, R. X., & Chen, X. (2014). Wavelets for fault diagnosis of rotary machines: A review with applications. *Signal Processing*, *96*, 1–15.

Yang, B., & Li, X. (2021). Bearing fault diagnosis under variable speed conditions using CNN–LSTM networks. *IEEE Access*, *9*, 97890–97902.

Yedurkar, D. P., et al. (2023). Early fault diagnosis of rolling bearing based on threshold acquisition U-Net. *Machines*, *11*(1), 119. doi: 10.3390/machines11010119

Zhang, L., & Chen, M. (2023). Multi-scale convolutional networks with CWT for intelligent fault diagnosis of bearings. *Mechanical Systems and Signal Processing*, *185*, 109735.

Zheng, H., & Xu, Y. (2023). Correlation-aware feature learning for vibration-based anomaly detection. *IEEE Transactions on Industrial Electronics*, *70*(7), 7412–7421.