

From Noise to Knowledge: A Comparative Study of Acoustic Anomaly Detection Models in Pumped-storage Hydropower Plants

Karim Khamaisi
University of St. Gallen
Dornbirn, Austria
karim.khamaisi@unisg.ch

Nicolas Keller
University of St. Gallen
St. Gallen, Switzerland
nicolas.keller@student.unisg.ch

Stefan Krummenacher
University of St. Gallen
St. Gallen, Switzerland
stefan.krummenacher@student.unisg.ch

Valentin Huber
University of St. Gallen
St. Gallen, Switzerland
valentin.huber@student.unisg.ch

Bernhard Fässler
illwerke vkw AG
Bregenz, Austria
bernhard.faessler@illwerkevkw.at

Bruno Rodrigues
University of St. Gallen
Dornbirn, Austria
bruno.rodrigues@unisg.ch

Abstract

In the context of industrial factories and energy producers, unplanned outages are highly costly and difficult to service. However, existing acoustic-anomaly detection studies largely rely on generic industrial or synthetic datasets, with few focused on hydropower plants due to limited access. This paper presents a comparative analysis of acoustic-based anomaly detection methods, as a way to improve predictive maintenance in hydropower plants. We address key challenges in the acoustic preprocessing under highly noisy conditions before extracting time- and frequency-domain features. Then, we benchmark three machine learning models: LSTM AE, K-Means, and OC-SVM, which are tested on two real-world datasets from the Rodundwerk II pumped-storage plant in Austria, one with induced anomalies and one with real-world conditions. The One-Class SVM achieved the best trade-off of accuracy (ROC AUC 0.997–0.998) and minimal training time, while the LSTM autoencoder delivered strong detection (ROC AUC 0.997–0.999) at the expense of higher computational cost.

CCS Concepts

• **Computing methodologies** → **Machine learning**; • **Applied computing** → *Industry and manufacturing*.

Keywords

Industrial Machinery, Acoustic Anomaly Detection, Predictive Maintenance, Machine learning

1 Introduction

Maintenance is a critical function in industrial environments due to the significant operational and financial impacts associated with equipment failure. Recent estimates highlight that unplanned downtime costs industries billions annually, with individual facilities losing millions per event [27]. Such unexpected interruptions are particularly consequential in energy production facilities like pumped-storage hydropower plants, where operational continuity is essential to maintain grid stability and minimize economic impacts.

Hydropower plant equipment, including turbines and generators, is central to plant operations. Failures in these mechanical

parts can result in costly downtime. Downtime costs are highly variable, primarily depending on fluctuating market energy prices. For instance, the Rodundwerk II pumped-storage (*cf.* Section 2) hydropower plant in Vorarlberg, Austria, has a full-load capacity of 295 MW, translating to 295 MWh of electricity production per hour. With a wholesale price of 97.30 Euro/MWh [9] (as of February, 2025), downtime can cost over 28,000 Euros per hour [12]. Moreover, since pumped-storage hydropower plants typically function as peak-load power plants, selling electricity predominantly at times of high market prices, actual downtime costs can significantly exceed this estimate.

Predictive maintenance (PdM) proactively anticipates equipment failures, enabling maintenance timing that avoids both unnecessary downtime and catastrophic failures [21, 32]. Among various PdM approaches, acoustic-based anomaly detection stands out for its non-intrusive nature, cost-effectiveness, and sensitivity to early-stage mechanical defects [7, 13, 15]. Acoustic signals carry vital diagnostic information, reflecting subtle deviations in machinery operations that precede potential failures. However, acoustic-anomaly detection methods often face challenges in industrial environments characterized by high noise levels, complex acoustic patterns, and limited labeled anomaly data [17, 28].

Although acoustic anomaly detection has been extensively studied, existing literature predominantly relies on generic industrial or synthetic datasets [2, 7, 8, 24], with limited representation from actual hydropower plant operations. This gap exists mostly due to restricted access to critical infrastructure, such as pumped-storage hydropower facilities, whose acoustic characteristics significantly differ from standard industrial machinery. Consequently, existing models [15, 17, 28] trained on generic data often fail to generalize effectively to the acoustic conditions and anomalies characteristic of hydropower plants.

We provide systematic comparative analysis of acoustic anomaly detection methods explicitly applied to a pumped-storage hydropower plant, Rodundwerk II in Austria managed and operated by illwerke vkw AG [12]. **Contributions** are summarized as follows:

- We present an acoustic preprocessing pipeline tailored to mitigate noise and handle the challenging acoustic environment of hydropower plants.
- Benchmark three prominent machine learning models: LSTM Autoencoder (LSTM AE), K-Means clustering, and One-Class SVM (OC-SVM), using two distinct real-world datasets



This work is licensed under a Creative Commons Attribution 4.0 International License.

Table 1: Selected Related Work on Acoustic Anomaly-detection

Study	Domain / Data	Feature Rep.	Model(s)	Key Insight
Müller et al. (2020) [16]	Factory machinery	MelSpec + CNN feats	IF, GMM, OC-SVM, AE	CNN features beat CAE in noise
Meire et al. (2019) [20]	Industrial bench setup	MFCC, MelSpec	OC-SVM, AE	HW-friendly AEs for edge devices
Bayram et al. (2021) [2]	Chemical process plant	MelSpec	Conv-LSTM-AE, CAE	Sequential AEs allow real-time
Ferraro et al. (2023) [11]	Multi-factory lines	MelSpec	LSTM-AE, CNN-AE	Sliding-window edge inference
Coelho et al. (2022) [4]	Factory & vehicle	MFEC, MelSpec	Dense/CNN/LSTM AEs	Depth vs. noise-robustness study
Duman et al. (2020) [8]	Industrial plant	MelSpec	CAE	Lightweight CAE for on-site use
Ahn et al. (2021) [1]	Vehicle acoustics	Raw audio	SVM, K-Means, CNN	Multi-mic array boosts recall
Tagawa et al. (2021) [28]	Rotating machinery	Spectral coeff.	SVM, Auto-corr.	Spectral coeff. key for bearings
Purohit et al. (2019) [24]	MIMII (public)	Log-MelSpec	VAE, AE	Widely used open benchmark

acquired from Rodundwerk II: one with induced anomalies under controlled conditions, and another featuring naturally occurring operational anomalies.

- We discuss strengths and weaknesses of each model, providing guidelines for practitioners on selecting appropriate methods based on operational constraints.

The remainder of this paper is organized as follows: Section 2 reviews existing literature on acoustic anomaly detection, focusing on both traditional signal processing and machine learning approaches. Section 3 presents the proposed framework, detailing its preprocessing pipeline and model architectures. Section 4 discusses experimental results and provides a comparative evaluation real-world datasets. Finally, Section 5 offers concluding remarks and outlines directions for future research.

2 Background and Related Work

Pumped storage hydropower plants (PSHP). Traditional hydropower plants generate electricity using water stored behind a dam, flowing downstream to drive turbines [14]. In contrast, PSHPs utilize mountainous topography with two water reservoirs [22]. During low demand, water is pumped from the lower reservoir to the upper reservoir using surplus energy; during high demand, stored water is released to generate electricity.

Unlike traditional hydropower plants, PSHPs require turbines to operate in dual mode (generation and pumping), leading to frequent operational changes that increase stresses, vibrations, and dynamic loads. These factors accelerate turbine fatigue and wear, causing issues like pump cavitation [31]. Anomaly detection facilitates early fault detection to prevent costly downtime [7, 13], with acoustic-based methods leveraging non-intrusive signals to capture subtle mechanical deviations prior to critical failures [17, 28].

Feature extraction techniques. Robust feature extraction is essential for accurately representing acoustic signals. Traditional methods like Short-Time Fourier Transform (STFT) and Mel-scale transformations convert raw audio signals into time-frequency spectrograms [2, 32]. Mel-spectrograms are widely adopted due to their perceptual alignment with human auditory systems [20], while Mel-Frequency Cepstral Coefficients (MFCCs) capture critical temporal and spectral characteristics that enhance anomaly detection [2, 4].

Machine learning models. Acoustic anomaly detection typically employs machine learning (ML) methods, broadly categorized into unsupervised and deep learning approaches. Density-based

models like Local Outlier Factor (LOF) identify anomalies based on local data densities without extensive labeled data [3, 29]. Clustering methods like K-Means offer computational simplicity but struggle with overlapping acoustic patterns [1]. Deep learning methods like Autoencoders (AEs) detect anomalies through reconstruction errors, proving effective in complex environments [8, 11], though they require substantial computational resources often unavailable in hydropower plants.

Table 1 provides a concise comparative overview of recent relevant works, highlighting feature extraction methods, ML models used, and key objectives. Existing work relies mainly on generic or synthetic datasets, with few incorporating real-world data from specific applications such as hydroelectric plants [2, 24]. In terms of feature extraction, Mel-spectrograms are the most commonly used technique [2, 4, 8, 11, 16, 20], followed by MFCCs and STFT [20]. This highlights the importance and effectiveness of converting raw audio signals into meaningful inputs for ML models.

The most used ML models are AEs, which are highly adaptable to specific audio signals and effectively detect abnormal sounds deviating from learned baselines. Various AE architectures appear across studies: [4] explores three different architectures, [11] compares LSTM-AE and CNN-AE models, while [2] utilizes Conv-LSTMAE and CAE. This diversity highlights standardization limitations, as no single model consistently outperforms others across all scenarios.

3 Proposed Method

Figure 2 illustrates the proposed acoustic anomaly detection pipeline. The pipeline addresses the challenges inherent in handling acoustic data from PSHP hydropower plants. It consists of four main stages, which are detailed in next subsections: Data Acquisition, Preprocessing, Model Benchmarking, and Evaluation Metrics.

The design of this pipeline reflects key considerations for handling real-world industrial acoustic-data, including noise resilience, efficient anomaly representation, and scalable deployment. The pipeline comprises four main stages:

- (1) **Data Acquisition:** audio recordings of machine operations in controlled and real-world environments.
- (2) **Preprocessing:** transformation of raw audio into normalized Mel-spectrograms.
- (3) **Model Training:** benchmarking and evaluation of five selected models.



Figure 1: Data Acquisition in PSHP Rodundwerk II: (a) audio recordings with induced anomalies (hammer and a shovel). (b) microphone stands in 4 corners capturing airborne audio. (c) shaft of the reversible francis-pump turbine; microphone stands were installed in each corner of the room.

- (4) **Evaluation Metrics:** identification of anomalies using reconstruction error and comparing with the benchmarks.

3.1 Data Acquisition

Two distinct datasets were produced on-site to benchmark models for the acoustic anomaly detection. This was achieved by recording the machinery in real operating modes and with induced anomalies, splitting them into separate .wav files, used for data preprocessing.

The dataset with **induced anomalies** was recorded on-site to help in the first stage of fine-tuning models (cf Figure 1 - a). We used a hammer striking a shovel to resemble real-world structural issues yet remain safe and repeatable. These synthetic knocks allow replication of realistic fault phenomena and thus form the basis for testing the framework under near-real-world conditions while preserving consistency in labeling.

The **real dataset** was collected from an operational pumped storage hydropower plant under strict operational constraints that

inherently limit data collection scope (in which the authors are thankful to the Rodundwerk’s technical team). Access to such facilities is restricted due to safety regulations, security protocols, and the need for coordination with plant operators.

3.2 Preprocessing

Prior to modelling, raw audio signals must be converted into representations that are both stable and discriminative in the context of acoustic anomaly detection. We use a Short-Time Fourier Transform (STFT) followed by a Mel-spectrogram conversion and a frame-based segmentation procedure. These steps are essential to transform the original time domain signals into a form more suited to machine learning methods. In addition, we used a combination of **noise reduction** to provide cleaner audio features and reduce the influence of noisy backgrounds; **normalization** using root mean square through average amplitude levels to ensure consistency while preserving feature stability; and **segmentation** splitting audio recordings into smaller, overlapping frames before being input to the ML model, facilitating feature extraction and maintaining balanced datasets [26].

To extract audio features, we first apply a standard discrete *Short-Time Fourier Transform (STFT)* [5, 19, 23] from the *librosa* library with a 1024-sample window and a hop length of 512. By decomposing the audio into short, overlapping time segments, we obtain local frequency-domain information, enabling the detection of short-duration events (e.g., knocks or impacts). This time-frequency representation is critical for capturing the subtle temporal patterns that characterize anomalies in industrial or mechanical systems. Mathematically, the STFT is defined in Equation 1, where we sum over windowed signal frames and apply the discrete Fourier transform.

$$STFT(n, k) = \sum_{m=0}^{L-1} x(n+m) w(m) e^{-j 2\pi \frac{k}{N} m} \quad (1)$$

From the STFT magnitude, we derive a 128-band *Mel-spectrogram* [6, 18, 19]. The Mel scale aligns more closely with human auditory

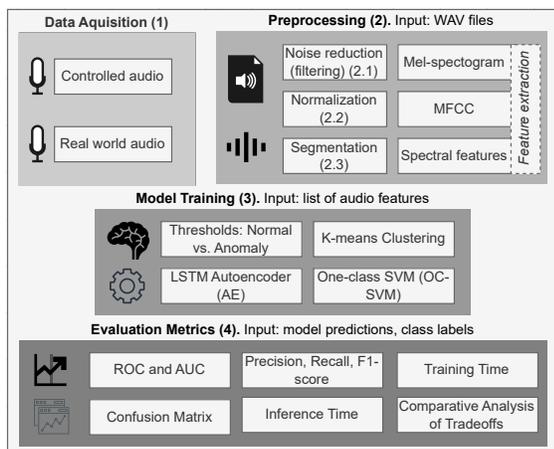


Figure 2: Overview of the acoustic anomaly detection framework

perception, ensuring that frequency regions vital to detecting anomalies (e.g., sudden high-frequency bursts) are not overshadowed by less relevant areas of the spectrum. By compressing the frequency axis in this perceptual manner, the model is better able to focus on critical frequency components that differentiate normal from abnormal operations.

$$M(m, n) = \sum_{k=0}^{K-1} H_m(k) |STFT(n, k)|^2 \quad (2)$$

In Equation 2, each Mel bin is computed by summing the STFT power according to a triangular filter bank H_m .

$$M_{DB}(m, n) = 10 \log_{10} \left(\frac{M(m, n)}{\max_{m, n}(M(m, n))} \right) \quad (3)$$

The Mel-spectrogram is converted to decibels using Equation 3, normalizing each cell by the global maximum and applying $10 \log_{10}$. Then, each Mel-spectrogram is min-max normalized [25] to a [0, 1] range as shown in Equation 4, which subtracts the minimum and divides by the overall range:

$$x_{\text{norm}} = \frac{x - \min(X)}{\max(X) - \min(X)} \quad (4)$$

This standardization step makes training more robust by reducing amplitude-related variability and ensuring uniform feature scales across the dataset. As a result, the model is not biased by unusually loud or soft recordings, improving generalization to new, unseen data.

```

1 time_per_frame = 0.6
2 hop_ratio      = 0.2
3 hop_length     = 512
4
5 def generate_mel_spectrogram(audio_path):
6     audio, sr = librosa.load(audio_path, sr=None)
7     stft = librosa.stft(audio, n_fft=1024, hop_length=
8         ↪ hop_length)
9     mel = librosa.feature.melspectrogram(S=np.abs(
10        ↪ stft)**2, sr=sr, n_mels=128)
11    mel_db = librosa.power_to_db(mel, ref=np.max)
12    mel_db_norm = (mel_db - mel_db.min()) / (mel_db.
13        ↪ max() - mel_db.min())
14    return mel_db_norm, sr
    
```

Listing 1: Code to Generate and Normalize Mel-Spectrograms

The Python code in Listing 1 illustrates how the first steps are implemented in practice, including loading the audio data, applying an STFT, generating and normalizing Mel-spectrograms. After computing the Mel-spectrogram, we segment each spectrogram into overlapping frames. In our setup:

- **Windowing:** A frame length of 0.512 s is chosen to capture transient acoustic events while retaining sufficient frequency resolution.
- **Hop Ratio:** A 20% overlap is maintained between consecutive frames (i.e., a hop ratio of 0.2). This partial overlap preserves temporal continuity and guards against losing information at frame boundaries—particularly important for detecting short, impulsive anomalies.

By transforming each spectrogram into a series of frames, we produce localized snapshots of the acoustic signal. These frames serve as input units for both training and evaluation, making the subsequent anomaly detection task more fine-grained and sensitive to local transient events.

```

1 def generate_frames(mel_spectrogram, frame_size,
2     ↪ hop_size):
3     num_frames = (mel_spectrogram.shape[1] -
4         ↪ frame_size) // hop_size + 1
5     frames = np.zeros((num_frames, mel_spectrogram.
6         ↪ shape[0], frame_size))
7     for i in range(num_frames):
8         start = i * hop_size
9         frames[i] = mel_spectrogram[:, start:start +
10            ↪ frame_size]
11    return frames
    
```

Listing 2: Code to Generate Overlapping Frames from a Mel-Spectrogram

The Python code in Listing 2 illustrates how the generation of frames is implemented in practice. Each set of frames is then saved in .npy format for subsequent training and evaluation. By separating the raw audio loading from the downstream tasks, we ensure a clear workflow in which all models operate on the same preprocessed feature set.

3.3 Anomaly Detection Models

We selected three machine learning models due to their distinct methodological strengths, computational efficiency, and suitability to handle the complexities inherent to acoustic anomaly detection in hydropower:

- **K-Means clustering:** has a minimal computational overhead and intuitive clustering based approach [1], making it suitable for real-time applications with limited computational resources.
- **Long Short-Term Memory Autoencoder (LSTM AE):** a deep learning model capable of capturing temporal dependencies in sequential acoustic signals, excelling at identifying deviations from normal behaviors, as can be seen in [10] where accuracy rates of 99% were achieved when detecting anomalies.
- **One-Class Support Vector Machine (OC-SVM):** a model that learns a tight decision boundary around normal (non-anomalous) data, flagging deviations from the decision boundary as anomalies [16, 20].

In addition, we compute each method’s outlier score on the validation set (normal-only) to pick a percentile-based **anomaly threshold**. We evaluated multiple thresholds between the 5th and 95th percentiles, and selected the ones that maximizes the F1-score [30].

4 Experimental Evaluation

This Section presents results of the proposed acoustic anomaly detection framework, evaluated on audio data collected with synthetic, induced anomalies, and in real operation mode. All audio was captured with a *Samson Meteor MIC* at a 16 kHz sampling rate and stored in .wav format for consistency. We examine three ML models representing key industrial trade-offs: fast but simple methods, balanced approaches, and high-performance but resource-heavy solutions.

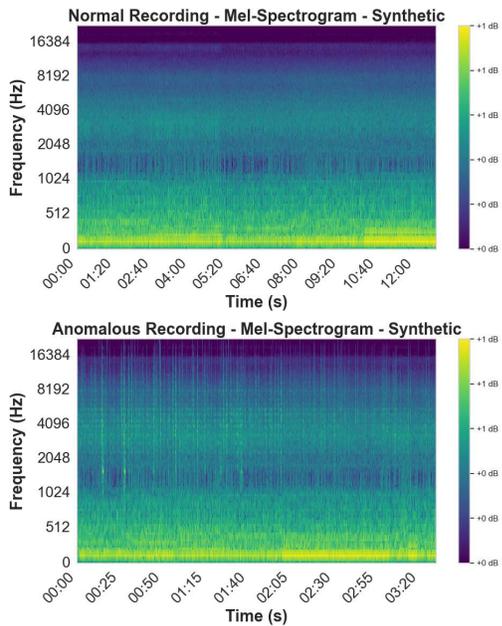


Figure 3: Mel-spectrograms for synthetic dataset: normal (top) vs. anomalous (bottom).

4.1 Induced Synthetic Anomalies

The first dataset was recorded in an operational industrial machine and comprises:

- **Normal operation (12 minutes 46 seconds):** Equipment functioning under routine conditions, reflecting regular machine activity.
- **Induced events (3 minutes 29 seconds):** Artificial mechanical faults introduced via controlled impacts from a hammer striking a shovel, selected to resemble real-world structural issues yet remain safe and repeatable.

These synthetic knocks allow replication of realistic fault phenomena and thus form the basis for testing the framework under near-real-world conditions while preserving consistency in labeling.

Evaluation metrics. Table 2 presents the **evaluation metrics** for the ML models on the synthetic anomaly dataset with focus on the K-Means, OC-SVM, and LSTM-AE. All models achieved near-perfect results, as reflected in the ROC AUC, precision, recall, and F1-scores. The **confusion matrices** in Figure 4 further highlight the strong classification performance. Overall, the LSTM-AE provides the most balanced outcome, maintaining both a low false positive rate (3.0%) and high anomaly recall (99.8%), making it well suited for deployment where precision and trust in anomaly alarms are essential. However, K-Means offers the fastest training and inference times, with performance metrics nearly matching those of LSTM-AE, and achieves the lowest overall misclassification count.

The **Mel-spectrograms** of the synthetic anomaly and normal datasets are presented in Figure 3. The normal recording (top) shows smooth, constant energy distribution below 2000 HZ. The anomalous Mel-spectrogram (bottom) reveals multiple energy bursts with

vertical stripes representing induced anomalies. It shows short high-energy bands in a high frequency range of above 2000 HZ, which correspond to the knocking sounds caused by a hammer hitting a shovel. Therefore, the normal operation noise is clearly visually distinguishable from the anomalous sound.

Figure 5 shows the **MFCC** coefficients for both recordings. The normal recording (top) exhibits relatively uniform spectral patterns suggesting a stable envelope, while the anomalous recording (bottom) reveals sudden blanks in coefficients over short intervals, especially in coefficient 0 representing average log-energy. However, spectral patterns remain fairly similar between recordings, likely due to hydroelectric plant background noise.

Figure 6 displays the **FFT Amplitude spectrum**, revealing amplitude distribution across frequencies. The normal recording (top) shows a concentrated spectral profile in the lower frequencies, while the anomalous recording (bottom) exhibits wider energy distribution with distinct amplitude spikes in higher frequencies, corresponding to induced anomalies.

Table 2: Evaluation metrics for the synthetic anomaly dataset.

Method	Train Time (s)	ROC AUC	Precision	Recall	F1-Score	Inference Time (s)
K-Means	0.3700	0.9974	0.9825	0.9989	0.9906	0.0275
OC-SVM	2.8284	0.9977	0.9615	0.9994	0.9801	2.1469
LSTM-AE	32.8777	0.9995	0.9814	0.9972	0.9893	0.4765

4.2 Real-world Operation

The second dataset was recorded in an operational settings and comprises:

- **Normal operation (59 minutes 53 seconds):** equipment functioning under routine conditions, reflecting regular machine activity.
- **Transient acoustic event (5 seconds within 59 minutes):** detected during the transition phase, *i.e.*, electricity generation to water pump.

According to operational insights from plant personnel, acoustic anomalies typically occur during specific operational transitions: startup sequences, mode switching between pumping and generation phases, and load adjustment periods. These transient events are **not** indicative of equipment failure but rather reflect the dynamic behavior of turbine systems designed to operate in dual modes. Also, they are brief by nature, as prolonged anomalies would indicate actual equipment malfunction or damage rather than normal operational variations. The short duration of real anomalies (5 seconds within 59 minutes recording) and the constrained dataset size reflect this operational reality where brief acoustic deviations are characteristic of healthy industrial systems, combined with the practical challenges of extended data collection in critical infrastructure environments.

The **Mel-spectrogram** of the real anomaly dataset as well as the normal and anomalous frames are shown in Figure 7. The operating modes are clearly visible through pronounced energy shifts, but these do not necessarily correspond to anomalies since normal mode transitions are prolonged with uniform frequency distribution. Conversely, the anomalous recording (bottom) exhibits inconsistent frequency changes without apparent pattern throughout the recording.

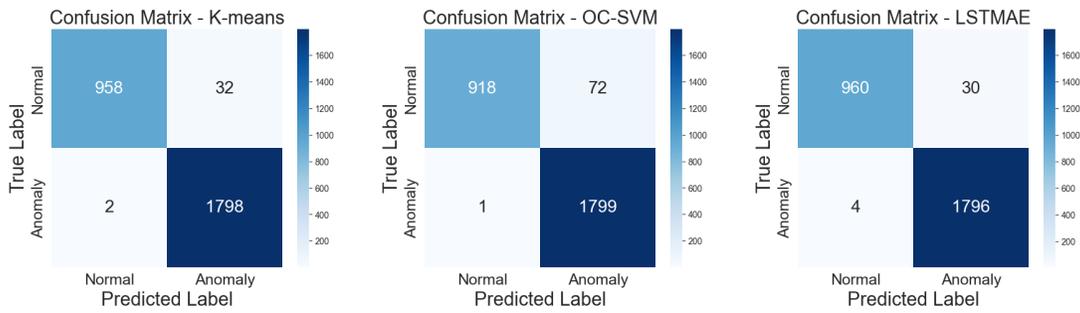


Figure 4: Confusion matrices: Induced (Synthetic) Anomalies Dataset

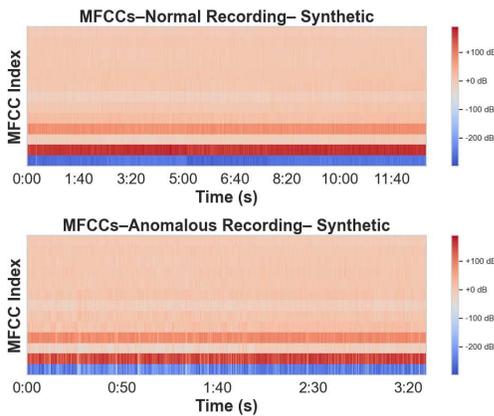


Figure 5: Comparison of Anomalous and Normal Recordings MFCCs

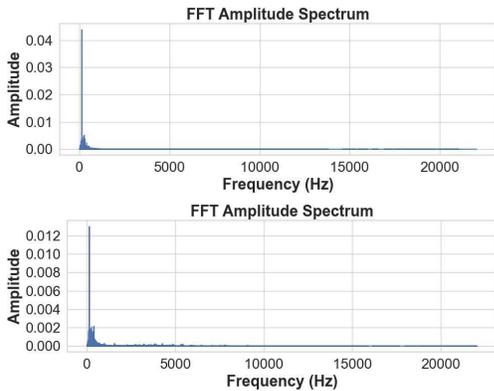


Figure 6: Comparison of Anomalous and Normal Recordings FFT Amplitudes

Evaluation metrics. In the real-world scenario, the confusion matrices (cf. Figure 8) confirm that all models successfully detected the transient event during the transition phase, with LSTM-AE and OC-SVM achieving perfect recall. LSTM-AE correctly identified all 43 transient events samples with 150 false positives, achieving the highest F1-score (0.360). OC-SVM also detected all anomalies but misclassified 243 normal samples (F1-score: 0.261). K-Means missed 6 anomalies with 156 false positives, achieving the lowest F1-score (0.314). Per Table 3, K-Means remains fastest for training and inference, while OC-SVM is significantly slower at inference.

The previous observations can also be seen in Figure 9, which shows the MFCCs of both recordings. The normal MFCC (top) shows consistent color bands in most coefficients in the normal recording indicating a stable spectral envelope. Apart from the evident the changes in operating modes in coefficient 0, each coefficient contains a stable spectral envelope. In contrast, the anomalous recording (bottom) exhibits higher variability in the MFCC coefficients. The coefficients 1-12 fluctuate more intensely during anomalies, showing unstable spectral patterns. Additionally, coefficient 0 shows a lower log-energy suggesting reduced energy in abnormal sounds.

Lastly, Figure 10 shows a comparison of the FFT Amplitude spectrum of both recordings. The amplitude and frequency distribution of the anomalous FFT (bottom) is significantly broader compared to the normal recording (top). The frequency distribution for the normal is much more compact, representing normal operating conditions. Furthermore, the amplitude peak in the normal recording is 0.010dB compared to a higher peak in the anomalous recording

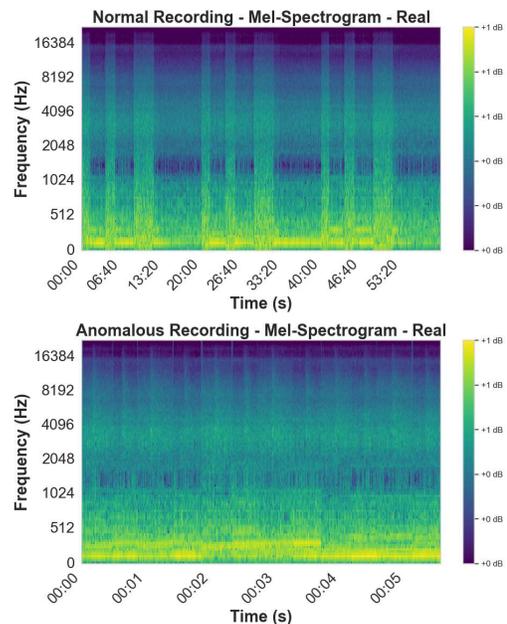


Figure 7: Mel-spectrograms for real dataset: normal (top) vs. anomalous (bottom).

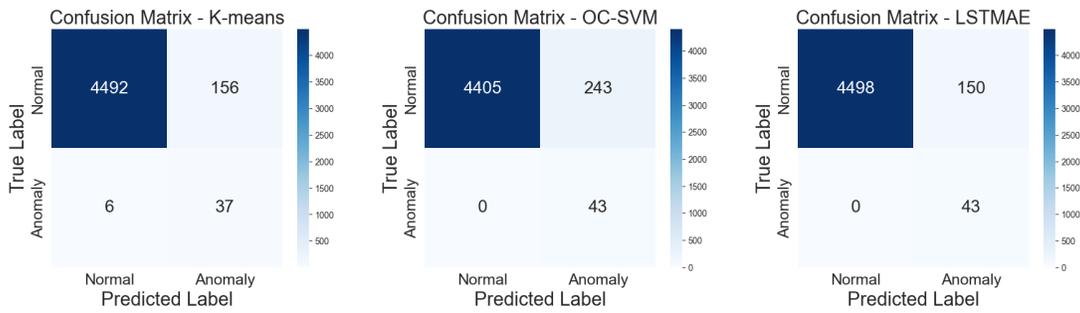


Figure 8: Confusion matrices: Real Industrial Dataset

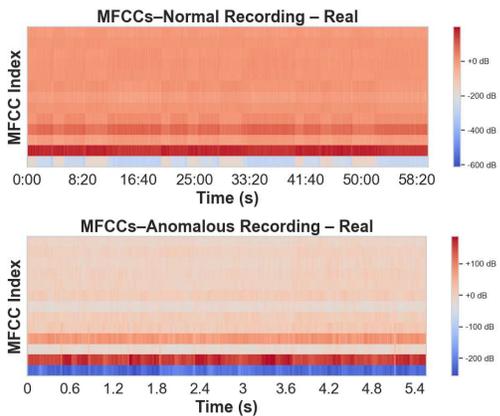


Figure 9: Comparison of Real Recordings MFCCs

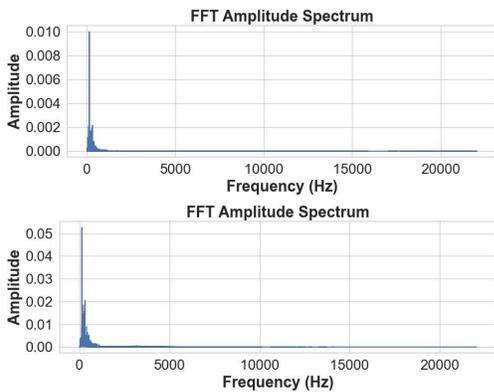


Figure 10: Comparison of Real Recordings FFT Amplitude

of over 0.050dB. This is due to the longer audio file for the normal recording and more distributed acoustic signals in contrast to the short and noisy signals in the anomalous recording.

Table 3: Evaluation metrics for the real world dataset.

Method	Train Time (s)	ROC AUC	Precision	Recall	F1-Score	Inference Time (s)
K-Means	5.25	0.962	0.192	0.861	0.314	0.043
OC-SVM	55.54	0.998	0.150	1.000	0.261	12.17
LSTM-AE	145.25	0.997	0.219	1.000	0.360	0.867

4.3 Comparative Analysis

The experimental evaluation outlines the crucial role of effective audio feature extraction. Mel-spectrograms and MFCCs provided key inputs by distinguishing subtle frequency variations, aligning with our exploratory data analysis. This preprocessing stage directly influenced model performance in detecting anomalies despite typical hydropower challenges like background noise and signal reflections.

Model selection involves notable **trade-offs** among accuracy, computational efficiency, and explainability:

- **K-means** offered fast inference (0.027–0.043s) for real-time applications but limited robustness to overlapping acoustic patterns (precision: 0.192).
- **OC-SVM** demonstrated strong detection (ROC AUC: 0.9976 – 0.998) with moderate computational overhead (training time: 2.8–55.5s) and intuitive boundary-based interpretability.
- **LSTM Autoencoder (AE)** captured complex temporal dependencies effectively (ROC AUC: 0.997–0.999) but required significantly higher computational complexity (training time: 145s). Its performance justifies deployment where accuracy is paramount, though explainability and maintenance may pose practical challenges.

5 Conclusions and Future Work

This paper presented an acoustic anomaly detection pipeline for predictive maintenance in PSHP power plants. We benchmark three ML models (K-Means, OC-SVM, and LSTM-AE) on real-world acoustic datasets from Rodundwerk II, analyzing audio feature selection and model trade-offs between accuracy, computational cost, and interpretability. OC-SVM achieved optimal efficiency-accuracy balance, while LSTM-AE excelled at detecting subtle temporal anomalies despite higher computational demands.

Future work includes developing a multi-modal anomaly detection framework that integrates acoustic signals with vibration data and operational parameters (e.g., rotational speed, bearing temperature, power output). Additionally, collecting expanded datasets from multiple microphone arrays positioned across different floors of the plant will enable extending detection capabilities from 2D to 3D, incorporating localization of anomalies.

Acknowledgments

This work was conducted in the context of a joint project with illwerke vkw AG, Bregenz, Austria, to whom the authors express gratitude for all support and discussions on technical aspects, and access to the Rodundwerk II hydropower plant.

References

- [1] Hoyjung Ahn and Inchoon Yeo. 2021. Deep-Learning-Based Approach to Anomaly Detection Techniques for Large Acoustic Data in Machine Operation. *Sensors* 21, 5446 (2021). doi:10.3390/s21165446
- [2] Barış Bayram, Taha Berkay Duman, and Gökhan Ince. 2021. Real time detection of acoustic anomalies in industrial processes using sequential autoencoders. *Expert Systems* 38, 1 (2021), e12564. doi:10.1111/exsy.12564
- [3] M. M. Breunig, H. P. Kriegel, R. T. Ng, and J. Sander. 2000. LOF: Identifying density-based local outliers. *ACM SIGMOD Record* 29, 2 (2000), 93–104. doi:10.1145/335191.335388
- [4] Gabriel Coelho, Luis Miguel Matos, Pedro José Pereira, André Ferreira, André Pilastri, and Paulo Cortez. 2022. Deep Autoencoders for Acoustic Anomaly Detection: Experiments with Working Machine and In-Vehicle Audio. *RepositóriUM* (2022). <https://repositorium.sdum.uminho.pt/bitstream/1822/81433/1/aeaad2.pdf> Accessed: 2025-02-16.
- [5] Leon Cohen. 1995. *Time-frequency analysis*. Vol. 778. Prentice Hall PTR New Jersey.
- [6] S. Davis and P. Mermelstein. 1980. Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. *IEEE Transactions on Acoustics, Speech, and Signal Processing* 28, 4 (1980), 357–366. doi:10.1109/TASSP.1980.1163420
- [7] E. Di Fiore, A. Ferraro, A. Galli, V. Moscato, and G. Sperli. 2022. An anomalous sound detection methodology for predictive maintenance. *Expert Systems with Applications* 209 (2022), 118324. doi:10.1016/j.eswa.2022.118324
- [8] T. B. Duman, B. Bayram, and G. Ince. 2020. Acoustic Anomaly Detection Using Convolutional Autoencoders in Industrial Processes. In *14th International Conference on Soft Computing Models in Industrial and Environmental Applications (SOCO 2019)*, F. Martínez Álvarez, A. Troncoso Lora, J. A. Sáez Muñoz, H. Quintián, and E. Corchado (Eds.). Springer International Publishing, 432–442. doi:10.1007/978-3-030-20055-8_41
- [9] E-control. 2025. Aktueller Marktpreis gemäß § 41 Abs. 1 Ökostromgesetz 2012. <https://www.e-control.at/industrie/oeko-energie/oekostrommarkt/marktpreise-gem-paragraph-20#:~:text=Dieser%20Marktpreis%20betr%C3%A4gt%20demzufolge%2097,Quartal%202024>.
- [10] Sarvarbek Ermiyazov, Yongmin Kim, M. Jaleel, and Chang Gyoon Lim. 2024. Comprehensive Analysis and Improved Techniques for Anomaly Detection in Time Series Data with Autoencoder Models. *International Journal on Advanced Science, Engineering and Information Technology* 14, 6 (Dec. 2024), 1861–1867. doi:10.18517/ijaseit.14.6.20451
- [11] A. Ferraro, A. Galli, V.L. Gatta, V. Moscato, M. Postiglione, G. Sperli, and F. Moscato. 2023. Unsupervised Anomaly Detection in Predictive Maintenance using Sound Data. *CEUR Workshop Proceedings* 3478 (2023), 449–458.
- [12] illwerke vkw. 2025. Rodundwerk II. <https://www.illwerkevkw.at/rodundwerk-ii>
- [13] M. Khanjari, A. Azarfar, M. H. Abardeh, and E. Alibeiki. 2024. Anomalous sound detection for machine condition monitoring using 3D tensor representation of sound and 3D deep convolutional neural network. *Multimedia Tools and Applications* 83, 15 (2024), 44101–44119. doi:10.1007/s11042-023-17043-9
- [14] Nand Kishor, RP Saini, and SP Singh. 2007. A review on hydropower plant models and control. *Renewable and Sustainable Energy Reviews* 11, 5 (2007), 776–796.
- [15] F. König, C. Sous, A. Ouald Chaib, and G. Jacobs. 2021. Machine learning-based anomaly detection and classification of acoustic emission events for wear monitoring in sliding bearing systems. *Tribology International* 155 (2021), 106811. doi:10.1016/j.triboint.2020.106811
- [16] Claudia Linnhoff-Popien, Steffen Illium, Fabian Ritz, and Robert Müller. 2021. Acoustic Anomaly Detection for Machine Sounds based on Image Transfer Learning. In *Proceedings of the 13th International Conference on Agents and Artificial Intelligence (Volume 2)*, Ana Paula Rocha, Luc Steels, and Jaap van den Herik (Eds.). SciTePress, Setúbal, Portugal, 49–56.
- [17] F. Lo Scudo, E. Ritacco, L. Caroprese, and G. Manco. 2023. Audio-based anomaly detection on edge devices via self-supervision and spectral analysis. *Journal of Intelligent Information Systems* 61, 3 (2023), 765–793. doi:10.1007/s10844-023-00792-2
- [18] Slaney Malcolm. 1998. Auditory toolbox version 2. <https://engineering.purdue.edu/~malcolm/interval/1998-010/> (1998).
- [19] Brian McFee, Colin Raffel, Dawen Liang, Daniel PW Ellis, Matt McVicar, Eric Battenberg, and Oriol Nieto. 2015. librosa: Audio and music signal analysis in python.. In *SciPy*. 18–24.
- [20] M. Meire and P. Karsmakers. 2019. Comparison of Deep Autoencoder Architectures for Real-time Acoustic Based Anomaly Detection in Assets. In *2019 10th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems (IDAACS)*, Vol. 2. 786–790. doi:10.1109/IDAACS.2019.8924301
- [21] R. Keith Mobley. 2002. 1 - Impact of Maintenance. In *An Introduction to Predictive Maintenance (Second Edition)* (second edition ed.), R. Keith Mobley (Ed.). Butterworth-Heinemann, Burlington, 1–22. doi:10.1016/B978-075067531-4/50001-4
- [22] Juan I Pérez-Díaz, Manuel Chazarra, Javier García-González, Giovanna Cavazzini, and Anna Stoppato. 2015. Trends and challenges in the operation of pumped-storage hydropower plants. *Renewable and Sustainable Energy Reviews* 44 (2015), 767–784.
- [23] M. Portnoff. 1980. Time-frequency representation of digital signals and systems based on short-time Fourier analysis. *IEEE Transactions on Acoustics, Speech, and Signal Processing* 28, 1 (1980), 55–69. doi:10.1109/TASSP.1980.1163359
- [24] Harsh Purohit, Ryo Tanabe, Takashi Ichige, Tomoya Endo, Yasunori Nikaido, Mofei Suefusa, and Yasue Kawaguchi. 2019. MIMI dataset: Sound dataset for malfunctioning industrial machine investigation and inspection. In *Proceedings of the Detection and Classification of Acoustic Scenes and Events 2019 Workshop (DCASE2019)*. doi:10.33682/m76f-d618
- [25] V N Ganapathi Raju, K Prasanna Lakshmi, Vinod Mahesh Jain, Archana Kalidindi, and V Padma. 2020. Study the Influence of Normalization/Transformation process on the Accuracy of Supervised Classification. In *2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT)*. 729–735. doi:10.1109/ICSSIT48917.2020.9214160
- [26] ScienceDirect. 2025. Audio Segmentation. <https://www.sciencedirect.com/topics/engineering/audio-segmentation> Accessed: 2025-02-13.
- [27] Siemens. 2023. The True Cost of Downtime: Identify, Avoid, and Overcome. <https://assets.new.siemens.com/siemens/assets/api/uid:3d606495-dbe0-43e4-80b1-d04e27ada920/dics-b10153-00-7600truecostofdowntime2022-144.pdf> Digital Industries, Customer Services.
- [28] Y. Tagawa, R. Maskeliūnas, and R. Damaševičius. 2021. Acoustic anomaly detection of mechanical failures in noisy real-life factory environments. *Electronics* 10, 19 (2021). doi:10.3390/electronics10192329
- [29] Han Wang, Dongdong Wang, Haoxiang Liu, and Gang Tang. 2022. A predictive sliding local outlier correction method with adaptive state change rate determining for bearing remaining useful life estimation. *Reliability Engineering and System Safety* 225 (2022), 108601. doi:10.1016/j.res.2022.108601
- [30] Yiming Yang. 1999. An Evaluation of Statistical Approaches to Text Categorization. *Information Retrieval* 1, 1-2 (1999), 69–90. doi:10.1023/A:1009982220290
- [31] Haoru Zhao, Baoshan Zhu, and Boshuang Jiang. 2025. Comprehensive assessment and analysis of cavitation scale effects on energy conversion and stability in pumped hydro energy storage units. *Energy Conversion and Management* 325 (2025), 119370.
- [32] Tiago Zonta, Cristiano André da Costa, Rodrigo da Rosa Righi, Miromar José de Lima, Eduardo Silveira da Trindade, and Guann Pyng Li. 2020. Predictive maintenance in the Industry 4.0: A systematic literature review. *Computers and Industrial Engineering* 150 (2020), 106889. doi:10.1016/j.cie.2020.106889