

Article

Data-Efficient Degradation Progression Modeling in Industrial Compressors via Baseline-Referenced Deep Feature Learning and Unsupervised Clustering

Gonca Öcalan ^{1,*}  and İbrahim Türkoğlu ² ¹ Department of Software Engineering, Graduate School of Natural and Applied Sciences, Firat University, Elazığ 23119, Türkiye² Department of Software Engineering, Faculty of Technology, Firat University, Elazığ 23119, Türkiye; iturkoglu@firat.edu.tr

* Correspondence: 192137201@firat.edu.tr

Abstract

Accurate modeling of degradation progression in rotating machinery remains challenging in real industrial systems, where data are inherently limited and imbalanced because of safety-critical operations and associated risks, and the cost of acquiring fault data is high. These conditions make it difficult for data-driven approaches to reliably capture the evolution of degradation over time. To address this challenge, this study proposes a hybrid framework that models degradation progression as a set of distinct behavioral regimes driven by loss of lubrication. The proposed framework first applies adaptive scaling guided by an α parameter derived from Root Mean Square (RMS) deviation of the vibration signals relative to the baseline condition, aiming to mitigate data leakage during preprocessing while improving robustness to data imbalance. It then performs baseline-referenced deep feature learning using a lightweight Long Short-Term Memory (LSTM) model trained only on baseline data. The trained model is subsequently used to encode the entire dataset into latent representations, which are finally clustered using Mini-Batch K-Means to organize distinct degradation-related behavioral regimes. Results on both real-world and experimental datasets demonstrate that the learned latent representations strongly agree with the degradation regimes associated with baseline characterization and α -guided progression patterns, achieving an Adjusted Rand Index (ARI) of 1.0 across both datasets with respect to the internally defined reference stages.

Keywords: condition monitoring; industrial compressors; loss of lubrication; unsupervised learning; LSTM-based feature learning; Mini-Batch K-Means; vibration signal analysis



Academic Editors: Qasim Zaheer,
Shi Qiu and Mohammad
AlHamaydeh

Received: 21 January 2026

Revised: 17 June 2026

Accepted: 3 July 2026

Published: 9 July 2026

Copyright: © 2026 by the authors.
Licensee MDPI, Basel, Switzerland.
This article is an open access article
distributed under the terms and
conditions of the [Creative Commons
Attribution \(CC BY\) license](https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Rotating machinery constitutes the backbone of modern industrial infrastructure, playing a critical role in power transmission, manufacturing, and energy systems. Despite their widespread use and mechanical sophistication, these systems remain vulnerable to unexpected failures, which can lead to costly downtime, safety risks, and reduced operational efficiency. Consequently, the early detection and monitoring of mechanical degradation has become a key challenge in predictive maintenance and reliability engineering.

Among the critical subsystems of rotating machinery, gear mechanisms are particularly susceptible to failure due to their high load conditions, frequent cyclic contact, and strong dependence on lubrication quality and adequacy. In particular, lubrication-related

faults typically evolve progressively over time, altering surface interactions and leading to increasingly severe degradation if not properly identified at early stages.

Loss of lubrication refers to the partial or complete breakdown of the lubricant film between contacting surfaces in mechanical components, resulting in direct metal-to-metal contact [1]. From a tribological perspective, this condition disrupts the fundamental functions of lubrication, including friction reduction, wear minimization, and heat dissipation, thereby accelerating surface degradation mechanisms [2]. In the absence of adequate lubrication, the system can no longer sustain stable hydrodynamic or boundary lubrication regimes and is forced to operate under mixed or dry friction conditions. This transition significantly accelerates the development of wear phenomena such as abrasive wear, adhesive wear, surface fatigue, pitting, and scuffing [1–7]. Consequently, prolonged lubrication loss leads to severe degradation, ultimately compromising the reliability, efficiency, and service life of mechanical systems.

Detecting loss of lubrication in rotating machinery remains a significant diagnostic challenge, as the condition typically progresses gradually and does not produce observable visual symptoms during operation. Consequently, several indirect sensing techniques have been developed to monitor lubrication status through its impact on system behavior. Thermal monitoring is widely used, as increased friction due to insufficient lubrication leads to localized heat generation, which can be detected using infrared thermography [8,9] or embedded temperature sensors [10]. Oil condition analysis provides another important approach by evaluating changes in lubricant properties, such as viscosity degradation, contamination, and the presence of metallic wear particles [6,11–13]. In addition, acoustic emission techniques enable the detection of high-frequency stress waves associated with micro-crack initiation and intensified surface interactions under inadequate lubrication conditions [14]. Among these methods, vibration analysis stands out due to its sensitivity to subtle mechanical variations induced by lubrication loss, making it particularly effective for detecting early-stage surface degradation [4,7]. This sensitivity has been consistently demonstrated in experimental studies, highlighting the strong relationship between lubrication conditions and vibration response. However, accurately characterizing and detecting lubrication loss-related degradation remains a significant challenge, motivating extensive research efforts in recent years.

2. Related Works

This section reviews existing studies by grouping them based on their methodological approaches and practical applicability, with particular emphasis on their limitations in real-world lubrication degradation scenarios.

2.1. Vibration Characteristics of Lubrication Degradation Under Lubrication Loss Conditions

Lubrication degradation in gear systems significantly influences friction, wear progression, and dynamic behavior, making it a critical factor in condition monitoring. Early studies largely concentrated on identifying vibration signatures associated with lubrication loss conditions, which represent a common form of lubrication degradation. For instance, Peng and Kessissoglou [15] demonstrated that lubrication loss in worm–crown gear systems results in elevated spectral baselines, increased sideband energy, and waveform distortions associated with intensified sliding contact and misalignment. Similarly, Banks et al. [16] reported that low-frequency vibration components exhibit limited sensitivity to lubrication variations, whereas high-frequency energy—particularly within the 15–24 kHz range—increases significantly as lubrication levels decrease.

Subsequent studies consistently demonstrate that lubrication degradation, particularly under lubrication loss conditions, fundamentally alters gear dynamics. For instance,

Brethee et al. [17] reported amplified gear mesh frequency (GMF) components and strengthened modulation sidebands under oil-starved conditions, indicating periodic impacts and nonlinear contact behavior. From a modeling perspective, Hu et al. [18] developed a nonlinear tribo-dynamic framework incorporating frictional effects, thermal dynamics, and progressive wear mechanisms, showing that lubrication loss introduces pronounced nonlinearities and alters both load-directional and non-load-directional motion.

Complementary tribological and data-driven investigations by Chen et al. [4] further revealed that reduced lubrication leads to increased friction coefficients, greater wear depth, and elevated high-frequency vibration content. By establishing a strong correlation between vibration-based features, particularly RMS, and tribological indicators, this study demonstrates the potential of vibration signals for data-driven characterization of lubrication degradation.

Despite this, because RMS is based on the overall energy of the signal, it tends to mask small impulsive changes such as micro-pitting or early-stage lubrication defects [19–21]. A significant increase in RMS generally occurs only when the damage has progressed sufficiently to elevate the total vibration energy. In contrast, kurtosis, by focusing on sudden shocks and impulsive events, is capable of detecting small faults and incipient degradation before they contribute noticeably to the overall energy level [19–21].

These findings collectively demonstrate that lubrication loss leads to distinct and measurable changes in vibration behavior, reflecting the underlying tribological and dynamic processes within the system, and providing a basis for data-driven interpretation of lubrication conditions.

2.2. Supervised Learning for Lubrication Degradation and Condition Monitoring

With the advancement of Machine Learning (ML), numerous studies have employed supervised learning techniques to classify lubrication conditions based on vibration signals. Saravanan and Ramachandran [22] utilized Artificial Neural Networks (ANN) combined with Discrete Wavelet Transform (DWT)-based feature extraction, achieving high classification accuracy across predefined lubrication states. Similarly, Keleşoğlu et al. [23] utilized frequency-domain features derived from Fast Fourier Transform (FFT) and Power Spectrum Density (PSD) analyses, which were subsequently employed as inputs to Multi-Layer Perceptron (MLP) and Neural Pattern Recognition (NPR) models for fault diagnosis under varying operating conditions and lubrication states. Additionally, Lupea et al. [24] confirmed that Support Vector Machines (SVM) based on gear mesh harmonics and modulation sidebands provide highly reliable fault detection.

More recent studies have explored Deep Learning (DL) architectures. Babu et al. [25] demonstrated that Deep Neural Networks (DNN) outperform conventional ANN models, achieving consistently high classification accuracy. Time–frequency-based approaches have also shown strong performance; Yu and Wei [7] and Lupea and Lupea [26] employed Continuous Wavelet Transform (CWT) in combination with Convolutional Neural Networks (CNN), achieving classification accuracies exceeding 99%. Subsequent work by Lupea et al. [27] revealed that raw vibration signals can be effectively processed using CNN architectures without the need for explicit feature extraction.

In addition to gearbox-focused studies, lubrication monitoring has also been investigated in bearing systems. Hemati [28] proposed a predictive maintenance framework based on vibration signals and LSTM networks using real industrial data collected over an extended period. The results indicate that lubrication loss introduces high-frequency spectral components and amplitude variations that can be effectively captured by DL models.

Although these supervised approaches achieve high classification accuracy, many existing studies treat lubrication conditions as discrete and predefined fault categories

(e.g., full, partial, and dry lubrication), thereby introducing artificial boundaries that limit the ability of such methods to capture transitional and overlapping states commonly observed in real industrial systems. Furthermore, most studies are conducted under controlled laboratory conditions, limiting their generalizability to real-world environments characterized by noise, non-stationarity, and varying operational loads.

In addition to lubrication-related studies, recent advances in supervised DL have demonstrated the effectiveness of combining time–frequency representations with attention-enhanced convolutional architectures for general fault diagnosis tasks. Approaches based on S-Transform and CWT have been integrated with lightweight and dual-stream CNN models to capture both temporal and spectral features [29,30]. The incorporation of attention mechanisms, such as Convolutional Block Attention Module (CBAM), further improves feature representation by emphasizing informative components while suppressing irrelevant information. These methods achieve high classification accuracy and robustness under noisy conditions. However, they are primarily formulated as supervised classification models and rely on labeled datasets and predefined fault categories.

These limitations highlight the need for alternative approaches that can operate without labeled data and better capture the complex and evolving nature of degradation processes.

2.3. Unsupervised and Data-Driven Anomaly Detection Methods

Unsupervised anomaly detection methods have gained increasing attention in recent years due to their ability to model normal system behavior without requiring labeled fault data. These methods can be broadly categorized into reconstruction-based, density-based, boundary-based and uncertainty-aware approaches, depending on how normal data characteristics are modeled, and are typically implemented using advanced DL architectures.

Reconstruction-based approaches are among the most widely adopted techniques, commonly implemented using Auto Encoder (AE) based DL models. These methods learn compact representations of normal operating conditions and detect anomalies based on reconstruction error. For instance, Li et al. [31] proposed a time–frequency convolutional AE that integrates CNN with spectral representations to enhance feature extraction under noisy conditions. Similarly, Zhong et al. [32] introduced a sliding-window convolutional AE capable of processing raw vibration signals using deep CNN architectures. In addition, Lee and Yu [33] proposed a hybrid anomaly detection framework that combines spectrogram-based convolutional AE with one-class classification techniques, enhancing robustness under non-stationary operating conditions.

Density-based approaches aim to model the underlying probability distribution of normal data and identify anomalies as low-probability events. These methods are typically implemented using probabilistic DL frameworks such as normalizing flows. Zhou et al. [34] proposed MTGFlow, a flow-based model that leverages deep neural networks to estimate data distributions and capture temporal dependencies in multivariate time series. Compared to reconstruction-based methods, density-based approaches provide a more expressive representation of system behavior.

Boundary-based approaches focus on learning a decision boundary that encloses normal data, enabling anomaly detection based on deviations from this boundary. These methods are closely related to one-class learning paradigms and are often implemented using deep representation learning techniques. Yin et al. [35] developed an adaptive hypersphere-based anomaly detection model using a Deep Separable Linear Attention AE (DSAAE), which integrates separable convolution and attention mechanisms to capture complex feature dependencies. The proposed approach learns compact feature represen-

tations and introduces adaptive, variable-radius decision boundaries in the latent space, while jointly optimizing reconstruction and boundary learning.

In addition to these core categories, uncertainty-aware approaches have been proposed to enhance detection robustness. Çekik and Turan [36] proposed a RoughLSTM-based model that integrates rough set theory with LSTM to explicitly model uncertainty in vibration signals.

Despite their effectiveness, these approaches are primarily designed for anomaly detection, focusing on distinguishing normal and abnormal conditions rather than uncovering the underlying structure of degradation. As a result, they tend to provide limited insight into the internal organization and progression of degradation patterns, particularly in complex industrial systems where degradation does not manifest as clearly separable states.

To facilitate a clear comparison of existing approaches, Table 1 presents a summary of the reviewed studies in terms of their methodologies, strengths, and limitations.

Table 1. Summary of Existing Approaches.

Study Type	Method	Strength	Limitation
Vibration-based (e.g., [15–18])	Signal analysis	Physically interpretable	Laboratory-based
Supervised ML and DL (e.g., [7,22–27])	FFT, PSD, DWT, CWT and raw signals ANN, SVM, CNN	High classification accuracy	Laboratory-based, label dependency
Real-world DL (e.g., [28])	Time, frequency and envelop LSTM	Industrial validation	label dependency
Advanced DL (e.g., [29,30])	S-Transform, CWT CNN + attention	Enhanced feature representation	Not lubrication-specific, label dependency
Unsupervised DL (e.g., [31–36])	AE + CNN, AE + CNN + attention, flow-based models, uncertainty-aware models	Label-free	Not lubrication-specific, No gradual degradation modeling

Although recent studies have significantly advanced AI-based fault diagnosis, as summarized in Table 1, the existing literature still exhibits a fundamental methodological limitation: both supervised and unsupervised approaches tend to constrain degradation behavior into easily separable states. Supervised methods [7,22–27] achieve high classification accuracy when sufficient labeled data and predefined fault categories are available, whereas unsupervised methods [31–35] reduce label dependency and offer better adaptability to non-stationary environments. However, in practice, both paradigms are still largely evaluated within formulations that distinguish either discrete classes or binary normal–abnormal conditions. These formulations may be insufficient for lubrication loss-related degradation, which typically evolves gradually rather than abruptly and may involve transitional stages with overlapping characteristics. As a result, the progression from healthy to degraded behavior cannot always be represented reliably through fixed class boundaries or binary anomaly labels. In addition, most existing studies are conducted under controlled laboratory conditions, where operating parameters, lubrication levels, and environmental variability are artificially regulated, limiting their transferability to real industrial systems.

Beyond these formulation-related limitations, real industrial datasets introduce further challenges that remain insufficiently addressed in the literature. Fault samples are inherently scarce, leading to severe class imbalance; preprocessing steps such as global normalization may unintentionally introduce data leakage and information loss [37–39]; and the resulting models often lack robustness under evolving operating conditions. Recent reviews [40,41] have also emphasized that practical industrial deployment increasingly requires interpretable, scalable, and physically meaningful models rather than highly accurate but isolated benchmark solutions.

Taken together, these limitations reveal a fundamental gap in the literature: the lack of robust frameworks capable of modeling lubrication loss as a gradual, transitional, and partially overlapping degradation process under realistic industrial conditions. Consequently, existing approaches typically report high classification or clustering performance, yet may not adequately capture degradation progression. This limitation is further challenged by limited and imbalanced data and the risk of data leakage during preprocessing, while many existing solutions remain insufficient in terms of scalability and physical interpretability.

This study addresses this gap by proposing a hybrid framework that combines representation learning with unsupervised clustering to model lubrication-loss degradation as a structured and progressively evolving process. To ensure reliable preprocessing, adaptive scaling is introduced, which prevents data leakage and improves robustness to data imbalance. Building on this, the framework employs an LSTM trained exclusively on baseline data to learn a reference representation and to encode system behavior relative to this baseline into a set of latent representations. The resulting latent representations are subsequently clustered to uncover data-driven behavioral regimes associated with lubrication loss, enabling the characterization of progressive degradation behavior and capturing the progression of lubrication-loss dynamics. Finally, the proposed method is evaluated on both a real-world compressor dataset and an experimental dataset, demonstrating its scalability and applicability across different operating conditions.

The main contributions of this study are as follows:

- A framework for modeling lubrication-loss degradation as a structured and progressively evolving process, without relying on predefined discrete or binary formulations.
- An adaptive class-wise scaling strategy is introduced to mitigate data leakage during preprocessing while preserving the temporal progression of degradation and reducing the adverse effects of data imbalance on feature representation.
- A baseline-referenced deep feature learning strategy is proposed, enabling system behavior to be represented relative to baseline conditions.

In addition to these contributions, the proposed framework offers favorable computational efficiency and scalability. By avoiding computationally intensive feature extraction and employing a lightweight LSTM architecture in conjunction with Mini-Batch K-Means, the framework maintains consistent clustering performance. These characteristics suggest that the proposed approach is suitable for large-scale monitoring applications under real-world computational constraints.

To implement the proposed framework, the methodology adopted in this study is presented in the following section.

3. Materials and Methods

This section presents the proposed data-driven diagnostic methodology for monitoring the progression of lubrication loss. The overall workflow of the proposed approach is illustrated in Figure 1 to provide a clear and comprehensive visual overview. The figure conceptualizes the proposed framework as a unified pipeline composed of four tightly coupled components: (i) adaptive class-wise scaling explicitly designed to mitigate data leakage and preserve regime-specific statistical structure, (ii) baseline-referenced deep feature learning via a lightweight LSTM architecture to capture intrinsic temporal dependencies, (iii) latent feature space representation that encodes the continuous degradation trajectory, and (iv) unsupervised identification of distinct yet progressively evolving degradation stages associated with lubrication loss. The detailed explanation of each step is provided in the following subsections.

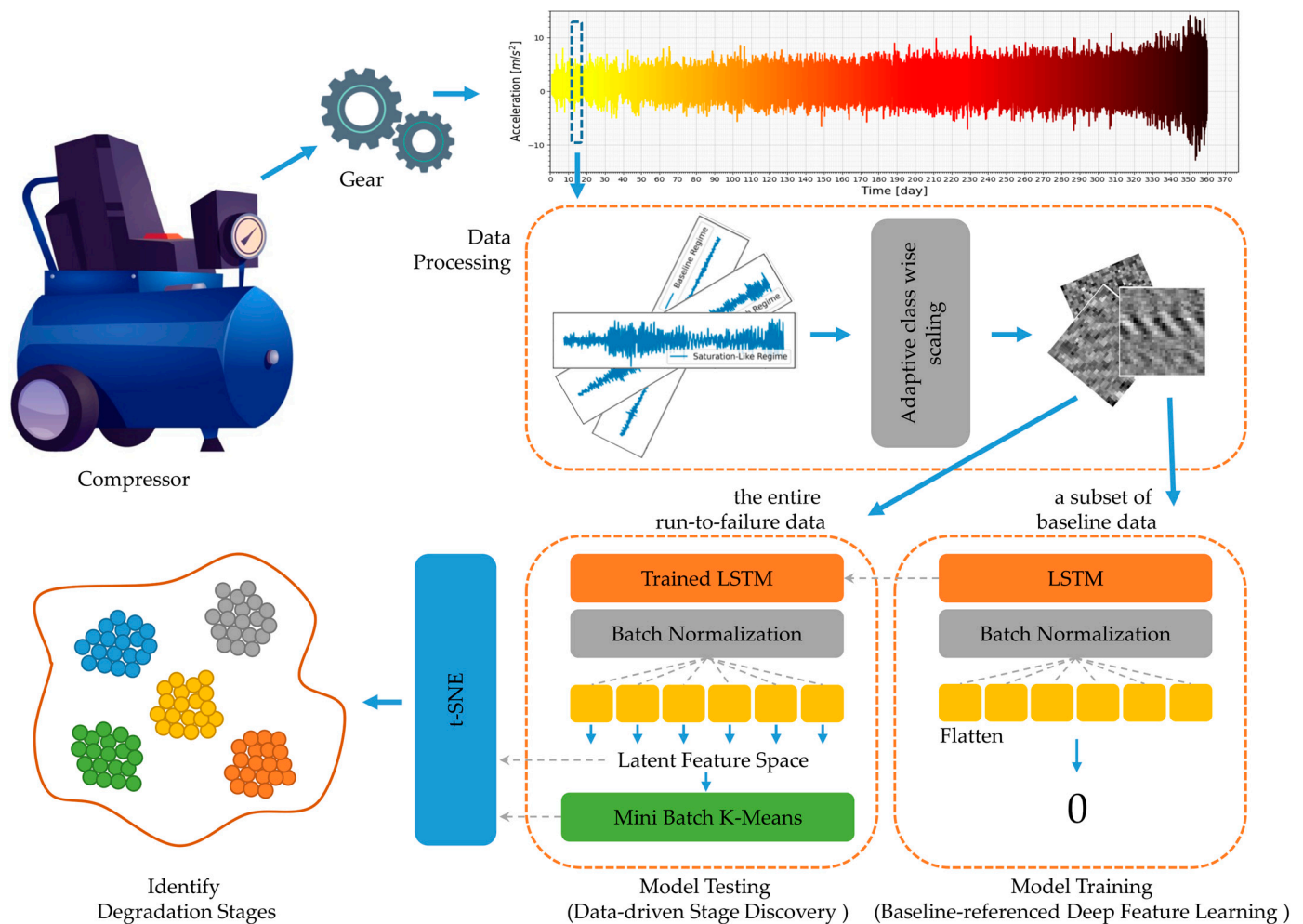


Figure 1. Overall workflow of the proposed data-driven diagnostic framework for lubrication loss progression. The different colored clusters correspond to different degradation stages. Certain minor graphical elements were adapted from designs by upklyak and rawpixel.com, via Freepik.

3.1. Data Preprocessing and Representation Strategy

The pattern recognition performance of AI-based models is directly influenced by the discriminative power and representational capacity of the learned features. Low-dimensional yet information-rich feature representations enable faster training convergence and improved generalization under varying operational conditions. However, traditional frequency and time frequency domain techniques, such as FFT, PSD [23] and CWT [7], require handcrafted feature extraction and impose substantial computational and memory demands, while their inherent assumptions including local stationarity, fixed window sizes, predefined wavelet bases and linearity may lead to information loss and distortion of subtle fault related structures, thereby limiting their practicality for real time machinery health monitoring. Therefore, the present study adopts an end-to-end representation learning strategy directly applied to raw vibration data, significantly reducing the reliance on manual preprocessing and explicit feature engineering. Within this framework, two key components are introduced to enhance representation quality and learning efficiency: an adaptive class-wise scaling mechanism and a 1D-to-2D representation conversion strategy.

3.1.1. Adaptive Class-Wise Scaling

Conventional normalization strategies, including global scaling, healthy-state-based normalization, and class-wise normalization, exhibit notable limitations when applied to

degradation modeling problems. Global normalization often compresses the variability across different degradation stages, leading to information loss and reduced sensitivity to early-stage faults. In addition, when scaling parameters are computed using the full run-to-failure dataset, it may introduce data leakage by implicitly incorporating information from future degradation states, which would not be available in real-world monitoring scenarios. This results in an unrealistic representation of early-stage data and further reduces sensitivity to incipient faults.

Healthy-state-based normalization, while providing a fixed reference, may fail to capture the evolving data distribution as degradation progresses, leading to a mismatch between early and advanced stages. Although class-wise normalization improves intra-class consistency, it typically assumes prior knowledge of class labels; however, it also provides a practical mechanism for handling imbalanced datasets by enabling class-specific scaling, thereby preventing dominant regimes from disproportionately influencing the feature space. In practical degradation scenarios, however, stage boundaries are not known a priori and can only be determined after observing the full degradation trajectory. Consequently, such approaches may implicitly rely on future information, leading to potential data leakage and limiting their applicability in real-world settings.

To address these limitations, an adaptive class-wise min–max scaling strategy is proposed, in which pseudo-classes representing degradation behavior are internally constructed using physically motivated signal characteristics, namely RMS and kurtosis. These features respectively capture the overall energy content and impulsive behavior of the vibration signal, enabling stage-aware normalization [34,42,43] without requiring prior knowledge of class boundaries, and ensuring both physical consistency and robust feature representation.

To reliably characterize the progression of lubrication degradation under real industrial conditions, a two-stage indicator framework is adopted, combining kurtosis-based onset detection with RMS-driven adaptive scaling.

In the first step, RMS and kurtosis values are computed for each vibration measurement. The mathematical definitions of RMS and kurtosis are given in Equations (1) and (2), respectively.

$$\text{RMS} = \sqrt{\frac{1}{N} \sum_{i=1}^N x_i^2} \quad (1)$$

$$\text{kurtosis} = \frac{1}{N} \sum_{i=1}^N \left(\frac{x_i - \mu}{\sigma} \right)^4 \quad (2)$$

where μ and σ denote the sample mean and sample standard deviation, respectively. The non-excess (Pearson) kurtosis formulation is adopted, for which a Gaussian distribution has a kurtosis value of 3.

In the initial stage, kurtosis is employed to identify the onset of degradation and to establish the early pseudo-class boundaries. Owing to its high sensitivity to impulsive events, kurtosis enables reliable discrimination between normal operation and incipient fault conditions, thereby guiding the formation of initial class-wise groupings.

Specifically, when the kurtosis value remains consistently close to the Gaussian reference level (≈ 3), the system is interpreted as operating under nominal conditions. As the kurtosis increases within the range of approximately 3.5 to 4.5, the system is inferred to transition into an early-stage degradation regime, reflecting the emergence of impulsive characteristics in the vibration signal. It should be noted that these thresholds are not treated as strict decision boundaries, but rather as empirical indicators used to guide the interpretation of data-driven behavioral regimes.

While kurtosis provides high sensitivity to impulsive events and is therefore well-suited for identifying the onset of degradation, its effectiveness diminishes beyond the early-stage regime due to saturation and increased variability. Consequently, beyond this range, the class-wise scaling strategy is no longer governed by kurtosis; instead, it is progressively driven by a nonlinear RMS-based transformation. More precisely, RMS values are normalized with respect to a robust baseline reference derived from the incipient degradation regime, and subsequently mapped to an adaptive scaling parameter through an exponential function, enabling a smooth and physically consistent transition across degradation stages. This parameter (α) is subsequently used to refine and extend the pseudo-class structure, enabling adaptive class-wise normalization across advanced degradation stages without requiring explicit labels.

First, the reference RMS value, denoted as RMS_H , is defined as a baseline quantity derived from the incipient degradation regime, as given in Equation (3)

$$\text{RMS}_H = \text{median}(\text{RMS}_{\text{base}}) \quad (3)$$

where RMS_{base} represents the RMS values obtained from the incipient degradation regime. The use of the median, rather than the mean, enhances robustness against impulsive noise and early-stage irregularities commonly observed in real industrial vibration signals.

Based on this reference, a normalized degradation indicator, ρ , is formulated as given in Equation (4):

$$\rho_i = \max\left(\frac{\text{RMS}_i}{\text{RMS}_H}, 1\right) \quad (4)$$

where RMS_i and ρ_i denote the RMS and ρ values computed from the i -th measurement, respectively. This ratio provides a dimensionless and physically interpretable measure of relative energy growth with respect to the baseline. Since the considered fault mechanism is associated with lubrication loss, which typically leads to increased friction and consequently elevated vibration energy, RMS values below the baseline were not interpreted as indicators of fault progression. Therefore, a lower bound of 1 was imposed to suppress minor sub-baseline fluctuations and ensure monotonic interpretability of the degradation indicator ρ .

Subsequently, the normalized indicator is mapped to an adaptive scaling parameter, denoted as α , through an exponential transformation as given in Equation (5):

$$\alpha_i = \alpha_{\min} + (\alpha_{\max} - \alpha_{\min}) \left(1 - e^{-\beta(\rho_i - 1)}\right) \quad (5)$$

where α_i denotes the α value computed from the i -th measurement, $\alpha_{\min}$, $\alpha_{\max}$ and β control the lower bound, upper bound, and sensitivity of the nonlinear mapping, respectively. By setting $\alpha_{\min} = 0$ and $\alpha_{\max} = 1$ the adaptive scaling parameter α is bounded within the interval $[0, 1]$. The exponential formulation is adopted to ensure a smooth, monotonic, and progressively saturating response of the scaling parameter, allowing low sensitivity to minor fluctuations in early-stage conditions while amplifying significant deviations in advanced degradation stages. Accordingly, β ($= -\ln(1 - 0.95) \approx 2.9957$) is calibrated such that the exponential mapping reaches 95% of its maximum response at $\rho = 2$, ensuring that moderate deviations from the baseline are sufficiently amplified while preventing excessive sensitivity at higher deviation levels.

To emulate class-wise normalization without relying on labeled data, the continuous α values are discretized into a finite set of levels:

$$\alpha \rightarrow \{\alpha^{(1)}, \alpha^{(2)}, \dots, \alpha^{(k)}\}$$

Each discrete level implicitly defines a pseudo-class corresponding to a specific degradation regime. Measurements exhibiting similar α -values are grouped accordingly. Prior to normalization, each measurement is partitioned into non-overlapping windows corresponding to one full shaft revolution, ensuring a physically meaningful and consistent segmentation of the vibration signal. Subsequently, class-wise min–max scaling is performed within each pseudo-class, where feature-wise scaling bounds are computed using only the windows belonging to the same group. For each pseudo-class k , the scaled representation is computed as given in Equation (6):

$$X_{scaled}^{(k)} = \frac{X^{(k)} - X_{min}^{(k)}}{X_{max}^{(k)} - X_{min}^{(k)}} \quad (6)$$

where $X_{min}^{(k)}$ and $X_{max}^{(k)}$ are feature-wise extrema obtained from windows belonging to the same α -level. This allows each degradation regime to be normalized according to its own statistical distribution, thereby mitigating the distortion caused by global normalization under non-stationary conditions.

By operating within physically consistent degradation regimes, the proposed approach reduces intra-regime variability while preserving inter-regime progression differences embedded in the vibration data, ultimately enhancing feature separability and facilitating subsequent representation learning and clustering.

It should be emphasized that the α parameter is not directly used during the training of the deep learning model or as an input feature for clustering. Instead, it serves as an adaptive scaling guide during preprocessing to reduce inter-regime scale disparities, mitigate potential data leakage effects, and improve robustness against data imbalance. The subsequent latent representations are learned by the LSTM model from the vibration signals themselves, while the clustering algorithm organizes the resulting latent space into degradation-related behavioral patterns beyond simple RMS-based partitioning. In this context, the α -guided pseudo-classes are used solely as internally constructed reference degradation stages for evaluating the consistency of the clustering results.

3.1.2. 1D-to-2D Representation Conversion

Following adaptive class-wise scaling, the one-dimensional (1D) vibration signal windows are reshaped into two-dimensional (2D) representations using a structured reshaping strategy [44–46]. Specifically, each scaled signal window corresponding to one complete machine rotation is reorganized into a matrix form with an equivalent number of elements. This transformation preserves the intrinsic characteristics of the original signal while enabling the model to capture local temporal dependencies as spatial patterns, thereby enhancing feature discrimination and facilitating the extraction of degradation-sensitive spatio-temporal features.

Let the scaled vibration signal corresponding to the i -th window be represented as a 1D sequence:

$$\tilde{x}_i = [x_{i,1}, x_{i,2}, \dots, x_{i,N}] \in \mathbb{R}^N$$

where N denotes the number of samples within a single window corresponding to one complete machine rotation.

To enable structured representation of the signal, the 1D sequence is reorganized into a 2D matrix as given in Equation (7):

$$X_i \in \mathbb{R}^{h \times w} \text{ such that } h \times w = N$$

$$X_i = \text{reshape}(\tilde{x}_i, h, w) \quad (7)$$

More precisely, the reshaping operation is performed in a row-wise manner and is defined element-wise as:

$$X_i(p, q) = x_{i,(p-1)w+q}, \quad p = 1, \dots, h, \quad q = 1, \dots, w$$

Accordingly, the equivalent 2D matrix representation is given by:

$$X_i = \begin{bmatrix} x_{i,1} & x_{i,2} & \cdots & x_{i,w} \\ x_{i,w+1} & x_{i,w+2} & \cdots & x_{i,2w} \\ \vdots & \vdots & \ddots & \vdots \\ x_{i,(h-1)w+1} & x_{i,(h-1)w+2} & \cdots & x_{i,hw} \end{bmatrix}$$

In this stage, each 2D matrix represents a structured temporal map, where each vibration signal sample is uniquely mapped to a matrix element while preserving the original temporal order without distortion. This representation retains the complete temporal dynamics of the signal and organizes adjacent samples into spatially coherent local regions. Such a structure enhances the capability of LSTM networks to learn both short-range and long-range temporal dependencies more effectively than when using unstructured 1D sequences. Additionally, this structured format promotes consistent memory allocation and improves the efficiency of batch-wise recurrent computations during the training phase.

3.2. Proposed Diagnostic Framework

The proposed diagnostic framework integrates deep temporal feature learning with unsupervised clustering to systematically model the progression of lubrication loss and categorize it into distinct degradation-related behavioral regimes. Using the constructed representative temporal maps, an LSTM network is employed to learn the intrinsic vibrational patterns associated with the baseline operating condition, defined by the incipient degradation regime. The trained LSTM network is then used to map the entire run-to-failure sequence into a compact and discriminative latent space, where each sample is represented by a latent feature vector. Finally, an unsupervised Mini-Batch K-Means algorithm is applied to these latent feature vectors to automatically identify distinct degradation regimes without requiring labeled fault data.

The proposed framework is further detailed through deep feature learning and unsupervised clustering in the following section.

3.2.1. Baseline-Referenced Deep Feature Learning via LSTM and Latent Space Projection

In this study, deep feature extraction is performed using an LSTM [47,48] network trained exclusively on baseline data. The baseline operating condition is defined by early-stage degradation regimes, enabling the model to learn the intrinsic temporal dynamics representative of baseline system behavior.

LSTM networks are a class of recurrent neural networks designed to capture long-term temporal dependencies in sequential data through gated memory mechanisms.

At each time step t , the input sequence x_t , previous hidden state h_{t-1} , and previous cell state c_{t-1} are used to update the internal memory of the network, as defined in Equations (8)–(13):

$$f_t = \sigma_s(W_f x_t + U_f h_{t-1} + b_f) \quad (8)$$

$$i_t = \sigma_s(W_i x_t + U_i h_{t-1} + b_i) \quad (9)$$

$$\tilde{c}_t = \tanh(W_{\tilde{c}} x_t + U_{\tilde{c}} h_{t-1} + b_{\tilde{c}}) \quad (10)$$

$$o_t = \sigma_s(W_o x_t + U_o h_{t-1} + b_o) \quad (11)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t \quad (12)$$

$$h_t = o_t \odot \tanh(c_t) \quad (13)$$

where f_t , i_t and o_t denote the forget, input, and output gates, respectively; $\tilde{c}_t$ represents the candidate cell state; σ_s is the sigmoid activation function; $\tanh$ hyperbolic tangent activation function and $\odot$ denotes element-wise multiplication. These gating mechanisms enable the LSTM to selectively retain relevant temporal information while suppressing noise and less informative signal variations.

To this end, the 2D temporal maps corresponding to baseline conditions are fed into the LSTM as sequences of feature vectors, allowing the network to learn the intrinsic temporal dynamics of baseline system behavior. Notably, the input sequences are provided to the LSTM without flattening, thereby preserving the inherent structural and temporal relationships within the data. Specifically, each row of the 2D temporal map is treated as a time step, while the corresponding row values form the feature vector at that step. This formulation enables the network to capture both short-term and long-term temporal dependencies, along with implicit spatial dependencies encoded in the structured representation.

The overall architecture of the proposed framework is illustrated in Figure 2. Unlike classification-oriented architectures commonly used in the literature, the proposed LSTM operates in a regression setting, where it is trained to reproduce a constant zero reference output. By enforcing the network to approximate this baseline response, the model internalizes the intrinsic temporal dynamics of baseline system behavior. This zero-reference formulation provides a consistent baseline representation, such that deviations from the learned response can be directly associated with changes in system condition. Through this learning process, the LSTM acts as an encoder, transforming the input sequences into compact and discriminative feature representations that capture the underlying temporal patterns of system behavior.

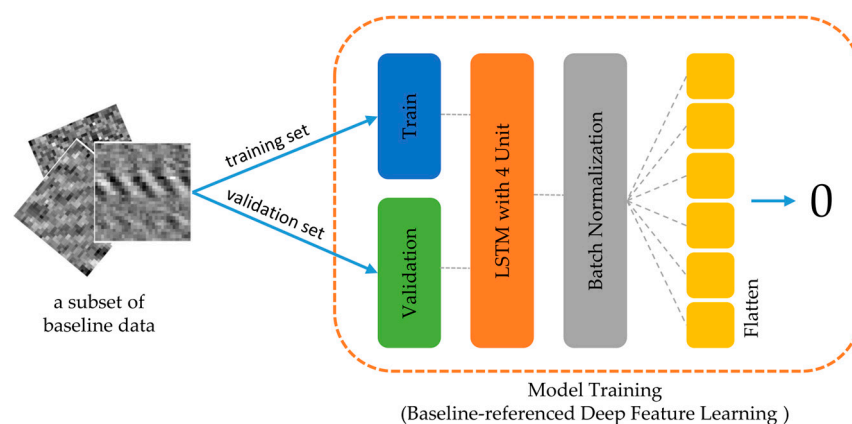


Figure 2. Architecture of the proposed LSTM-based deep feature learning model used to extract temporal representations from 2D structured vibration maps.

Unlike conventional reconstruction- or prediction-based approaches, the proposed method directly enforces a reference output, enabling the model to learn a baseline-consistent representation without requiring explicit reconstruction or forecasting. This formulation simplifies the learning objective, allowing deviations from the reference response to be directly interpreted as indicators of degradation.

To operationalize this concept, the trained LSTM network is applied to the entire run-to-failure representative temporal maps, generating a compact latent feature vector for each map, as illustrated in Figure 3. These vectors encapsulate the temporal dynamics learned from baseline operation and provide a discriminative embedding that characterizes

machinery-specific behavior. The collection of all latent vectors defines a latent feature space, within which deviations associated with lubrication degradation become distinguishable from the baseline structure. When the entire run-to-failure sequence is projected into this space, samples corresponding to progressively deteriorating lubrication conditions increasingly diverge from the baseline trajectory, forming a coherent degradation pathway. This latent manifold reveals a structured progression of lubrication loss, rendering the underlying degradation process both visually interpretable through low-dimensional projections and analytically tractable, and thereby facilitating the subsequent unsupervised identification of discrete degradation stages.

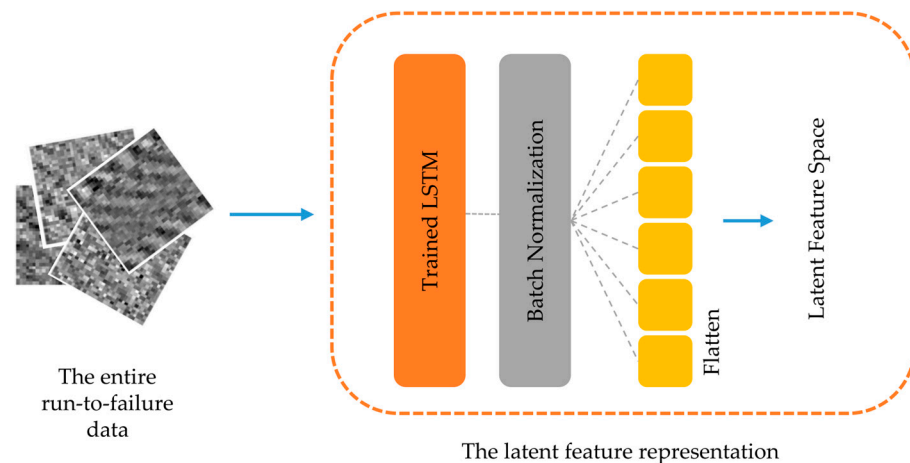


Figure 3. Transformation of run-to-failure representative temporal maps through the trained LSTM model into compact latent feature vectors, which collectively form the latent feature space used for degradation analysis.

The proposed LSTM architecture comprises a single LSTM layer with 4 units, followed by a batch normalization layer and a flattening layer that prepares the extracted features for subsequent processing. The LSTM layer employs the hyperbolic tangent activation function, and its weights are initialized using the Truncated Normal distribution with a fixed random seed of zero to ensure reproducibility. Furthermore, the `return_sequences` parameter is set to `True` so that the LSTM outputs a feature vector at every time step, enabling the extraction of the full temporal structure encoded in each representative map. The latent feature dimension depends on the input map size and is determined as:

$$D_{latent} = T \times U \quad (14)$$

where T denotes the number of timesteps and U the number of LSTM units.

During model training, the Log-Cosh loss function [49] is employed to minimize the deviation between the latent representations and the zero-reference target. This encourages the network to model the baseline behavior as a stable reference, allowing deviations in the output to reflect degradation-induced changes in system dynamics. The model parameters are then optimized by minimizing this loss using the Stochastic Gradient Descent (SGD) algorithm.

In the training phase, the learning rate of SGD was set to 1×10^{-2} , and the momentum coefficient was chosen as 0.9 to enhance convergence stability. The training procedure was conducted for 100 epochs, with the batch size set equal to the size of the training dataset composed exclusively of baseline data; given the limited number of samples, this corresponds to full-batch training, which ensures stable gradient estimation and avoids unnecessary stochasticity introduced by mini-batch training.

To ensure optimal generalization performance, the ModelCheckpoint callback from the Keras library was used, enabling automatic storage of the model weights whenever the validation loss reached its minimum during training. After the training phase, these best-performing weights were reloaded for the test stage to guarantee evaluation under the most favorable configuration.

The proposed model was implemented in Python 3.11.4 (Python Software Foundation), and the computational framework relied on TensorFlow 2.14 [50] (Google LLC, Mountain View, CA, USA) with the integrated Keras API, Scikit-learn 1.3.0 [51] (Scikit-learn Developers), Numpy 1.24.3 [52] (NumPy Developers) and Pandas 1.5.3 [53] (Pandas Development Team) libraries for deep learning, numerical optimization, and auxiliary processing.

3.2.2. Unsupervised Clustering of the Latent Feature Space

To obtain a discrete characterization of the lubrication-loss degradation process, the latent feature space is partitioned using an unsupervised clustering approach. The complete set of latent feature vectors, generated by passing the run-to-failure representative temporal maps through the trained LSTM network, is used as input to the clustering algorithm. By operating directly on this machinery-specific embedding, the proposed method groups latent feature vectors according to their intrinsic deviations from baseline operating dynamics, rather than relying on predefined labels. This enables the identification of structurally coherent degradation regimes that reflect the progressive evolution of lubrication loss, while facilitating the distinction between baseline-consistent and degradation-induced patterns.

A Mini-Batch K-Means algorithm is employed to partition the latent feature space due to its computational efficiency and scalability for large-scale, high-dimensional machinery datasets. Unlike standard K-Means, which requires repeated full passes over the dataset and becomes increasingly inefficient as the sample size grows, Mini-Batch K-Means updates cluster centroids using small, randomly sampled mini-batches [54,55].

Mini-Batch K-Means follows the same objective function as classical K-Means, while updating cluster centroids incrementally using stochastic approximations based on randomly sampled mini-batches [56]. The Mini-Batch K-Means algorithm minimizes the within-cluster variance objective defined in Equation (14):

$$\min_{\{\mu_k\}} \sum_{i=1}^N \|x_i - \mu_{c_i}\|^2 \quad (15)$$

In this formulation, N denotes the total number of data samples, while $x_i \in \mathbb{R}^d$ represents the i -th data point in the feature space, corresponding to a latent feature vector in the proposed framework. The term μ_k denotes the centroid of the k -th cluster, and c_i indicates the cluster assignment index for the i -th sample. The expression $\|x_i - \mu_{c_i}\|^2$ represents the squared Euclidean distance between a data point and the centroid of its assigned cluster. The objective function therefore minimizes the total within-cluster variance by assigning each sample to the nearest centroid and iteratively updating the cluster centers [56]. Mini-Batch K-Means is particularly advantageous in the present study, as it efficiently handles the large number of latent vectors generated from extended run-to-failure sequences, enabling scalable characterization of the degradation trajectory in the latent space.

For the clustering stage, the Mini-Batch K-Means algorithm was configured with the initialization method set to k-means++ and a fixed random state of 42 to ensure reproducibility. The batch size was defined as multiples of the number of windows extracted from each measurement. Under this configuration, the most favorable clustering results were obtained with batch sizes of 1547 (91×17) for Case Study I and 330 (110×3) for Case Study II.

3.3. Evaluation Metrics and Visualization

To quantitatively evaluate the agreement between the unsupervised cluster assignments and the reference degradation stages, three widely used external clustering validity indices were employed: ARI, Normalized Mutual Information (NMI) and Fowlkes–Mallows Index (FMI) [49]. ARI measures the similarity between two partitions while correcting for chance, providing a robust indicator of labeling consistency [49]. NMI quantifies the amount of shared information between the predicted clusters and the reference labels, with higher values indicating stronger correspondence [49]. FMI evaluates the balance between precision and recall in the clustering result, thereby reflecting the degree of correct pairwise assignment [49]. Together, these metrics offer a comprehensive assessment of clustering fidelity from complementary statistical perspectives.

In addition to these external metrics, internal clustering validity indices were also considered to assess the intrinsic geometric structure of the learned feature space. Specifically, the Silhouette coefficient [49] and the Davies–Bouldin Index [49] were employed. The Silhouette coefficient measures the compactness and separation of clusters by comparing intra-cluster cohesion with inter-cluster dissimilarity, whereas the Davies–Bouldin Index evaluates the average similarity between each cluster and its most similar counterpart, with lower values indicating better-defined clusters.

Furthermore, t-distributed Stochastic Neighbor Embedding (t-SNE) [57] is employed to qualitatively examine the structural organization of the latent feature space and to visually interpret the clustering patterns obtained from Mini-Batch K-Means. While ARI, NMI, and FMI quantify alignment with reference labels, and Silhouette and DBI assess geometric cluster quality, t-SNE provides an intuitive visualization of the local neighborhood structure in the learned latent representations. To further facilitate interpretation, the t-SNE projections are color-coded according to the cluster assignments obtained from Mini-Batch K-Means, enabling a direct visual comparison between the learned feature representation and the unsupervised clustering results. This process is illustrated in Figure 4, which summarizes how latent feature extraction, unsupervised clustering, and t-SNE visualization jointly reveal the underlying degradation stages. Well-separated and consistently colored cluster regions therefore indicate a well-structured latent space.

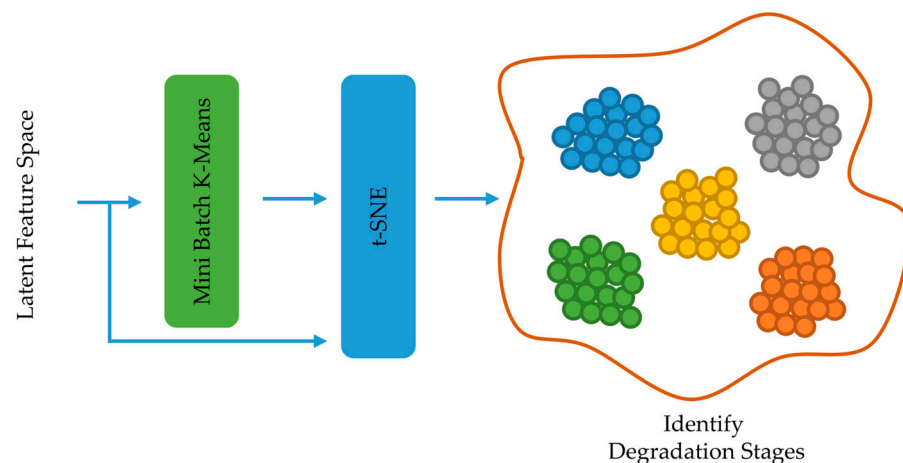


Figure 4. Visualization of degradation stage formation through Mini-Batch K-Means clustering of the latent feature space and its two-dimensional projection using t-SNE. Each colored cluster represents a distinct degradation stage.

To enable a direct, sample-wise comparison between the unsupervised cluster assignments and the reference degradation stages, the predicted cluster labels are post-processed using the Hungarian matching [58,59] algorithm. Because clustering labels are inherently

permutation-invariant, this alignment procedure is used to establish an optimal one-to-one mapping between predicted cluster labels and pseudo-class labels for interpretability purposes. Following this alignment, the correspondence between the predicted and reference labels is visualized through a confusion matrix. This representation provides an interpretable, class-level summary of clustering performance, highlighting both the dominant correctly identified stages and the cross-stage ambiguities.

All computational experiments are conducted on a computer system equipped with a twelfth-generation Intel Core i7 processor (Intel Corporation, Santa Clara, CA, USA) and an NVIDIA GeForce RTX 3060 graphics unit (NVIDIA Corporation, Santa Clara, CA, USA).

3.4. Algorithm of the Proposed Framework

To provide a clear and reproducible description of the proposed framework, the overall methodology is summarized in Algorithm 1.

Algorithm 1. The Proposed Framework

Input:

Raw vibration measurement $X = \{x_1, x_2, x_3 \dots x_N\} \in \mathbb{R}^N$

Output:

Cluster labels C ,

Evaluation metrics (ARI, NMI, FMI, Silhouette, Davies–Bouldin Index),

t-SNE projections of the latent feature space,

the training time of the LSTM model

the inference time of the LSTM model,

the inference time of the Mini-Batch K-Means

1: Compute kurtosis k_i and RMS for each measurement x_i , $i = 1, \dots, N$

2: Measurements satisfying $3.5 \leq k_i < 4.5$ are assigned to the baseline regime.

3: A transition from the baseline regime is declared when $k_i \geq 4.5$ persists over three consecutive measurements

4: The adaptive scaling parameter α_i is defined as the RMS-based deviation of each measurement x_i from baseline regime.

5: Each measurement x_i is assigned a pseudo-class y_i according to its (k_i, α_i) values and regime transitions are validated only if the corresponding boundary condition is satisfied for three consecutive measurements:

$$y_i = \begin{cases} \text{Baseline Regime,} & 3.5 \leq k_i < 4.5 \\ \text{Bounded Regime,} & k_i \geq 4.5 \wedge \alpha_i < 0.1 \\ \text{Transition Regime,} & 0.1 \leq \alpha_i < 0.4 \\ \text{Consistent Growth Regime,} & 0.4 \leq \alpha_i < 0.7 \\ \text{Saturation-LikeRegime,} & \alpha_i \geq 0.7 \end{cases}$$

6: Within each pseudo-class, the measurements are segmented into non-overlapping windows.

7: Within each pseudo-class, the windows are normalized.

8: Within each pseudo-class, the normalized windows are reshaped into 2D temporal maps.

9: The training and validation sets are constructed from subsets of the baseline 2D maps.

10: The test set is constructed from all 2D maps covering the entire run-to-failure data

11: The LSTM model is trained using the training and validation sets, where parameters are optimized via SGD by minimizing the Log-Cosh loss between the predicted outputs and the zero-centered baseline reference.

12: Measure the training time of the LSTM model

13: Pass the test data through the trained LSTM model to extract latent features, forming the latent feature space

14: Measure the inference time of the LSTM model

15: Apply Mini-Batch K-Means to the latent feature space to obtain cluster labels

16: Measure the inference time of the Mini-Batch K-Means

17: Compute external clustering validation metrics (ARI, NMI, FMI) by comparing cluster labels with pseudo-class labels.

18: Compute internal validation metrics (Silhouette score and Davies–Bouldin Index) using the latent feature space.

19: Apply the Hungarian algorithm to obtain the best cluster-to-stage matching

20: Project the latent feature space into a two-dimensional space using t-SNE for visualization purposes.

21: Compute the confusion matrix by optimally aligning cluster labels with pseudo-class labels.

4. Experimental Validation and Case Studies

This section presents the experimental validation of the proposed framework through multiple case studies designed to evaluate its effectiveness across distinct yet related lubrication-driven degradation regimes. The aim is to assess the capability of the proposed approach to consistently model degradation processes under varying lubrication conditions.

In particular, the evaluation is conducted across two complementary scenarios. The first case study is based on an industrial run-to-failure lubrication degradation dataset, where degradation evolves naturally due to the progressive loss of lubrication during real operating conditions. The second case study considers an experimental run-to-failure gear wear dataset [60,61] obtained under dry operating conditions, representing a controlled setting in which degradation is dominated by direct mechanical wear in the absence of lubrication. Together, these case studies span a broad spectrum of lubrication conditions, ranging from naturally evolving lubrication degradation in industrial environments to wear-dominated degradation under controlled dry conditions. This enables a comprehensive evaluation of the robustness and generalization capability of the proposed framework across different degradation behaviors.

For each case study, the data are processed following the methodology described in Section 3, and the proposed model is applied to extract latent representations and identify degradation patterns. The performance is evaluated using the quantitative metrics introduced earlier, while latent space visualization is employed as a complementary qualitative tool to support the interpretation of the learned feature representations and clustering behavior.

4.1. Case Study I: Industrial Run-to-Failure Lubrication Loss Dataset

4.1.1. System Description and Data Acquisition for Case Study I

The monitored system was a motor-driven Roots-type positive displacement compressor. Mechanical power was supplied by a 50 hp, 1400 rpm variable frequency drive motor that enabled speed regulation during normal service operation. Torque transmission between the blower drive shaft and the blower-driven shaft was provided by a spur gear pair located at the end of the compressor. Both gears contained 58 teeth and rotated at a one-to-one speed ratio, ensuring synchronized motion of the twin-shaft assembly throughout the operational cycle. The vibration data from the gearbox system were collected at a sampling frequency of 23.438 kHz, with approximately 376 separate measurement records acquired, each consisting of four seconds of continuous data during the run-to-failure operation. Figure 5 presents the long-term vibration measurements collected throughout the entire run-to-failure operation.

As shown in Figure 5, as the lubrication progressively degrades, a clear upward trend in vibration energy becomes evident, reflecting the deteriorating mechanical condition of the gearbox system. In addition to this consistent degradation trajectory, a small number of measurement records exhibit abrupt amplitude spikes that deviate from the natural progression observed in the signal. These atypical portions originate from transient disturbances or sensor malfunctions rather than lubrication-related deterioration. To ensure a reliable and unbiased learning process focused on physically meaningful dynamics, these corrupted measurements were identified and excluded prior to further analysis.

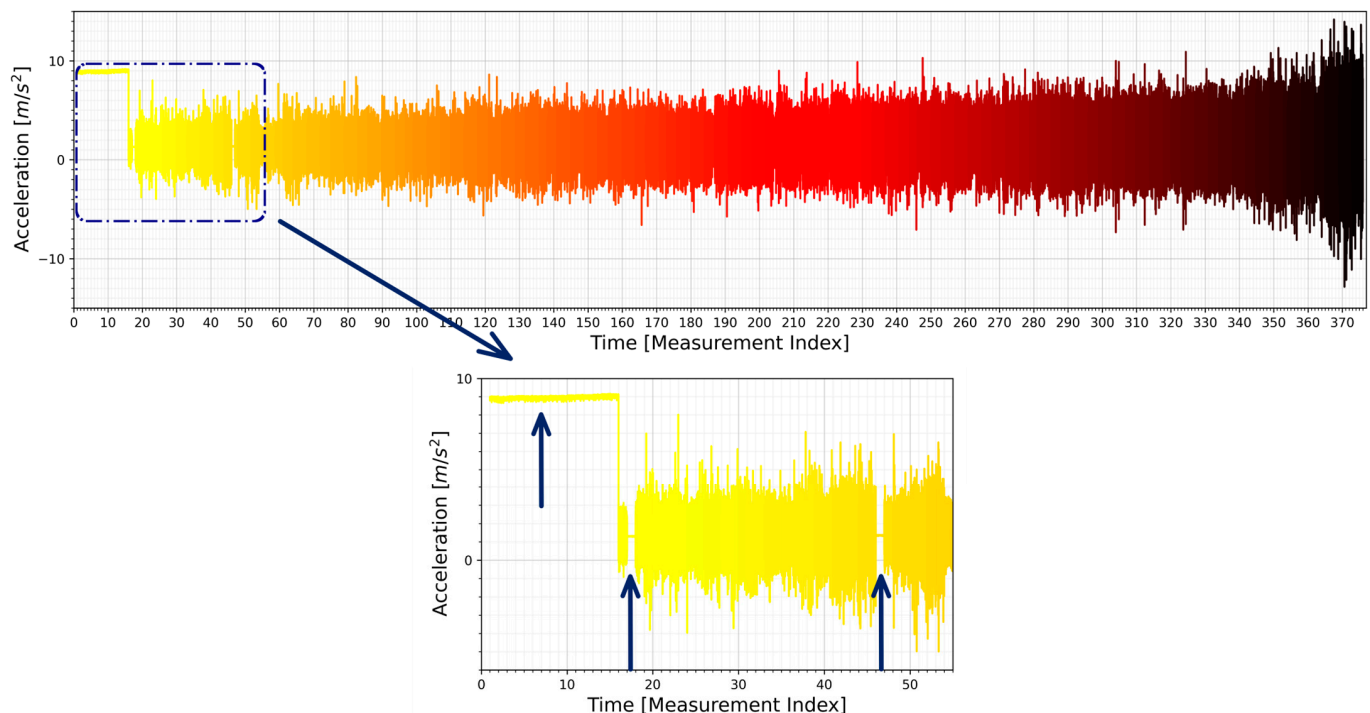


Figure 5. Vibration measurements recorded throughout the run-to-failure test, showing the progressive increase in vibration amplitude due to lubrication degradation. Corrupted records with abnormal spikes were removed from the dataset. The color gradient from yellow to dark red/black represents the progressive increase in fault severity over successive measurement indices. The inset provides a magnified view of the initial measurement region, where the arrows indicate erroneous (invalid) measurements that were excluded from further analysis.

4.1.2. Data Preprocessing and Feature Representation for Case Study I

To prepare the raw sensor data for effective feature learning and to preserve degradation-related characteristics, a structured preprocessing pipeline is employed. In this context, the adaptive class-wise scaling strategy introduced in Section 3 is first applied to normalize the data while maintaining class-specific statistical properties, thereby enhancing the representation of subtle variations associated with different degradation stages.

As an initial step of this strategy, the pseudo-class (stage) boundaries are determined based on key statistical indicators and the adaptive scaling parameter α . The temporal evolution of the key statistical indicators, kurtosis and RMS, along with the α , is depicted in Figure 6, highlighting their distinct yet complementary roles in characterizing the progression of system degradation over the run-to-failure process. Since Case Study I was conducted using industrial compressor data acquired under real operating conditions, the measured data were inherently affected by environmental and operational noise. To reduce the risk of spurious stage transitions caused by isolated boundary exceedances, a stage transition was confirmed only when the corresponding indicator remained consistently above the relevant boundary for three consecutive observations. This temporal consistency criterion enhances the robustness of stage assignment by suppressing transient fluctuations that do not reflect a genuine degradation progression.

Among key statistical indicators, kurtosis provides the earliest response to the onset of degradation. This is because kurtosis is highly sensitive to impulsive and non-Gaussian components in the vibration signal, which typically emerge during the incipient stage of fault development.

At the beginning of the degradation process, localized surface irregularities, intermittent asperity contacts, or micro-scale damage mechanisms generate transient shock-like

events. These isolated impulsive components cause a noticeable increase in kurtosis even when the overall signal energy remains relatively unchanged. However, as the fault progresses, the vibration response gradually shifts from isolated impulsive behavior to more frequent and temporally widespread excitation. In this stage, repeated impacts, sustained friction, and widespread surface deterioration produce a signal in which energy is no longer concentrated in a limited number of sharp transients but is instead spread more broadly over time. As a consequence, the impulsiveness of the signal becomes less pronounced in a statistical sense, and the discriminative power of kurtosis correspondingly decreases. Therefore, although kurtosis is highly effective for detecting the early onset of degradation, its sensitivity tends to diminish as the fault evolves into a more advanced and continuously active regime.

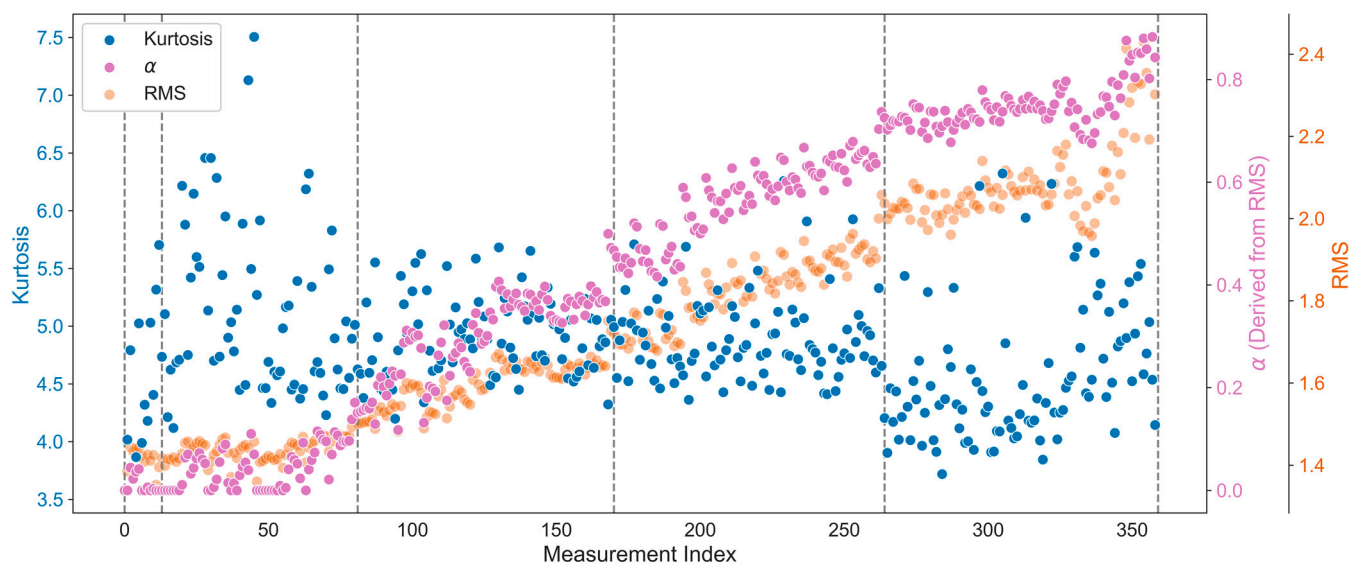


Figure 6. Temporal evolution of kurtosis, RMS, and the RMS-derived α parameter during the run-to-failure process, illustrating the progression of degradation and the corresponding stage boundaries (S1–S5) indicated by gray dashed lines.

The baseline region is defined as all observations preceding the first occurrence of three consecutive measurements in which the kurtosis value exceeds 4.5. Within this region, kurtosis does not exhibit sustained growth, indicating a quasi-stationary regime with no persistent impulsive escalation, despite remaining above the ideal Gaussian reference level. It should be noted that, in the considered real industrial dataset, the vibration signal does not exhibit a strictly healthy regime, as kurtosis values exceed 3.5 from the initial stage. Therefore, the baseline is defined in a relative sense, corresponding to the lowest observable degradation level rather than an ideal fault-free condition. In addition, the RMS values remain relatively stable, showing no significant increasing trend, which further supports the characterization of this region as a baseline regime.

Following the identification of the baseline region, the system states are defined such that the baseline regime corresponds to Stage S1, and the progression of degradation is represented by Stages S2, S3, S4, and S5. The boundaries between these stages are clearly shown in Figure 6 as gray dashed lines.

Beyond the baseline region, the determination of stage boundaries is governed by the behavior (i.e., increasing trend) of the adaptive scaling parameter α rather than kurtosis. As observed in Figure 6, in the first stage following the baseline, defined as Stage S2, kurtosis exhibits pronounced increases with intermittent high peaks; however, these variations are irregular and do not form a sustained or consistent trend. In contrast, both

RMS and α remain confined within a relatively narrow band without exhibiting a clear increasing trend.

The transition to Stage S3 is identified when α exceeds 0.1 for three successive observations. At this point, α transitions from a bounded regime to a state characterized by a clear increasing trend, albeit with local fluctuations, indicating the onset of systematic deviation from baseline conditions.

As observed in Figure 6, the transition from Stage S3 to Stage S4 occurs when the α parameter exceeds a threshold of 0.4. In Stage S3, α exhibits an increasing trend with noticeable local fluctuations, indicating an emerging but not yet fully established degradation pattern. In contrast, once α surpasses 0.4, the trend becomes more pronounced and consistent, reflecting a more stable and intensified deviation from baseline conditions. This transition is further supported by the behavior of RMS, which shows a more evident and sustained increase in Stage S4 compared to Stage S3, indicating that the energy associated with the vibration signal has become systematically elevated.

In Stage S5, the α parameter remains at elevated levels and continues to increase; however, the slope becomes noticeably reduced compared to Stage S4, exhibiting a tendency toward saturation with local fluctuations. The transition to Stage S5 is defined when the α parameter reaches the threshold of 0.7 for three successive observations.

Overall, the degradation process is characterized by the evolution of the α parameter, which transitions from a bounded regime to a progressively increasing trend and ultimately exhibits saturation-like behavior.

To quantitatively support the proposed stage segmentation, the statistical properties of the α parameter were evaluated for each stage using four metrics: the mean value ($Mean_\alpha$), standard deviation (Std_α), linear slope ($Slope_\alpha$) [62], and coefficient of determination (R_α^2) [62]. The $Mean_\alpha$ represents the average level of deviation from the baseline within each stage, while the Std_α reflects the variability of α , indicating the degree of fluctuation. The $Slope_\alpha$ provides a measure of the trend of α over time and captures the increasing trend of degradation within each stage. Finally, the R_α^2 evaluates the consistency of the increasing trend, reflecting how reliably α follows a linear trend within each stage.

The statistical results presented in Table 2 confirm the validity of the proposed stage definitions. The $Mean_\alpha$ values increase monotonically from S1 to S5, demonstrating a clear separation between the stages and confirming the progressive nature of degradation.

With regard to trend behavior, the $Slope_\alpha$ reaches its maximum value in Stage S3, indicating that this stage corresponds to the most active phase of degradation growth. In Stage S4, although the slope is slightly reduced, the higher R_α^2 value indicates that the increasing trend becomes more stable and consistent.

Table 2. Stage-wise statistical characteristics of the α parameter and corresponding reference stage intervals (S1–S5).

Stage	Measurement Index Range	The Stage Transition Rule	$Mean_\alpha$	Std_α	$Slope_\alpha$	R_α^2	Description
S1	0–12	$3.5 \leq \text{kurtosis} < 4.5$	0.0117	0.0176	−0.0018	0.1656	Baseline Regime
S2	13–80		0.0445	0.0429	0.0012	0.3058	Bounded Regime
S3	81–169	$0.1 \leq \alpha < 0.4$	0.3003	0.0838	0.0027	0.7145	Transition Regime
S4	170–263	$0.4 \leq \alpha < 0.7$	0.5605	0.0748	0.0025	0.7900	Consistent Growth Regime
S5	264–359	$\alpha \geq 0.7$	0.7481	0.0461	0.0010	0.3900	Saturation-Like Regime

In contrast, in Stage S5, both the $Slope_\alpha$ and R_α^2 decrease, indicating a reduction in the growth rate and increased variability, which is consistent with a saturation-like behavior. This observation is further supported by the Std_α , which peaks in Stage S3 and decreases

in subsequent stages, suggesting that the system transitions from a highly dynamic regime to a more stabilized but elevated degradation state.

As presented in Table 2, each stage implicitly defines a pseudo-class corresponding to a specific degradation-related behavioral regime. The preprocessing framework introduced in Section 3 is then applied, where the signals are segmented into non-overlapping windows corresponding to one complete shaft revolution and subsequently normalized using class-wise min–max scaling within each pseudo-class.

The window length is determined based on the sampling rate and rotational speed of the system. Based on a sampling rate of 23.438 kHz and a shaft rotational frequency of ~23 Hz, a full rotation theoretically contains ~1019 samples. To standardize processing and maintain computational efficiency, the window size was set to 1024 samples, representing a negligible deviation while achieving full compatibility with deep learning batch operations.

Subsequently, as described in Section 3, each normalized 1D window of 1024 samples is reshaped into a 2D temporal map with an equivalent number of elements, enabling the extraction of degradation-sensitive spatio-temporal features. To investigate the effect of temporal map configuration on deep feature learning and clustering performance, five reshaping configurations are considered, representing progressively increasing emphasis on temporal continuity: 32×32 , 64×16 , 128×8 , 256×4 and 512×2 .

To provide an intuitive illustration of how this representation reflects degradation-related changes in the vibration signal, Figure 7 presents example 32×32 temporal maps extracted from multiple stages of the run-to-failure process.

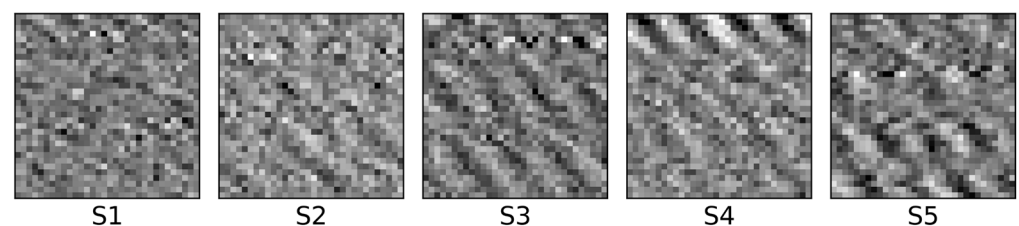


Figure 7. Representative 32×32 temporal maps showing the evolution of vibration patterns across different stages (S1–S5).

Following the construction of the 2D temporal maps, the dataset is partitioned into training, validation, and test subsets. For LSTM training, only data from the baseline regime are utilized to ensure that the model learns the intrinsic characteristics of baseline operating conditions. Specifically, the first 10 measurement records within the baseline regime—each consisting of 91 2D temporal maps—are used for training, resulting in a total of 910 samples. For validation, the subsequent two measurement records from the baseline regime are employed, corresponding to 182 2D temporal maps. The trained model is then evaluated on the complete run-to-failure dataset, which is used as the test set and comprises 32,669 2D temporal maps.

4.1.3. Experimental Procedure for Case Study I

The experimental procedure is structured to assess the effectiveness of the proposed framework under realistic operating conditions, where only baseline data is available during training. Following the learning phase, the model is applied sequentially to the entire run-to-failure dataset, enabling a continuous mapping of the degradation process in the latent feature space. Model parameters are optimized using the training and validation subsets, while the test set is utilized to analyze how the learned latent representations evolve across the full degradation trajectory under previously unseen conditions. It is important to emphasize that no explicit stage labels or prior assumptions regarding degradation levels are incorporated at any stage of the analysis. Instead, the objective is to examine whether

the progression of lubrication loss can be inherently organized into distinct states based solely on the data-driven representations learned by the model.

The application of the proposed framework yields a sequence of latent feature representations that encode the temporal evolution of the system under progressive degradation. In order to interpret these representations, a comprehensive analysis is conducted by examining both their structural organization in the latent space and the clustering patterns that emerge from the vibration data. Particular attention is given to whether the learned feature space reflects a meaningful separation of degradation-related behavioral regimes and preserves consistency with the internally constructed reference degradation stages. The results presented in the following section therefore focus on evaluating the interpretability of the latent representations and the ability of the unsupervised clustering to reveal distinct behavioral regimes associated with lubrication loss.

4.1.4. Results and Degradation Analysis for Case Study I Spectral Characteristics and Tribological Interpretation of Degradation

It should be noted that frequency domain analysis is not part of the proposed diagnostic framework and is employed solely for validation purposes to confirm the presence of lubrication loss in the collected dataset. To this end, the vibration responses are examined in both the time and frequency domains to provide qualitative evidence of degradation behavior.

To provide physical context for the frequency domain features, the concept of gear mesh frequency and its relevance to lubrication conditions is briefly outlined below.

In geared systems, the gear mesh frequency corresponds to the fundamental rate at which gear teeth engage and is determined by the product of the shaft rotational frequency and the number of gear teeth, expressed as

$$f_{mesh} = N \times f_{shaft} \quad (16)$$

where N denotes the number of teeth and f_{shaft} represents the shaft rotational frequency. Changes in lubrication conditions alter the contact mechanics at the gear interface, often leading to increased mesh frequency harmonics and elevated high-frequency vibration content, which are commonly associated with lubricant film breakdown.

Figure 8 illustrates the evolution of vibration responses under three distinct operating regimes, namely the baseline regime, the consistent growth regime, and the saturation-like regime, in both time and frequency domains. A clear and progressive transition is observed across these stages, reflecting the underlying degradation dynamics associated with lubrication loss.

In the baseline regime, the time-domain signal is characterized by low-amplitude oscillations with a relatively consistent temporal structure, accompanied by minor fluctuations and low-level variability. Impulsive events are infrequent and remain temporally localized, indicating that irregular disturbances are limited and do not significantly affect the overall signal structure. Notably, the presence of small irregularities suggests that the system may already exhibit characteristics associated with incipient degradation. Correspondingly, in the frequency domain, the spectrum remains relatively sparse, although several low-amplitude components are present. These components are not strongly amplified and do not exhibit pronounced harmonic structures or a wide distribution across the frequency range. This spectral pattern suggests that meshing-related excitations are present but not dominant, and that strong dynamic disturbances have not yet developed. From a tribological perspective, this regime suggests that lubrication is still relatively effective in limiting severe surface interactions, although early-stage degradation effects may already be present. The absence of strongly amplified frequency components and the limited

irregularity observed in the time-domain response indicate that friction-induced impacts and extensive asperity contact are not yet dominant.

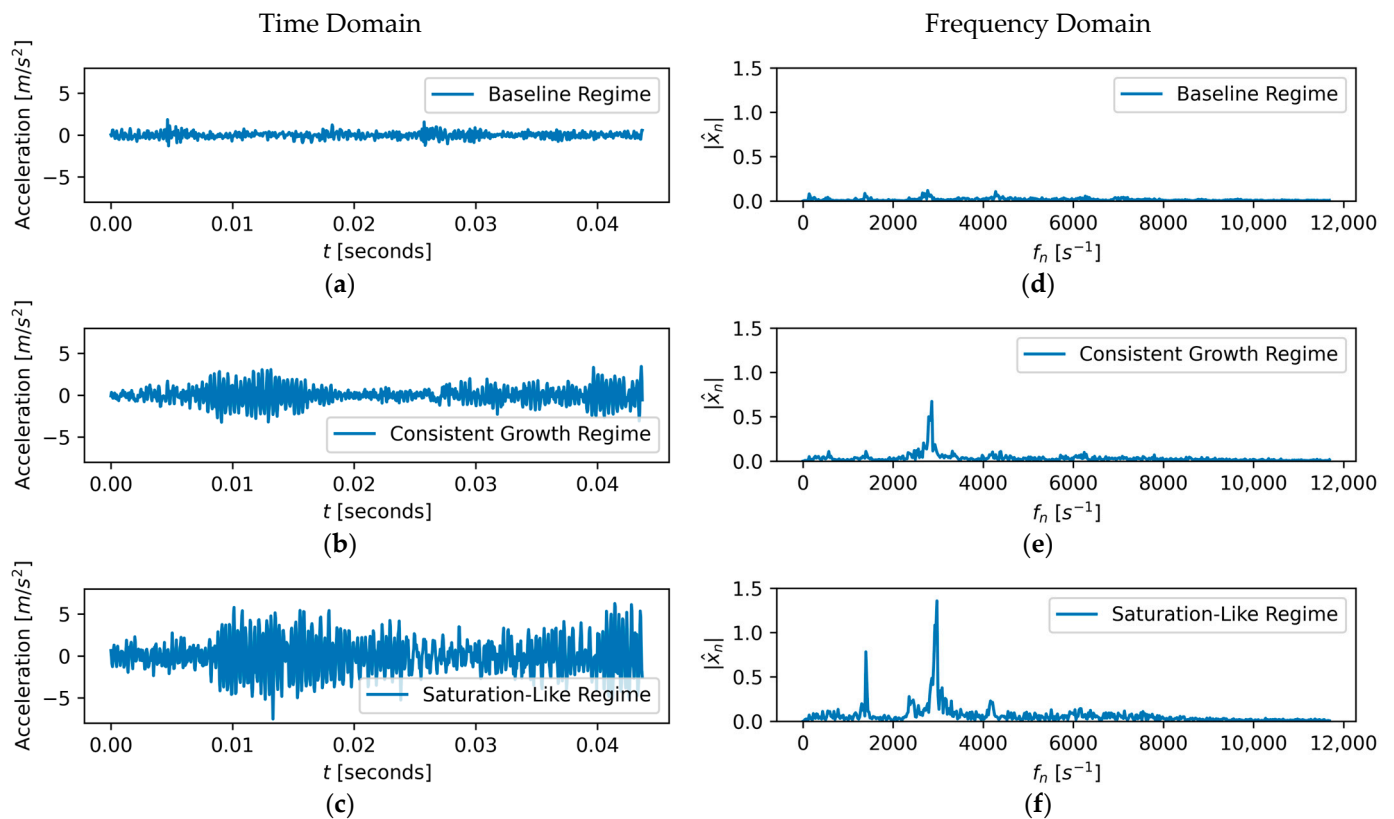


Figure 8. Evolution of vibration responses across different degradation regimes in both time and frequency domains: (a,d) baseline, (b,e) consistent growth and (c,f) saturation-like regimes.

As the system transitions into the consistent growth regime, the time-domain signal exhibits increased amplitude and more pronounced irregular fluctuations compared to the baseline condition. Impulsive events become more frequent and extend over a broader portion of the time axis, indicating that excitation events are no longer isolated but increasingly influence the overall temporal response. Notably, this change reflects a progressive deviation from the initial temporal structure, suggesting that degradation-related effects are becoming more prominent. In the frequency domain, this transition is characterized by the amplification of spectral components near the mesh frequency and its harmonics, accompanied by the appearance of additional frequency components over a wider range. The spectrum becomes more densely populated, reflecting the coexistence of dominant periodic excitations and additional dynamic disturbances. From a tribological perspective, this regime may be associated with a transition toward mixed lubrication conditions, where partial degradation of the lubricant film leads to increased asperity interactions. These interactions can introduce additional excitation sources, contributing to the amplification of harmonic components and the emergence of additional frequency components.

In the saturation-like regime, the time-domain signal is dominated by high-amplitude fluctuations and pronounced irregularities. Impulsive events are highly frequent and temporally widespread, occupying a substantial portion of the signal and leading to a highly disturbed temporal structure. Notably, the distinction between individual events becomes less clear, indicating that excitation mechanisms are no longer intermittent but pervasive throughout the signal. Correspondingly, in the frequency domain, the spectrum exhibits strongly amplified components, particularly around the mesh frequency and

its harmonics, along with the presence of additional frequency components extending over a wider range. Compared to earlier regimes, the spectrum appears more densely populated, reflecting the contribution of multiple excitation sources. From a tribological perspective, these observations may be associated with advanced lubrication degradation, where the lubrication condition approaches boundary regimes and asperity interactions become more pronounced. Such conditions can lead to increased frictional interactions, intermittent impacts, and localized stress concentrations, which are reflected in the irregular time-domain response and the presence of multiple amplified frequency components.

Overall, the observed progression across the three regimes suggests a gradual transition from localized and infrequent disturbances to highly frequent and temporally widespread excitation behavior. From a physical and tribological perspective, this evolution may be associated with progressively degraded lubrication conditions, consistent with increased frictional activity and intermittent surface interactions. However, it should be emphasized that vibration-based observations alone are not sufficient to uniquely distinguish between specific tribological failure mechanisms, particularly in the absence of direct physical or surface-level evidence.

Effect of Temporal Map Configuration on Latent Representation and Clustering Performance for Case Study I

To systematically investigate the influence of the 2D temporal-map configuration on the quality of the learned latent representations and the subsequent clustering performance, five alternative reshaping schemes (32×32 , 64×16 , 128×8 , 256×4 , and 512×2) were evaluated. For all configurations, the LSTM architecture, training protocol, and Mini-Batch K-Means parameters were held constant to ensure a controlled comparison, allowing performance variations to be primarily attributed to differences in the underlying temporal representation. Table 3 presents the clustering performance and computational efficiency for the evaluated temporal-map configurations, as assessed using both external (ARI, NMI, and FMI) and internal (Silhouette and Davies–Bouldin) validation metrics, along with the corresponding LSTM training and inference times and Mini-Batch K-Means clustering time.

Table 3. Comparison of Clustering Performance and Computational Efficiency Across Different Temporal-Map Configurations for Case Study I.

Temporal Map Configuration	32×32	64×16	128×8	256×4	512×2
ARI	0.86	0.94	0.99	1.0	1.0
NMI	0.89	0.95	0.99	1.0	1.0
FMI	0.89	0.96	0.99	1.0	1.0
Silhouette	0.32	0.40	0.44	0.47	0.44
Davies–Bouldin	2.53	1.82	1.10	0.96	0.98
LSTM training time (baseline data) [s]	5.73	5.99	8.56	11.84	19.90
LSTM inference time (the entire data) [s]	0.29	0.48	0.74	1.41	2.68
Mini-Batch K-Means clustering time (the entire data) [s]	0.28	0.30	0.37	0.51	0.83
LSTM output shape	32×4	64×4	128×4	256×4	512×4
Flattened latent feature dimension	128	256	512	1024	2048

The results presented in Table 3 demonstrate that the temporal-map configuration has a substantial impact on both clustering performance and computational efficiency. As the representation transitions from compact to more temporally extended configurations, a consistent improvement is observed in the external clustering metrics (ARI, NMI, and FMI), indicating that increasing temporal resolution enhances the discriminative capability of the learned latent space. In particular, the 256×4 configuration achieves perfect agreement with the reference degradation stages (ARI = NMI = FMI = 1.0), while also providing the most favorable internal validation scores among all evaluated configurations. The observed

improvement in clustering performance is also consistent with the increase in latent feature dimensionality induced by higher temporal resolutions.

Nevertheless, the moderate Silhouette and Davies–Bouldin values suggest that the latent space does not exhibit strongly separated cluster structures, but rather reflects a smoothly varying structure. This behavior is physically consistent with lubrication-loss degradation, where the underlying fault mechanism remains unchanged and progressively intensifies over time, resulting in a continuous system evolution. Accordingly, this evolution can be interpreted as the partitioning of an underlying degradation manifold, primarily governed by the physical process and further supported by the adopted adaptive normalization strategy. Therefore, although the clustering results demonstrate near-perfect agreement with the reference degradation stage, this agreement is more appropriately interpreted as the partitioning of an underlying continuous degradation manifold, rather than evidence of inherently disjoint and well-separated clusters. Although the 512×2 configuration achieves performance results largely comparable to those of the 256×4 configuration, it is associated with an observable increase in computational cost. Overall, these findings indicate that temporal continuity is the primary factor governing the effectiveness of the learned representation, and that 256×4 constitutes an optimal resolution for practical deployment.

The quantitative clustering results presented in Table 3 are complemented by the t-SNE visualizations shown in Figure 9, which provide a qualitative interpretation of the latent feature distributions across the evaluated configurations. All evaluated configurations exhibit structurally organized latent representations with visually distinguishable groups. However, the correspondence between the visually separated regions and the predicted cluster assignments differs across configurations. In the 32×32 and 64×16 configurations, certain visually distinguishable groups are assigned to the same cluster label, suggesting that the clustering algorithm does not fully recover the underlying latent grouping structure in these configurations. In contrast, the 256×4 configuration demonstrates a more consistent correspondence between the predicted cluster assignments and the more visually distinguishable latent groups. Although the 512×2 configuration produces more compact clusters, the overall latent structure remains similar to that of the 256×4 configuration, with the primary difference being a denser cluster geometry.

The confusion matrices shown in Figure 10 indicate a progressive improvement in clustering consistency with increasing temporal resolution. In the lower-resolution configurations (32×32 and 64×16), noticeable misclassifications are observed between several degradation stages. In particular, a substantial portion of the S1 samples is incorrectly assigned to the S5 cluster, while a considerable number of S3 samples are assigned to the S1 cluster in the 32×32 configuration. Although the misclassification level decreases in the 64×16 configuration, confusion between S1 and S5 still persists, indicating that the clustering algorithm does not fully recover the underlying latent grouping structure at lower temporal resolutions. In contrast, the 128×8 configuration achieves near-perfect alignment, while the 256×4 and 512×2 configurations provide complete correspondence between the predicted cluster assignments and the reference degradation stages.

The results are consistently supported by quantitative metrics, confusion matrices, and t-SNE visualizations, which collectively confirm the effectiveness of the proposed framework in capturing the progression of degradation. Overall, the 256×4 configuration provides the most favorable overall performance in terms of clustering consistency, latent representation quality, and computational efficiency, and is therefore identified as the most suitable configuration among the evaluated settings. Accordingly, the comparative evaluation of alternative clustering algorithms was conducted using the 256×4 latent representation, allowing the influence of the clustering strategy to be analyzed under the most computationally efficient and structurally stable configuration identified in this study.

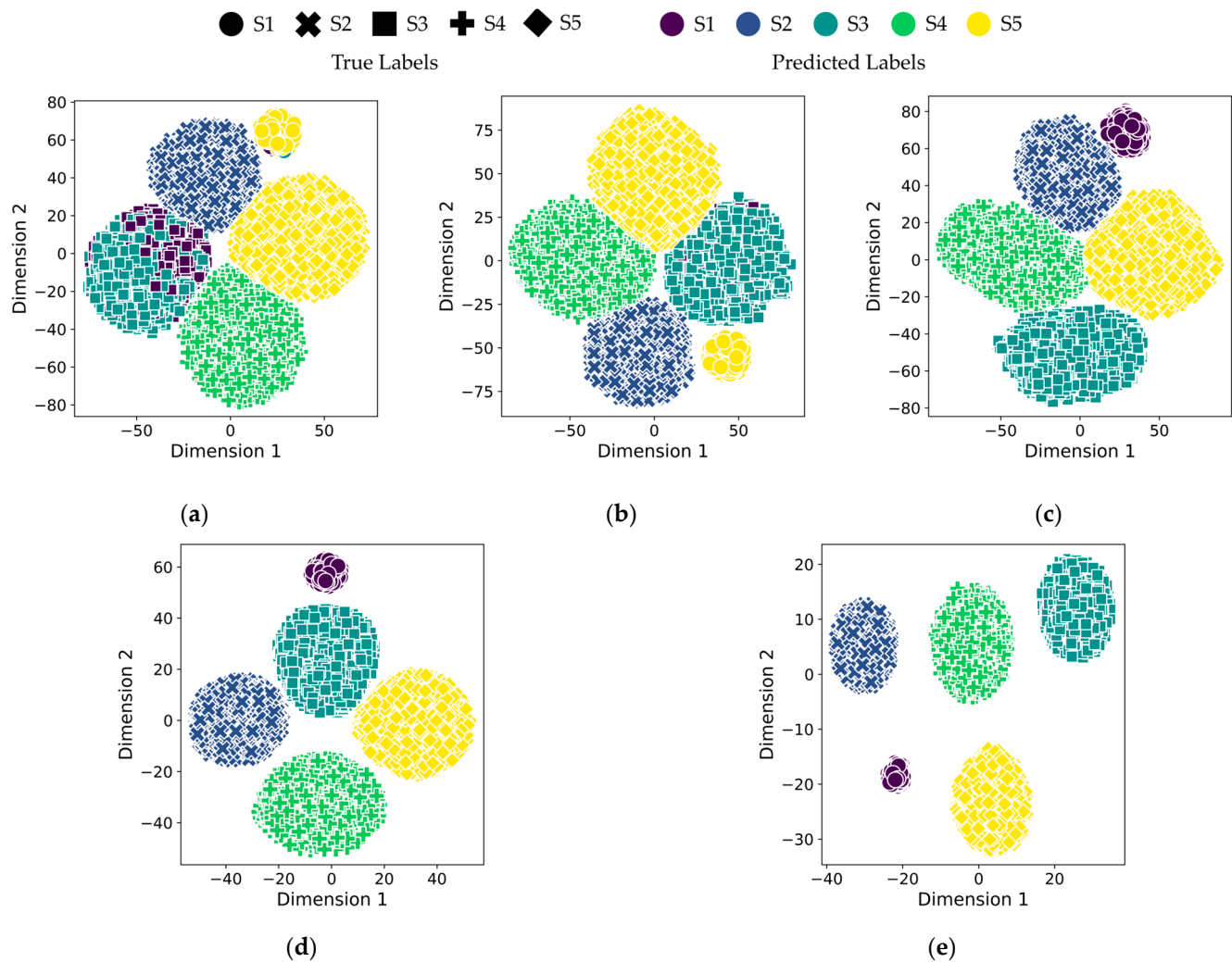


Figure 9. t-SNE visualization of the latent feature space showing both the discriminative structure of the learned representations and the cluster assignments produced by Mini-Batch K-Means. Subfigures correspond to different temporal-map configurations: (a) 32×32 , (b) 64×16 , (c) 128×8 , (d) 256×4 , (e) 512×2 .

Performance Comparison with Alternative Clustering Algorithms for Case Study I

To assess the robustness, efficiency and generalizability of the proposed latent space clustering framework, several widely used unsupervised algorithms were evaluated alongside Mini-Batch K-Means. Specifically, Birch, Bisecting K-Means, standard K-Means, Spectral Clustering and Agglomerative Clustering were applied to the same set of latent feature vectors extracted from the LSTM. These methods collectively span centroid-driven, hierarchical and graph-theoretic clustering paradigms, thereby providing a comprehensive methodological spectrum for examining how different clustering principles interact with the geometry of the learned latent representations. All clustering algorithms were evaluated using their default parameter settings, except for the Birch method, for which the threshold parameter was set to 0.05 to ensure stable cluster formation within the learned latent space.

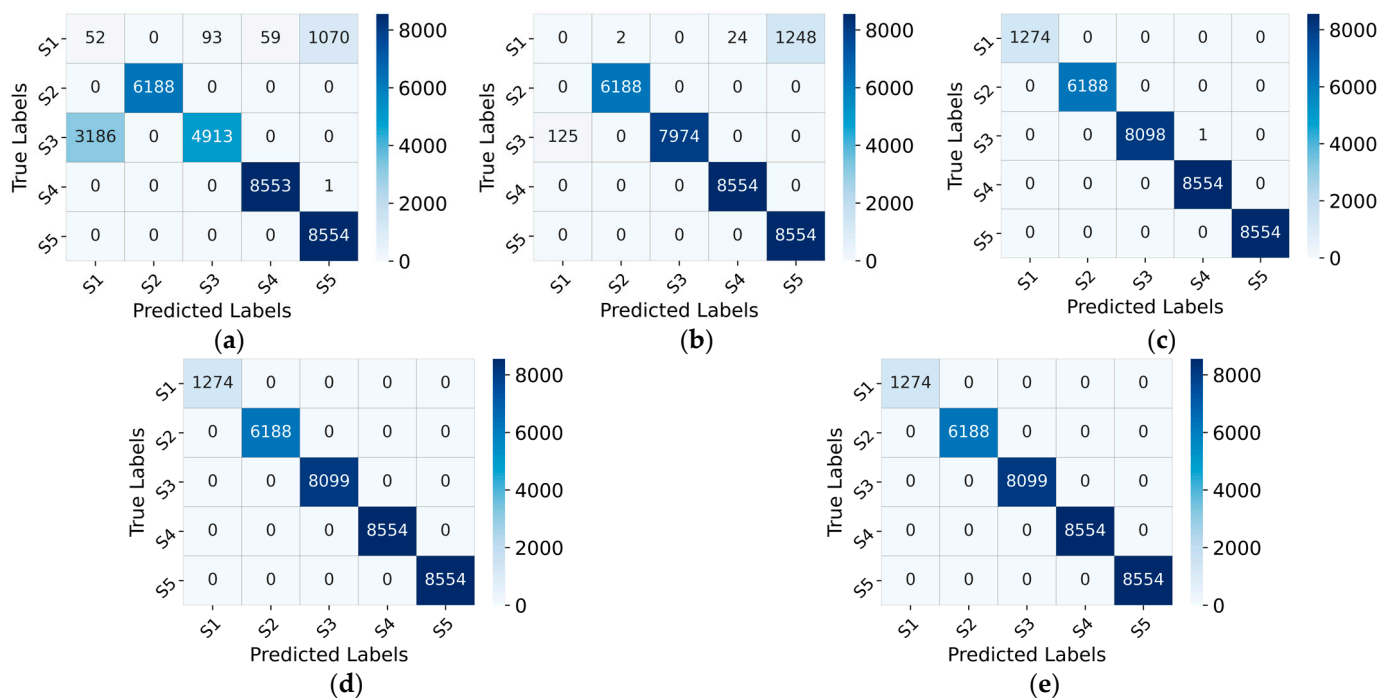


Figure 10. Confusion matrices illustrating the alignment between predicted degradation stages and the reference labels for different temporal map configurations: (a) 32×32 , (b) 64×16 , (c) 128×8 , (d) 256×4 , (e) 512×2 .

The comparative analysis of multiple unsupervised algorithms is conducted using the 256×4 2D map configuration that yields the most favorable latent representation quality, as summarized in Table 4, enabling a balanced evaluation of both clustering fidelity and computational cost. The table reports the clustering outcomes in terms of ARI, NMI, FMI, Silhouette, and Davies–Bouldin metrics, together with the corresponding computational times. The results show that most clustering algorithms achieve near-perfect external validation scores (ARI, NMI, FMI ≈ 1.0), indicating strong agreement with the reference degradation stages. However, the moderate Silhouette (≈ 0.47) and Davies–Bouldin values (≈ 0.96) indicate limited cluster separability with a tendency toward gradual structural transitions, suggesting that the latent space reflects a progressive degradation process rather than sharply defined states.

Table 4. Comparison of Unsupervised Clustering Algorithms Using the Most Favorable 256×4 2D Map Configurations.

Unsupervised Algorithms	ARI	NMI	FMI	Silhouette	Davies–Bouldin	Clustering Time (100% Data) [s]
Birch	1.0	1.0	1.0	0.47	0.96	82.99
Bisecting K-Means	0.86	0.89	0.89	0.33	3.32	0.74
K-Means	1.0	1.0	1.0	0.47	0.96	2.65
Spectral Clustering	1.0	1.0	1.0	0.47	0.96	250.45
Agglomerative Clustering	1.0	1.0	1.0	0.47	0.96	297.87
Mini-Batch K-Means (proposed)	1.0	1.0	1.0	0.47	0.96	0.51

Despite this similarity in clustering outcomes, the algorithms exhibit substantial differences in computational cost. Spectral Clustering and Agglomerative Clustering incur very high execution times due to their graph-based and hierarchical operations, whereas standard K-Means offers moderate computational efficiency. In contrast, Mini-

Batch K-Means achieves clustering performance identical to the best-performing alternative algorithms at substantially lower computational cost, thereby emerging as the most efficient and practically viable solution.

Comparative Evaluation of Clustering Algorithms for Unsupervised Degradation Analysis Without DL Models for Case Study I

Table 5 presents a comprehensive comparative evaluation of several widely used unsupervised clustering algorithms applied directly to the original measurement space, independently of the proposed LSTM-based feature learning framework. Since conventional clustering algorithms operate on vectorized inputs and do not inherently exploit spatial dependencies, the 2D temporal map transformation is not employed in this analysis. This design enables an objective assessment of each algorithm's intrinsic capability based solely on raw and manually derived feature spaces. The clustering performance is quantitatively assessed using standard external validation metrics, including ARI, NMI, and FMI, along with internal validation metrics such as Silhouette score and Davies–Bouldin index. In addition, the computational efficiency of each method is evaluated in terms of clustering time using the full dataset.

Table 5. Comparative clustering performance of unsupervised algorithms using different feature representations without LSTM-based feature learning for Case Study I.

Unsupervised Algorithms	Features	ARI	NMI	FMI	Silhouette	Davies–Bouldin	Clustering Time (100% Data) [s]
Birch	RMS	0.63	0.74	0.75	0.60	0.44	0.26
	α	0.47	0.59	0.62	0.66	0.41	0.25
	x_{raw}	0.21	0.33	0.45	0.08	5.92	274
	x_{scaled}	0.99	0.99	0.99	0.20	1.94	272
Bisecting K-Means	RMS	0.56	0.66	0.66	0.63	0.43	0.20
	α	0.70	0.76	0.77	0.60	0.47	0.20
	x_{raw}	0.26	0.40	0.48	0.1	4.1	1.41
	x_{scaled}	1.0	1.0	1.0	0.20	1.95	1.14
K-Means	RMS	0.62	0.70	0.70	0.65	0.43	0.19
	α	0.70	0.75	0.76	0.63	0.48	0.20
	x_{raw}	0.25	0.36	0.45	0.07	4.48	14.46
	x_{scaled}	1.0	1.0	1.0	0.20	1.95	4.22
Spectral Clustering	RMS	0.49	0.62	0.63	0.36	0.49	0.30
	α	0.84	0.83	0.88	0.44	0.48	0.39
	x_{raw} *	-	-	-	-	-	-
	x_{scaled}	0.04	0.19	0.50	0.18	14.29	287
Agglomerative Clustering	RMS	0.71	0.75	0.78	0.62	0.4	0.002
	α	0.67	0.73	0.75	0.62	0.47	0.002
	x_{raw}	0.22	0.33	0.45	0.08	5.92	245
	x_{scaled}	0.99	0.99	0.99	0.20	1.95	295
Mini-Batch K-Means	RMS	0.61	0.70	0.70	0.64	0.46	0.25
	α	0.69	0.74	0.76	0.63	0.48	0.26
	x_{raw}	0.26	0.37	0.43	−0.03	5.37	0.94
	x_{scaled}	1.0	1.0	1.0	0.20	1.95	0.47

* Spectral Clustering could not be reliably performed on the raw high-dimensional representation (x_{raw}), due to computational and connectivity issues.

The analysis is conducted using four different feature representations: the statistical indicators RMS, class wise scaling parameter α , the raw high-dimensional vibration windows ($x_{raw} \in \mathbb{R}^{1024}$), and their adaptively scaled counterparts ($x_{scaled} \in \mathbb{R}^{1024}$). The RMS feature is computed for each measurement, yielding a one-dimensional representation of 356 values, while the α parameter, derived from RMS behavior, serves as a class-wise

adaptive scaling coefficient reflecting deviations from baseline conditions. On the other hand, the high-dimensional representation x_{raw} is obtained by segmenting each measurement into non-overlapping windows of length 1024, where each window corresponds to one complete machine rotation. This results in a set of 1024-dimensional vectors preserving the temporal structure of the vibration signal. The scaled representation x_{scaled} is subsequently constructed by applying the proposed class-wise adaptive scaling strategy to each 1024-dimensional window, thereby normalizing the data while preserving degradation-sensitive variations across different operating conditions.

The results obtained from Table 5 indicate that clustering performance is predominantly governed by the choice of feature representation rather than the clustering algorithm itself. This is evident from the observation that, for a fixed feature space, different clustering algorithms yield relatively small performance variations, whereas substantial changes are observed when the feature representation is altered within the same algorithm.

A comparison of the low-dimensional feature representations in Table 5 reveals that both RMS and the α parameter exhibit similar clustering behavior, characterized by relatively high Silhouette scores and low Davies–Bouldin indices across multiple algorithms. However, despite this apparent geometric separation, these representations yield lower ARI, NMI, and FMI scores, indicating limited consistency with the reference degradation stage. This behavior can be attributed to the nature of simplified feature representations, where increased numerical spread may artificially enhance cluster separation without adequately capturing the underlying degradation structure. In particular, features such as RMS and the α parameter primarily emphasize amplitude-level variations, leading to well-separated clusters that do not necessarily reflect the progression of degradation states.

In contrast, the raw high-dimensional representation (x_{raw}) consistently exhibits poor clustering performance across all algorithms, as reflected by low Silhouette scores and high Davies–Bouldin indices. This behavior indicates that the intrinsic cluster structure is not well preserved in the raw feature space, making it difficult for distance-based clustering methods to identify meaningful groupings. Meanwhile, the adaptively scaled representation (x_{scaled}) yields near-perfect clustering scores (ARI, NMI, and FMI ≈ 1.0), despite exhibiting lower Silhouette values for several algorithms. This highlights that the scaling process enhances the structural consistency of the feature space, thereby enabling a more accurate representation of degradation behavior, even when cluster separation appears reduced in a purely geometric sense. However, when compared with the LSTM-based learned representations, x_{scaled} yields lower Silhouette scores and higher Davies–Bouldin indices. This suggests that the LSTM-based feature learning further enhances the compactness and separability of the feature space beyond what can be achieved through scaling alone.

In terms of computational efficiency, the results reveal a clear dependency on both the dimensionality of the feature space and the underlying algorithmic structure. While all clustering methods exhibit negligible execution times for low-dimensional features such as RMS and α , a substantial increase in computational cost is observed when high-dimensional representations are employed. In particular, hierarchical and tree-based methods, such as Agglomerative Clustering and BIRCH, demonstrate poor scalability, with execution times increasing dramatically. Spectral Clustering, on the other hand, shows limited practical applicability due to its high computational burden and sensitivity to data structure. In particular, Spectral Clustering could not be reliably performed on the raw high-dimensional representation (x_{raw}), due to the construction of a non-fully connected affinity graph.

Overall, the results demonstrate that the primary performance gain arises from the learned feature representation, as the LSTM-based framework organizes the feature space in a more discriminative and structured manner, enabling degradation-related characteristics

to become more distinguishable while reducing overlap between clusters and improving their compactness and separability.

Comparative Evaluation of DL Models for Unsupervised Degradation Clustering for Case Study I

To systematically evaluate the effectiveness of different deep feature learning strategies for unsupervised degradation analysis, a diverse set of DL architectures was employed, including convolutional models (Conv2D, SeparableConv2D, and DepthwiseConv2D), sequence-based models including Gated Recurrent Unit (GRU) and Temporal Convolutional Network (TCN), and hybrid spatiotemporal architectures (ConvLSTM2D and LSTM with Multi-Head Attention). The architectural details and hyperparameter settings of the considered models are summarized in Table 6, while all training hyperparameters not explicitly listed were kept identical to those used in the proposed model described in Section 3 to ensure a fair comparison. In addition, each model was evaluated using its optimal 2D map configuration and a minimal network architecture to avoid unnecessary complexity. These models were employed to learn a latent feature space from the input data, and the resulting representations were subsequently clustered using the Mini-Batch K-Means algorithm to assess their discriminative capability.

Table 6. The Architectural Details and Hyperparameter Settings of DL Models Used in the Comparative Analysis.

DL Algorithms	Architecture Details	Key Hyperparameters
Conv2D	Conv2D + BatchNormalization + Flatten	Filters: 4, kernel_size: (4, 4) Padding: same activation: LeakyReLU()
SeparableConv2D	SeparableConv2D + BatchNormalization + Flatten	Filters: 4, kernel_size: (2, 2) Padding: same activation: LeakyReLU()
DepthwiseConv2D	DepthwiseConv2D + BatchNormalization + Flatten	Filters: 4, kernel_size: (2, 2) Padding: same activation: LeakyReLU()
ConvLSTM2D	ConvLSTM2D + BatchNormalization + Flatten	Filters: 8, kernel_size: (2, 2) Padding: same activation: ThresholdedReLU(theta = 0.01)
GRU	GRU + BatchNormalization + Flatten	units: 4, activation = tanh return_sequences = True
TCN	TCN + BatchNormalization + Flatten	nb_filters = 3, kernel_size = 3, dilations = (2, 8, 16), return_sequences = True
LSTM + MultiHeadAttention	LSTM + MultiHeadAttention + BatchNormalization + Flatten	units: 4, activation = tanh return_sequences = True num_heads = 1, key_dim = 3

The results obtained from each architecture are reported in terms of both clustering quality (ARI, NMI, FMI, Silhouette, and Davies–Bouldin indices) and computational efficiency (inference time), enabling a comprehensive assessment of the trade-off between representation quality and processing cost. The corresponding results are presented in Table 7. The DL column denotes the feature space construction time of the DL model (inference time for latent representation extraction), while the Cluster column represents the inference time of the clustering algorithm applied to the extracted feature space. It should be noted that latent feature dimensionality is reported for transparency as it reflects the

representation capacity induced by each model–input configuration. However, it should not be interpreted as an independent variable for performance comparison, since it is a by-product of the model architecture and input representation design.

Table 7. Performance Comparison of DL Models for Unsupervised Degradation Clustering for Case Study I.

DL Algorithms	2D Map Configurations	Latent Feature Dimension	ARI	NMI	FMI	Silhouette	Davies–Bouldin	Inference Time (100% Data) [s]	
								DL	Cluster
Conv2D	32 × 32	4096	1.0	1.0	1.0	0.40	1.10	0.55	1.39
SeparableConv2D	32 × 32	4096	1.0	1.0	1.0	0.31	1.39	0.45	1.32
DepthwiseConv2D	32 × 32	256	0.99	0.99	0.99	0.21	1.93	0.20	0.36
ConvLSTM2D	91 × 32 × 32	8192	1.0	1.0	1.0	0.51	0.80	8.45	0.57
GRU	256 × 4	1024	1.0	1.0	1.0	0.24	1.73	1.29	0.96
TCN	256 × 4	768	1.0	1.0	1.0	0.19	1.81	1.49	0.91
LSTM + MultiHeadAttention	256 × 4	1024	1.0	1.0	1.0	0.47	0.96	8.14	0.50
LSTM (proposed)	256 × 4	1024	1.0	1.0	1.0	0.47	0.96	1.41	0.51

In terms of performance evaluation, although all models achieve perfect external validation scores (ARI, NMI, and FMI equal to 1.0), the internal validation metrics and inference times reveal substantial differences in latent space quality and computational cost. ConvLSTM2D provides the most compact and well-separated clusters, as indicated by the highest Silhouette score and lowest Davies–Bouldin index, but at the expense of significantly increased inference time (8.45 s). In contrast, DepthwiseConv2D achieves the highest computational efficiency (0.20 s) while producing less discriminative representations (Silhouette = 0.21, Davies–Bouldin = 1.93).

The proposed LSTM model provides a balanced trade-off, achieving competitive clustering performance (Silhouette = 0.47, Davies–Bouldin = 0.96) with lower computational cost (1.41 s). This indicates that explicitly modeling temporal dependencies is sufficient to capture degradation-related patterns without requiring complex spatiotemporal architectures. Furthermore, the incorporation of MultiHeadAttention does not yield meaningful performance gains, as the clustering metrics remain nearly unchanged while the inference time increases dramatically (8.14 s), rendering it inefficient from a cost–benefit perspective.

Importantly, the results highlight a strong dependency between data representation and model architecture. Convolutional models perform better on compact 2D maps (e.g., 32 × 32), where local spatial patterns dominate, whereas sequence-based models (LSTM, GRU, TCN) benefit from configurations with higher temporal resolution (e.g., 256 × 4), enabling more effective modeling of progressive degradation dynamics.

Taken together, these findings demonstrate that although increased architectural complexity improves cluster separability, it also leads to a substantial increase in computational cost, which limits its practical applicability. The trade-off between clustering discriminability and computational efficiency highlights that model selection should be guided by application-specific requirements. From a practical standpoint, ConvLSTM2D is most suitable when maximizing cluster separability is the primary objective; DepthwiseConv2D is advantageous in time-critical scenarios due to its high computational efficiency, whereas the proposed LSTM model is well suited for scenarios requiring a balance between clustering performance and inference efficiency, making it a practical and scalable solution for unsupervised degradation analysis.

4.2. Case Study II: Experimental Run-to-Failure Gear Wear Dataset Under Dry Conditions

4.2.1. System Description and Data Acquisition for Case Study II

In Case Study II, to assess the effectiveness of the proposed methodology, a publicly available run-to-failure gearbox dataset is employed. The dataset is acquired from controlled experiments conducted on a spur gearbox test rig at the University of New South Wales (UNSW), Sydney, NSW, Australia [60,61], where gears with 19 and 52 teeth are primarily operated under constant load and nominal speed conditions, with additional measurements collected at varying rotational speeds. Vibration signals are recorded using high-sensitivity accelerometers at a sampling frequency of 100 kHz, along with tachometer and encoder signals for phase reference. Two operating conditions are considered, namely a dry test (10 Hz, 5 Nm) to induce abrasive wear and a lubricated test (16 Hz, 20 Nm) to promote fatigue pitting. The data are collected periodically, while the system is intermittently stopped to capture surface replicas, enabling the tracking and validation of wear evolution [60,61].

Within this case study, the analysis focuses on the dry test condition, representing a loss of lubrication scenario, where the absence of lubrication leads to accelerated abrasive wear and progressive surface degradation. For the present study, only the vibration data obtained from the B&K 4394 sensor (vib2) were utilized.

4.2.2. Data Preprocessing and Feature Representation for Case Study II

To prepare the raw sensor data for effective feature learning under dry conditions that result in progressive abrasive wear, a structured preprocessing pipeline is employed. In this context, the adaptive class-wise scaling strategy introduced in Section 3 is applied to normalize the vibration data while preserving degradation-relevant variations, thereby enhancing the representation of subtle changes throughout the degradation progression.

As an initial step, pseudo-class (stage) boundaries are determined based on key statistical indicators and the adaptive scaling parameter α . Figure 11 illustrates the temporal evolution of kurtosis, RMS, and α , highlighting their distinct yet jointly informative roles in characterizing degradation progression. The kurtosis initially reaches a pronounced peak and then gradually decreases as degradation progresses. This pattern indicates that the early stage is dominated by localized impulsive events, whereas the subsequent decline indicates the transition from localized surface interactions to more spatially distributed surface damage. In contrast, the RMS exhibits an overall increasing trend, capturing the growth in vibration energy as degradation intensifies, although noticeable fluctuations emerge at later stages. Meanwhile, the adaptive scaling parameter α increases rapidly during the transition from localized impulsive behavior to more widespread surface degradation and subsequently saturates, providing a more stable and robust representation of the progressive deviation from the baseline condition. This contrast highlights the ability of α to capture the underlying degradation trend more consistently, despite the variability observed in RMS.

The pseudo-class (stage) boundaries are defined using α -based threshold values ($\alpha = 0.1, 0.4$, and 0.7), which are directly adopted from Case Study I without any modification, ensuring methodological consistency across different case studies. Nevertheless, the three successive observation criteria used in Case Study I were omitted because the controlled experimental conditions and low noise levels in Case Study II reduced the need for additional temporal confirmation of threshold crossings. It is observed that the lower α threshold ($\alpha = 0.1$) is not activated in this case study, and similarly, the kurtosis does not exceed the higher threshold level (e.g., 4.5), preventing the corresponding transition from being triggered. This indicates that the degradation process in this dataset does not fully span the range defined by the proposed stage framework. It should also be noted that the

publicly available dataset represents only a subset of the full run-to-failure experiment, which may limit the observable range of degradation.

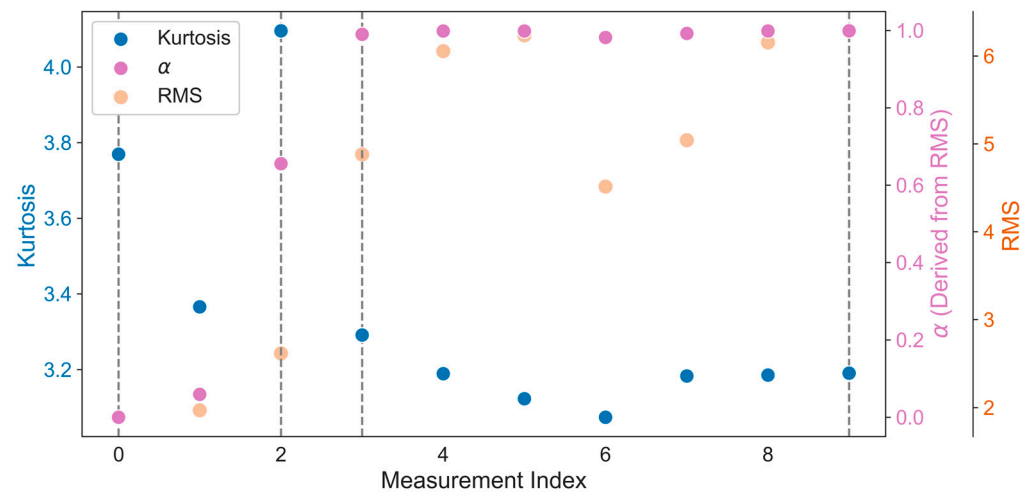


Figure 11. Temporal evolution of kurtosis, RMS, and α under dry test conditions with stage boundaries indicated by gray dashed lines.

Based on the predefined α -based threshold values, the degradation process is partitioned into distinct pseudo-classes, as summarized in Table 8. The pseudo-classes (stages) are defined such that Stage S1 represents the baseline regime, whereas Stages S2 and S3 represent the progressive degradation behavior. The corresponding stage boundaries are clearly illustrated in Figure 11 by gray dashed lines.

Table 8. Definition of pseudo-classes (stages), measurement index ranges, and associated degradation regimes.

Stage	Measurement Index Range	Description
S1	0–1	Baseline Regime
S2	2	Transition Regime
S3	3–9	Saturation-Like Regime

The preprocessing framework introduced in Section 3 is then applied, where the signals are segmented into non-overlapping windows corresponding to one shaft revolution and subsequently normalized using the adaptive class-wise scaling strategy within each pseudo-class. The window length is determined based on the sampling rate and the rotational speed of the system. Given a sampling rate of 100 kHz and a shaft rotational frequency of 10 Hz, one full rotation theoretically corresponds to 10,000 samples. Subsequently, as outlined in Section 3, each normalized 1D window (10,000 samples) is mapped into a 2D temporal representation of equal size, allowing the model to capture degradation-sensitive spatio-temporal characteristics.

Following the construction of the 2D temporal maps, the dataset is partitioned into training, validation, and test subsets. For LSTM training, only data from the baseline regime are utilized to ensure that the model learns the intrinsic characteristics of baseline operating conditions. Specifically, the first measurement record within the baseline regime—consisting of 110 2D temporal maps—is used for training. For validation, the subsequent measurement record is employed, corresponding to 110 2D temporal maps. The trained model is then evaluated on the complete run-to-failure dataset, which is used as the test set and comprises 1100 2D temporal maps.

4.2.3. Experimental Procedure for Case Study II

The experimental procedure is designed to evaluate the effectiveness of the proposed framework under a controlled experimental setting, where only baseline data are available during training. Following the learning phase, the model is applied sequentially to the complete run-to-failure dataset, enabling a continuous mapping of the degradation process within the latent feature space. Model parameters are optimized using the training and validation subsets, while the test set is used to assess how the learned representations evolve under previously unseen conditions. No prior assumptions or stage labels are incorporated, and the analysis is conducted in a data-driven manner, aiming to determine whether the progression of abrasive wear under dry conditions can be intrinsically organized into distinct degradation states within the learned feature space.

The application of the proposed framework produces a sequence of latent feature representations that capture the temporal evolution of the system under progressive degradation. These representations are subsequently analyzed in terms of their structural organization and clustering behavior within the latent space. This analysis aims to examine whether the learned feature space preserves the underlying progression dynamics and supports the identification of distinct degradation states associated with abrasive wear under dry conditions.

4.2.4. Results and Degradation Analysis for Case Study II

Effect of Temporal Map Configuration on Latent Representation and Clustering Performance for Case Study II

To assess the effect of 2D temporal-map configuration, five reshaping schemes were evaluated while keeping the LSTM architecture and Mini-Batch K-Means settings consistent with those described in Section 3. As summarized in Table 9, all evaluated temporal-map configurations achieve identical external clustering consistency with the reference degradation stages (ARI, NMI, FMI = 1.0), indicating that the proposed framework consistently preserves the overall degradation progression structure across different temporal-map resolutions. Although configurations with increasing temporal resolution exhibit slightly improved internal validation performance, these gains are accompanied by increased computational cost in both LSTM training and inference. Overall, the 200×50 configuration provides the most favorable overall behavior by maintaining strong latent representation quality together with substantially lower computational cost compared with configurations having increasing temporal resolution.

Performance Comparison with Alternative Clustering Algorithms for Case Study II

To further assess the robustness and generalizability of the proposed framework, a set of widely used unsupervised clustering algorithms, including Birch, Bisecting K-Means, K-Means, Spectral Clustering, and Agglomerative Clustering, were applied to the latent feature representations obtained from the LSTM. All methods were evaluated using their default parameter settings, with the exception of Birch, for which the threshold parameter was fixed at 0.05 to ensure stable clustering. The comparative clustering results corresponding to the optimal 200×50 2D map configurations for each algorithm are reported in Table 10.

As presented in Table 10, most clustering algorithms produce identical external clustering consistency with the reference degradation stages (ARI, NMI, FMI = 1.0) together with highly similar internal validation results (Silhouette = 0.26; Davies–Bouldin = 1.95), indicating that the learned latent representation is structurally stable and coherently organized. It should be noted that, consistent with the findings in Case Study I, the computational cost of most algorithms increases substantially with data volume, whereas Mini-Batch K-Means

demonstrates more scalable behavior. These results suggest that clustering performance is primarily governed by the quality of the learned latent representations, while the choice of algorithm mainly affects computational efficiency.

Table 9. Comparison of Clustering Performance and Computational Efficiency Across Different Temporal-Map Configurations for Case Study II.

Temporal Map Configuration	100 × 100	200 × 50	250 × 40	400 × 25	500 × 20
ARI	1.0	1.0	1.0	1.0	1.0
NMI	1.0	1.0	1.0	1.0	1.0
FMI	1.0	1.0	1.0	1.0	1.0
Silhouette	0.21	0.26	0.23	0.25	0.26
Davies–Bouldin	2.26	1.95	1.95	1.79	1.72
LSTM training time (baseline data) [s]	5.66	6.78	7.62	9.09	9.82
LSTM inference time (the entire data) [s]	0.27	0.29	0.32	0.43	0.64
Mini-Batch K-Means clustering time (the entire data) [s]	0.29	0.29	0.30	0.33	0.45
LSTM output shape	100 × 4	200 × 4	250 × 4	400 × 4	500 × 4
Flattened latent feature dimension	400	800	1000	1600	2000

Table 10. Comparison of Unsupervised Clustering Algorithms Using the Most Favorable 200 × 50 2D Map Configurations.

Unsupervised Algorithms	ARI	NMI	FMI	Silhouette	Davies–Bouldin	Clustering Time (100% Data) [s]
Birch	1.0	1.0	1.0	0.26	1.95	0.37
Bisecting K-Means	0.88	0.83	0.95	0.04	4.05	0.15
K-Means	1.0	1.0	1.0	0.26	1.95	0.22
Spectral Clustering	1.0	1.0	1.0	0.26	1.95	0.27
Agglomerative Clustering	1.0	1.0	1.0	0.26	1.95	0.17
Mini-Batch K-Means (proposed)	1.0	1.0	1.0	0.26	1.95	0.29

Comparative Evaluation of Clustering Algorithms for Unsupervised Degradation Analysis Without DL Models for Case Study II

Table 11 presents a comparative evaluation of several widely used unsupervised clustering algorithms directly applied to the Case Study II measurement space, without employing the proposed LSTM-based feature learning framework. Similar to the Case Study I analysis, the clustering methods operate on vectorized representations without using the 2D temporal map transformation. The performance of each method is evaluated using external validation metrics (ARI, NMI, FMI), internal validation metrics (Silhouette score and Davies–Bouldin Index), and clustering time measured on the complete dataset.

The analysis is conducted using four different feature representations: RMS, the class-wise scaling parameter α , the raw vibration windows ($x_{raw} \in \mathbb{R}^{10,000}$), and their adaptively scaled counterparts ($x_{scaled} \in \mathbb{R}^{10,000}$). The RMS feature provides a one-dimensional representation computed for each measurement, while α is derived from RMS behavior and used as a baseline-relative adaptive scaling parameter. The high-dimensional representation x_{raw} is constructed by segmenting the vibration signals into non-overlapping windows corresponding to one complete machine rotation, thereby preserving the temporal structure of the measurements. The representation x_{scaled} is then obtained by applying the proposed adaptive scaling strategy to these 10,000-dimensional windows in order to normalize the data while retaining degradation-related variations.

Table 11 indicates that the clustering behavior in Case Study II differs substantially from that observed in Case Study I, particularly with respect to the effectiveness of the

selected feature representations. In Case Study II, the α parameter achieves consistently perfect clustering performance across all evaluated algorithms, yielding ARI, NMI, and FMI values of 1.0 together with highly favorable internal validation metrics. However, this behavior is not equally observed in Case Study I, where the α parameter provides comparatively lower clustering consistency. One possible explanation is the substantially smaller number of measurements in Case Study II, which may allow the baseline-relative α parameter to more directly capture the overall degradation progression trend.

In contrast, the adaptively scaled high-dimensional representation x_{scaled} exhibits the opposite behavior. While x_{scaled} produces highly successful clustering results in Case Study I, where each vibration window consists of 1024 samples, its performance decreases considerably in Case Study II, where the window dimensionality increases to 10,000 samples. These findings suggest that direct clustering in the original measurement space becomes increasingly challenging as the dimensionality of the vibration representation increases, leading to weaker separability between degradation regimes, reduced cluster compactness, and lower clustering consistency with the reference degradation stages.

In contrast to the direct clustering approaches, whose performance varies considerably across datasets, the proposed framework demonstrates comparatively stable clustering behavior despite substantial differences in data volume and vibration-window dimensionality, while consistently maintaining perfect agreement with the reference degradation stages (ARI = NMI = FMI = 1.0) across both datasets.

Table 11. Comparative clustering performance of unsupervised algorithms using different feature representations without LSTM-based feature learning for Case Study II.

Unsupervised Algorithms	Features	ARI	NMI	FMI	Silhouette	Davies–Bouldin	Clustering Time (100% Data) [s]
Birch	<i>RMS</i>	0.37	0.65	0.62	0.83	0.20	0.19
	α	1.0	1.0	1.0	0.87	0.04	0.20
	x_{raw}	0.05	0.30	0.49	0.06	4.92	3.11
	x_{scaled}	0.48	0.65	0.71	0.01	4.73	3.58
Bisecting K-Means	<i>RMS</i>	0.37	0.65	0.62	0.83	0.20	0.15
	α	1.0	1.0	1.0	0.87	0.04	0.18
	x_{raw}	−0.003	0.002	0.73	0.05	0.79	0.36
	x_{scaled}	0.38	0.56	0.66	0.03	4.77	0.38
K-Means	<i>RMS</i>	0.37	0.65	0.62	0.83	0.20	0.20
	α	1.0	1.0	1.0	0.87	0.04	0.17
	x_{raw}	−0.01	0.20	0.46	0.06	4.43	1.25
	x_{scaled}	0.44	0.64	0.70	0.02	4.45	1.58
Spectral Clustering	<i>RMS</i>	0.37	0.65	0.62	0.83	0.20	0.19
	α	1.0	1.0	1.0	0.87	0.04	0.20
	x_{raw}	−0.05	0.21	0.45	0.09	4.26	0.43
	x_{scaled}	0.01	0.01	0.73	0.19	0.67	197.91
Agglomerative Clustering	<i>RMS</i>	0.37	0.65	0.62	0.83	0.20	0.0004
	α	1.0	1.0	1.0	0.87	0.04	0.0004
	x_{raw}	0.12	0.34	0.51	0.05	5.29	2.51
	x_{scaled}	0.48	0.65	0.71	0.01	4.73	2.62
Mini-Batch K-Means	<i>RMS</i>	0.37	0.65	0.62	0.83	0.20	0.20
	α	1.0	1.0	1.0	0.87	0.04	0.20
	x_{raw}	−0.05	0.11	0.53	0.09	3.46	0.59
	x_{scaled}	0.11	0.12	0.49	0.02	5.12	0.41

Comparative Evaluation of DL Models for Unsupervised Degradation Clustering for Case Study II

To further evaluate the robustness of the proposed framework, multiple DL architectures were applied to learn latent feature representations from the input data. The DL models detailed in Table 6 were employed in this analysis. To ensure a fair comparison, all training hyperparameters not explicitly listed were kept identical to those of the proposed model described in Section 3. In addition, each model was evaluated using its optimal 2D map configuration and a minimal network architecture to avoid unnecessary complexity. The resulting latent features were subsequently clustered using Mini-Batch K-Means to assess their effectiveness in representing the degradation process. Table 12 presents the performance comparison of DL models for unsupervised degradation clustering. The DL column denotes the feature space construction time of the DL model (inference time for latent representation extraction), while the Cluster column represents the inference time of the clustering algorithm applied to the extracted feature space. Similar to Case Study I, latent feature dimensionality is reported for transparency and should be interpreted as a descriptive characteristic of the learned representation rather than an independent performance variable.

Table 12. Performance Comparison of DL Models for Unsupervised Degradation Clustering for Case Study II.

DL Algorithms	2D Map Configurations	Latent Feature Dimension	ARI	NMI	FMI	Silhouette	Davies–Bouldin	Inference Time (100% Data) [s]	
								DL	Cluster
Conv2D	400 × 25	40,000	1.0	1.0	1.0	0.17	2.77	0.19	5.78
SeparableConv2D	200 × 50	40,000	1.0	1.0	1.0	0.26	2.28	0.14	3.28
DepthwiseConv2D	400 × 25	2600	0.39	0.55	0.66	0.02	5.39	0.07	0.66
ConvLSTM2D	110 × 200 × 50	80,000	0.65	0.69	0.85	0.17	0.60	4.50	0.37
GRU	200 × 50	800	1.0	1.0	1.0	0.26	1.82	0.31	0.31
TCN	400 × 25	1200	1.0	1.0	1.0	0.12	3.04	0.16	0.37
LSTM + MultiHeadAttention	200 × 50	800	1.0	1.0	1.0	0.26	1.95	0.44	0.33
LSTM (proposed)	200 × 50	800	1.0	1.0	1.0	0.26	1.95	0.29	0.29

As shown in Table 12, the performance of the evaluated DL architectures varies considerably under the larger 2D map configurations employed in Case Study II. While several models demonstrated strong performance in Case Study I, their effectiveness declines as the temporal-map resolution increases, particularly for some convolutional (DepthwiseConv2D) and hybrid (ConvLSTM2D) architectures. Although Conv2D and SeparableConv2D achieved perfect external clustering agreement, their substantially longer clustering times, particularly under higher-resolution temporal 2D maps, suggest lower computational favorability for clustering.

Across both Case Study I and Case Study II, GRU, TCN, and the proposed LSTM consistently achieved stable and perfect external clustering performance (ARI/NMI/FMI = 1.0), indicating reliable discrimination capability across different temporal 2D map configurations. However, the proposed LSTM demonstrated the most robust overall behavior when internal clustering quality and computational efficiency are jointly considered. In particular, LSTM exhibited more consistent internal clustering characteristics (Silhouette and Davies–Bouldin scores) and lower clustering time compared to GRU and TCN, especially in Case Study I, while maintaining comparable efficiency in Case Study II.

Across both case studies, the LSTM and LSTM with MultiHead Attention achieved identical performance in terms of external clustering (ARI/NMI/FMI = 1.0) and internal

clustering metrics, indicating that the attention mechanism does not provide additional discriminative or structural benefits for the learned latent representations in this setting. However, the attention-augmented model introduces a substantial computational overhead, where the DL inference time increases significantly without corresponding performance gains. Overall, these results suggest that the attention mechanism does not improve clustering effectiveness but instead increases computational cost.

Across both Case Study I and Case Study II, the proposed LSTM maintains consistently stable performance despite variations in data volume and temporal 2D map resolution. It achieves perfect external clustering performance (ARI/NMI/FMI = 1.0) in all configurations, while also exhibiting stable and competitive internal clustering behavior across both cases. Additionally, the proposed LSTM demonstrates favorable scalability in terms of computational cost, maintaining efficient performance under varying data volume and temporal resolution. Overall, these results indicate that the proposed LSTM exhibits strong robustness to changes in data volume and temporal representation complexity, maintaining both high clustering reliability and computational efficiency across heterogeneous experimental conditions.

5. Discussion

This study proposed a hybrid framework designed to model lubrication loss-related degradation progression as a gradual process involving distinct yet partially overlapping behavioral regimes. Within this framework, adaptive class-wise scaling was utilized to mitigate data leakage during preprocessing while improving robustness to data imbalance. Subsequently, lightweight LSTM-based deep feature learning trained exclusively on baseline data was employed to construct a latent feature space from vibration measurements, and Mini-Batch K-Means clustering was applied to organize the learned latent space into structurally coherent degradation regimes. Consistent results obtained across both the real-world compressor dataset and the experimental dataset further demonstrate the robustness and generalizability of the proposed framework in lubrication loss-related degradation modeling.

It should be noted that, in both datasets, the baseline data is likely to include incipient degradation rather than representing idealized healthy conditions, as the kurtosis values exceed 3.5 from the initial observations, indicating the presence of impulsive behavior. Despite this, the LSTM-based feature learning stage, trained on a limited subset of baseline data, produces a latent representation that remains relatively stable under baseline operating conditions while exhibiting deviating as degradation advances. This behavior suggests that the learned representation encodes degradation-related variations in the vibration data.

Across both datasets, the proposed framework achieved perfect agreement with the internally defined reference degradation stages, with ARI, NMI, and FMI values all equal to 1.0, indicating highly consistent clustering behavior. In contrast, the observed Silhouette and Davies–Bouldin values reflect relatively limited cluster separation within the latent feature space. This behavior can be interpreted in light of several contributing factors. First, degradation is inherently a progressive process, in which the system evolves through gradual and potentially overlapping transitions rather than discrete states. Such behavior naturally limits the formation of sharply separated clusters. Second, the adopted class-wise scaling strategy, in which each regime is independently scaled to the [0, 1] range, reduces inter-stage variance while preserving intra-stage distributions, thereby limiting geometric separability. Finally, the baseline data may already contain incipient degradation rather than representing strictly healthy conditions, which further blurs the boundaries between degradation states. Together, these factors can contribute to reduced inter-stage separability.

In addition to these findings, the results obtained in Case Study I further indicate that geometric cluster separation alone does not guarantee a meaningful representation of degradation. This is further supported by direct clustering performed on various input representations, including RMS, the class-wise scaling parameter α , the raw high-dimensional vibration windows x_{raw} , and their adaptively scaled counterparts x_{scaled} . Although RMS and α produce well-separated clusters, they primarily reflect amplitude-driven variations and may not fully capture the actual progression of degradation, leading to weaker agreement with the underlying structure. In contrast, the poor performance of x_{raw} highlights the necessity of addressing scale and distributional complexity. Adaptive scaling appears to improve structural consistency by reorganizing the feature space according to relative system changes, enabling more accurate clustering despite moderate geometric separation. Although the x_{scaled} representation demonstrates strong performance among these inputs, it still exhibits lower Silhouette scores and higher Davies–Bouldin indices compared to the proposed LSTM-based approach, indicating limited cluster separability. However, the LSTM-based representation offers a further improvement by embedding temporal dependencies, resulting in a more coherent organization of degradation states. These findings suggest that effective degradation analysis depends primarily on representation quality rather than cluster separability alone.

The comparative analyses performed on both the real-world compressor dataset and the experimental dataset reveal that the effectiveness of directly clustered RMS, α parameter, x_{raw} and their adaptively scaled counterparts x_{scaled} is strongly dependent on dataset characteristics, particularly data volume and vibration-window dimensionality. In the real-world compressor dataset, where the vibration windows consist of 1024 samples, x_{scaled} achieves highly favorable clustering consistency across several algorithms. However, this behavior does not generalize to the experimental dataset, where the vibration-window dimensionality increases substantially to 10,000 samples. Under these conditions, the clustering performance of x_{scaled} decreases considerably, suggesting that direct clustering in high-dimensional vibration spaces becomes increasingly difficult, leading to weaker separability between degradation regimes, reduced cluster compactness, and lower clustering consistency with the reference degradation stages. Conversely, the α parameter achieves highly favorable clustering behavior in the experimental dataset, but does not exhibit the same level of consistency in the real-world compressor dataset, which contains a substantially larger data volume. These observations indicate that no single handcrafted or directly clustered representation consistently maintains stable behavior across datasets with different characteristics, data volumes, and window dimensionalities.

Furthermore, while certain clustering algorithms (e.g., K-Means, Spectral Clustering, and Agglomerative Clustering) achieve performance comparable to the proposed model across all evaluation metrics in the experimental dataset and offer improved computational efficiency, this advantage does not persist under real-world conditions. As the data volume increases, their computational cost grows markedly, revealing inherent limitations in scalability and reducing their practical applicability.

The comparative analysis across the real-world compressor dataset and the experimental dataset reveals that the evaluated DL architectures respond differently to variations in data volume and temporal 2D map resolution. While the real-world compressor dataset involves higher data volume with lower-resolution temporal representations, the experimental dataset represents a contrasting setting with lower data volume but higher-resolution temporal maps. Despite these opposing configurations, the proposed LSTM consistently maintains perfect external clustering performance (ARI/NMI/FMI = 1.0) across both cases, indicating strong robustness in degradation state discrimination.

In terms of internal clustering quality, recurrent-based architectures (GRU, TCN, and the proposed LSTM) demonstrate more stable behavior compared to convolution-based models, which exhibit greater sensitivity to changes in temporal representation resolution. Although certain CNN-based models achieve perfect external clustering in both cases, their clustering times increase significantly under higher-resolution temporal maps, suggesting reduced efficiency in clustering-friendly feature representation.

From a computational perspective, the proposed LSTM maintains consistently low inference and clustering times across both datasets, indicating favorable scalability under varying input configurations. In contrast, in both case studies, the attention-enhanced LSTM incurs higher computational cost than the proposed LSTM, particularly in Case Study I, characterized by higher data volume and lower temporal resolution.

Another key advantage of the proposed framework lies in its computational efficiency and robustness. Unlike many existing approaches, no intensive feature extraction techniques are employed during preprocessing; instead, following class-wise adaptive scaling, the raw input data are directly reshaped into 2D maps. Furthermore, the adopted LSTM architecture is lightweight, consisting of only four units. In addition, both the LSTM model and the Mini-Batch K-Means clustering algorithm exhibit stable performance with consistent and scalable computational behavior across both the real-world compressor and experimental datasets, despite variations in 2D map resolutions and input data volume. These characteristics collectively highlight the scalability and efficiency of the proposed approach.

Despite the promising performance of the proposed framework, several limitations should be acknowledged. The method relies on an initial baseline regime that may already contain incipient degradation, making a strict separation from truly healthy conditions inherently ambiguous and potentially blurring the distinction between early-stage behaviors. In addition, under strong non-stationary operating conditions (e.g., variations in load, speed, or environmental factors), changes observed in the signal may not be attributable solely to degradation, which can affect the consistency and interpretability of the learned representation. Another limitation of this study is that the reference degradation stages are constructed using a pseudo-class strategy rather than externally validated ground truth annotations. In addition, the lack of externally grounded validation using independent datasets or expert-annotated operational data limits the assessment of generalizability. These factors define the practical boundaries of applicability of the proposed approach and indicate directions for future improvements.

A further limitation of the proposed framework is that the number of clusters is determined based on the observed behavior of kurtosis and the adaptive scaling parameter α , rather than through a fully automated data-driven criterion. While automated approaches such as change-point detection or trend-based analysis could be employed for this purpose, the primary objective of this study is not to obtain a purely statistically optimal partition, but to derive a physically interpretable representation of degradation progression. In this context, the use of kurtosis and α provides a meaningful basis for capturing the evolution of degradation. Future work may focus on integrating automated cluster selection strategies, such as cluster validity-based criteria, probabilistic model selection, or change-point detection techniques, to further enhance the adaptability of the framework.

Nevertheless, it should be noted that Case Study II includes a limited number of independent measurement records, representing a small-scale experimental validation. However, it serves to evaluate the robustness of the proposed framework under different temporal representation conditions rather than to claim broad generalization across datasets. Furthermore, Case Study II was included as a supplementary validation study using a

gearbox test rig to assess the applicability of the proposed framework beyond compressor-specific operating conditions.

6. Conclusions

This study addresses a fundamental challenge in industrial condition monitoring: the modeling of degradation progression induced by loss of lubrication without labeled data, under non-ideal baseline conditions, continuous fault evolution and imbalanced data distributions, while explicitly preventing data leakage during preprocessing. In this context, the proposed framework combines adaptive preprocessing, baseline-referenced LSTM-based deep feature learning, and Mini-Batch K-Means clustering to organize complex vibration behavior into structurally coherent degradation-related behavioral regimes.

Despite these challenges, the results demonstrate that the proposed method consistently captures the internally constructed degradation progression patterns across datasets with different characteristics, yielding ARI, NMI, and FMI scores of 1.0 and indicating that the learned latent representations effectively preserve the underlying degradation behavior embedded in the vibration data. In addition, the proposed framework enables the characterization of degradation progression through deviations from a baseline reference, organizing it into structurally coherent and interpretable behavioral regimes.

From an implementation perspective, the proposed framework demonstrates strong computational efficiency and scalability. By avoiding computationally intensive feature extraction and employing a lightweight LSTM architecture together with Mini-Batch K-Means, the method maintains consistent clustering performance with scalable computational behavior under varying data volume and 2D map resolutions. These characteristics make it a promising approach for large-scale monitoring applications under practical computational constraints.

Overall, this study indicates that effective unsupervised degradation analysis in real industrial systems depends less on strong cluster separability and more on the ability to consistently characterize degradation progression relative to a realistic baseline reference. In this context, the proposed framework provides a scalable and practically grounded approach for interpreting complex system behavior under limited supervision.

The study has several limitations, including the reliance on pseudo-class-based reference labels, the absence of externally grounded validation, and the limited scale of Case Study II. Future work will address these issues through validation on larger and externally annotated industrial datasets.

Author Contributions: Conceptualization, G.Ö.; methodology, G.Ö.; software, G.Ö.; validation, İ.T.; formal analysis, G.Ö.; investigation, G.Ö.; writing—original draft preparation, G.Ö.; visualization, G.Ö.; supervision, İ.T. All authors have read and agreed to the published version of the manuscript.

Funding: This study was funded by the Firat University Scientific Research Projects Unit (FÜBAP) under Project Number TEKF.26.39. The authors gratefully acknowledge the financial support provided by FÜBAP, which partially covered the article processing charge (APC).

Data Availability Statement: Restrictions apply to the availability of these data. The dataset was obtained from Eric Bechhoefer and is distributed directly by him upon reasonable request (contact: ebechhoefer@gmail.com).

Acknowledgments: This study is supported by TÜBİTAK—BİDEB 2211/C National PhD Scholarship Program in the Priority Fields in Science and Technology, 100/2000 Council of Higher Education (Yükseköğretim Kurulu—YÖK) Doctoral Scholarship Program. We would like to thank TÜBİTAK and YÖK for their support. The authors would like to express their sincere gratitude to Eric Bechhoefer for providing the industrial dataset that made this study possible and for his valuable technical discussions, insightful comments, and constructive suggestions throughout the preparation of this

manuscript. This article is derived from the doctoral thesis entitled “Development of AI-Based Online Vibration Analysis Methods for Assessing the Health of Rotating Machines,” submitted by Gonca Öcalan to the Graduate School of Natural and Applied Sciences at Firat University.

Conflicts of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AE	Auto Encoder
AI	Artificial Intelligence
ARI	Adjusted Rand Index
CNN	Convolutional Neural Network
CWT	Continuous Wavelet Transform
DL	Deep Learning
FFT	Fast Fourier Transform
FMI	Fowlkes–Mallows Index
GMF	Gear Mesh Frequency
LSTM	Long Short-Term Memory
ML	Machine Learning
MLP	Multi-Layer Perceptron
NMI	Normalized Mutual Information
NPR	Neural Pattern Recognition
PSD	Power Spectrum Density
RMS	Root Mean Square
SGD	Stochastic Gradient Descent
SVM	Support Vector Machine
t-SNE	t-distributed Stochastic Neighbour Embedding
1D	One-Dimensional
2D	Two-dimensional

References

1. Braumann, L.; de Viteri, V.S.; Morhard, B.; Lohner, T.; Ochoa, J.; Amri, H. Tribology Technologies for Gears in Loss of Lubrication Conditions: A Review. *J. Mater. Sci. Mater. Eng.* **2025**, *20*, 29. [\[CrossRef\]](#)
2. Morhard, B.; Lohner, T.; Stahl, K. Influence of Surface and Material Technologies on the Loss of Lubrication Performance of Gears. *J. Mater. Sci. Mater. Eng.* **2024**, *19*, 1. [\[CrossRef\]](#)
3. Herwig, N.; Peng, Z.; Borghesani, P. Bridging the Trust Gap: Evaluating Feature Relevance in Neural Network-Based Gear Wear Mechanism Analysis with Explainable AI. *Tribol. Int.* **2023**, *187*, 108670. [\[CrossRef\]](#)
4. Chen, Q.; Cen, J.; Chen, K.; Zhang, X.; Feng, C.; Zhang, D.; Ge, S. Evolution Behavior on Friction-Induced Vibration and Lubrication State Identification. *Tribol. Int.* **2025**, *210*, 110817. [\[CrossRef\]](#)
5. Song, J.; Dong, C.; Bai, X.; Yuan, C. Friction-Coefficient Prediction and Wear-State Diagnosis of Rotary-Vane Steering Gear Seals Using CNN-LSTM and LightGBM. *Tribol. Int.* **2025**, *212*, 110991. [\[CrossRef\]](#)
6. Rigolli, D.; Pompei, L.; Manfredini, M.; Vignoli, M.; La Battaglia, V.; Giorgetti, A. The Development of a Predictive Maintenance System for Gearboxes Through a Statistical Diagnostic Analysis of Lubricating Oil and Artificial Intelligence. *Machines* **2025**, *13*, 693. [\[CrossRef\]](#)
7. Yu, H.; Wei, H. Lubrication State Identification of Vibration Time-Frequency Characteristics Based on CWT and CNN. *Sci. Rep.* **2025**, *15*, 28936. [\[CrossRef\]](#) [\[PubMed\]](#)
8. Lu, X.; Li, P. Research on Gearbox Temperature Field Image Fault Diagnosis Method Based on Transfer Learning and Deep Belief Network. *Sci. Rep.* **2023**, *13*, 6664. [\[CrossRef\]](#) [\[PubMed\]](#)
9. Resendiz-Ochoa, E.; Saucedo-Dorantes, J.J.; Benitez-Rangel, J.P.; Osornio-Rios, R.A.; Morales-Hernandez, L.A. Novel Methodology for Condition Monitoring of Gear Wear Using Supervised Learning and Infrared Thermography. *Appl. Sci.* **2020**, *10*, 506. [\[CrossRef\]](#)
10. Touret, T.; Changenet, C.; Ville, F.; Lalmi, M.; Becquerelle, S. On the Use of Temperature for Online Condition Monitoring of Geared Systems—A Review. *Mech. Syst. Signal Process.* **2018**, *101*, 197–210. [\[CrossRef\]](#)

11. Kumar Gupta, K.; Muzakkir, S.M. Prediction of Gearbox Oil Degradation Based on Online Sensor Data and Machine Learning Algorithms. *Tribol. Ind.* **2023**, *45*, 487–502. [[CrossRef](#)]
12. Grebenişan, G.; Salem, N.; Bogdan, S.; Negrău, D.C. Oil Condition Monitoring, an AI Application Study Using the Classification Learner Technics. *IOP Conf. Ser. Mater. Sci. Eng.* **2021**, *1169*, 012012. [[CrossRef](#)]
13. Chen, Y.; Shen, J.; Zhou, F.; Li, H.; Yang, K.; Wang, L. Multi-Sensor Fusion and Deep Learning for Predictive Lubricant Health Assessment. *Lubricants* **2025**, *13*, 364. [[CrossRef](#)]
14. Leaman, F. A Review on Acoustic Emissions of Gear Transmissions: Source, Influencing Parameters, Applications and Modeling. *J. Vib. Eng. Technol.* **2024**, *12*, 7835–7846. [[CrossRef](#)]
15. Peng, Z.; Kessissoglou, N. An Integrated Approach to Fault Diagnosis of Machinery Using Wear Debris and Vibration Analysis. *Wear* **2003**, *255*, 1221–1232. [[CrossRef](#)]
16. Banks, J.C.; Reichard, K.M.; Brought, M.S. Lubrication Level Diagnostics Using Vibration Analysis. In Proceedings of the 2004 IEEE Aerospace Conference Proceedings (IEEE Cat. No.04TH8720), Big Sky, MT, USA, 6–13 March 2004; Volume 6, pp. 3528–3534. [[CrossRef](#)]
17. Brethee, K.F.; Zhang, R.; Hamad, N.; Gu, F.; Ball, A.D. Influence of Lubricant Starvation on Gearbox Vibration Signatures for Condition Monitoring. In Proceedings of the COMADEM 2016, the 29th International Congress on Condition Monitoring and Diagnostic Engineering Management, Xi'an, China, 20–22 August 2016; pp. 1–13.
18. Hu, B.; Zhou, C.; Wang, H.; Chen, S. Nonlinear Tribo-Dynamic Model and Experimental Verification of a Spur Gear Drive under Loss-of-Lubrication Condition. *Mech. Syst. Signal Process.* **2021**, *153*, 107509. [[CrossRef](#)]
19. Islam, T.; Joarder, A.; Hossain, K. A Comprehensive Study on Gearbox Condition Monitoring Using Vibration Analysis. *N. Am. Acad. Res.* **2024**, *7*, 1–11. [[CrossRef](#)]
20. Aher, H.R.; Ghuge, N.C. Vibration-Based Condition Monitoring of Tapered Roller Bearings Using Kurtosis and ANOVA. *Tribol. Ind.* **2025**, *47*, 370–388. [[CrossRef](#)]
21. Kumar, S.; Agarwal, P.K. Experimental Investigation of Gearbox Faults Using Vibration Analysis and Statistical Methods. *Noise Vib. Worldw.* **2025**, *56*, 507–517. [[CrossRef](#)]
22. Saravanan, N.; Ramachandran, K.I. Incipient Gear Box Fault Diagnosis Using Discrete Wavelet Transform (DWT) for Feature Extraction and Classification Using Artificial Neural Network (ANN). *Expert Syst. Appl.* **2010**, *37*, 4168–4181. [[CrossRef](#)]
23. Keleşoğlu, C.; Küçük, H.; Demetgül, M. Fault Diagnosis of Bevel Gears Using Neural Pattern Recognition and MLP Neural Network Algorithms. *Int. J. Precis. Eng. Manuf.* **2020**, *21*, 843–856. [[CrossRef](#)]
24. Lupea, I.; Lupea, M.; Coroian, A. Helical Gearbox Defect Detection with Machine Learning Using Regular Mesh Components and Sidebands. *Sensors* **2024**, *24*, 3337. [[CrossRef](#)] [[PubMed](#)]
25. Babu, T.N.; Ali, P.S.N.; Prabha, D.R.; Mohammed, V.N.; Wahab, R.S.; Vijayalakshmi, S. Fault Diagnosis in Bevel Gearbox Using Coiflet Wavelet and Fault Classification Based on ANN Including DNN. *Arab. J. Sci. Eng.* **2022**, *47*, 15823–15849. [[CrossRef](#)]
26. Lupea, I.; Lupea, M. Continuous Wavelet Transform and CNN for Fault Detection in a Helical Gearbox. *Appl. Sci.* **2025**, *15*, 950. [[CrossRef](#)]
27. Lupea, I.; Lupea, M. Detecting Helical Gearbox Defects from Raw Vibration Signal Using Convolutional Neural Networks. *Sensors* **2023**, *23*, 8769. [[CrossRef](#)] [[PubMed](#)]
28. Hemati, A. Bearing Failure Prevention Through Predictive Lubrication Monitoring Using Vibration Signals and Deep Learning. *J. Fail. Anal. Prev.* **2025**, *25*, 2165–2175. [[CrossRef](#)]
29. Hou, Z.; Wang, B.; Liu, J.; He, Y.; Yao, Y. Physics-Inspired Time-Frequency Feature Extraction and Lightweight Neural Network for Power Quality Disturbance Classification. *Front. Phys.* **2025**, *13*, 1616367. [[CrossRef](#)]
30. Hou, Z.; Liu, J.; Yu, S. Enhanced Analog Circuit Fault Diagnosis via Continuous Wavelet Transform and Dual-Stream Convolutional Fusion. *Sci. Rep.* **2025**, *15*, 19828. [[CrossRef](#)] [[PubMed](#)]
31. Li, J.; Liu, X.; Wu, X.; Wang, D.; Xu, K.; Li, Y. A Time and Frequency Convolutional Autoencoder for Anomaly Detection in Industrial Robots Based on Inertial Measurement Unit Error Calibration. *Eng. Appl. Artif. Intell.* **2025**, *162*, 112269. [[CrossRef](#)]
32. Zhong, Z.; Zhao, Y.; Yang, A.; Zhang, H.; Qiao, D.; Zhang, Z. Industrial Robot Vibration Anomaly Detection Based on Sliding Window One-Dimensional Convolution Autoencoder. *Shock Vib.* **2022**, *2022*, 1179192. [[CrossRef](#)]
33. Lee, H.; Yu, J. A Fault Detection Framework for Rotating Machinery with a Spectrogram and Convolutional Autoencoder. *Appl. Sci.* **2025**, *15*, 7698. [[CrossRef](#)]
34. Zhou, Q.; He, S.; Liu, H.; Chen, J.; Meng, W. Label-Free Multivariate Time Series Anomaly Detection. *IEEE Trans. Knowl. Data Eng.* **2024**, *36*, 3166–3179. [[CrossRef](#)]
35. Yin, K.; Chen, C.; Yang, L.; Zhang, Y.; Ou, F. Unsupervised State Evaluation of Rotating Machinery Using a Lightweight Autoencoder With Adaptive Decision Boundaries. *IEEE Sens. J.* **2025**, *25*, 36569–36579. [[CrossRef](#)]
36. Çekik, R.; Turan, A. Deep Learning for Anomaly Detection in CNC Machine Vibration Data: A RoughLSTM-Based Approach. *Appl. Sci.* **2025**, *15*, 3179. [[CrossRef](#)]

37. Seo, S.; Han, B. Re-Evaluating Group Robustness via Adaptive Class-Specific Scaling. In *Proceedings of the 2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), Tucson, AZ, USA, 26 February–6 March 2025*; IEEE: Piscataway, NJ, USA, 2025; pp. 8379–8388.
38. Francazi, E.; Baity-Jesi, M.; Lucchi, A. A Theoretical Analysis of the Learning Dynamics under Class Imbalance. *Proc. Mach. Learn. Res.* **2024**, *202*, 10285–10322.
39. Kapoor, S.; Narayanan, A. Leakage and the Reproducibility Crisis in Machine-Learning-Based Science. *Patterns* **2023**, *4*, 100804. [[CrossRef](#)] [[PubMed](#)]
40. Zaidi, S.H.H.; Shenfield, A.; Zhang, H.; Ikpehai, A. A Systematic Review of Anomaly and Fault Detection Using Machine Learning for Industrial Machinery. *Algorithms* **2026**, *19*, 108. [[CrossRef](#)]
41. Sufi, F. Beyond the Sensor: A Systematic Review of AI's Role in Next-Generation Machine Health Monitoring. *Appl. Sci.* **2025**, *15*, 10494. [[CrossRef](#)]
42. Iftee, M.A.R.; Hasan, S.M.A.; Hossain, M.S.; Rajib, R.H.; Ali, A.A.; Rahman, A.M.; Mistry, S.; Bhuyan, M. PFedBBN: A Personalized Federated Test-Time Adaptation with Balanced Batch Normalization for Class-Imbalanced Data. *arXiv* **2025**, arXiv:2511.18066.
43. Raghu, S.T.P.; MacIsaac, D.T.; Scheme, E.J. Self-Supervised Representation Learning with Continuous Training Data Improves the Feel and Performance of Myoelectric Control. *Comput. Biol. Med.* **2025**, *197*, 111029. [[CrossRef](#)] [[PubMed](#)]
44. Hoang, D.-T.; Kang, H.-J. Rolling Element Bearing Fault Diagnosis Using Convolutional Neural Network and Vibration Image. *Cogn. Syst. Res.* **2019**, *53*, 42–50. [[CrossRef](#)]
45. Wang, Y.; Cao, G.; Han, J. A Combination of Dilated Self-Attention Capsule Networks and Bidirectional Long- and Short-Term Memory Networks for Vibration Signal Denoising. *Machines* **2022**, *10*, 840. [[CrossRef](#)]
46. Gong, W.; Chen, H.; Zhang, Z.; Zhang, M.; Wang, R.; Guan, C.; Wang, Q. A Novel Deep Learning Method for Intelligent Fault Diagnosis of Rotating Machinery Based on Improved CNN-SVM and Multichannel Data Fusion. *Sensors* **2019**, *19*, 1693. [[CrossRef](#)] [[PubMed](#)]
47. Bharatheedasan, K.; Maity, T.; Kumaraswamidhas, L.A.; Durairaj, M. Enhanced Fault Diagnosis and Remaining Useful Life Prediction of Rolling Bearings Using a Hybrid Multilayer Perceptron and LSTM Network Model. *Alex. Eng. J.* **2025**, *115*, 355–369. [[CrossRef](#)]
48. Vos, K.; Peng, Z.; Jenkins, C.; Shahriar, M.R.; Borghesani, P.; Wang, W. Vibration-Based Anomaly Detection Using LSTM/SVM Approaches. *Mech. Syst. Signal Process.* **2022**, *169*, 108752. [[CrossRef](#)]
49. Lique, B.; Moka, S.; Nazareth, Y. Principles of Machine Learning. In *Mathematical Engineering of Deep Learning*; Chapman and Hall/CRC: Boca Raton, FL, USA, 2024; pp. 27–73. ISBN 9789819753321.
50. Abadi, M.; Agarwal, A.; Barham, P.; Brevdo, E.; Chen, Z.; Citro, C.; Corrado, G.S.; Davis, A.; Dean, J.; Devin, M.; et al. *TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems*; Google LLC: Mountain View, CA, USA, 2015. Available online: <https://www.tensorflow.org/> (accessed on 2 July 2026).
51. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-Learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
52. Harris, C.R.; Millman, K.J.; van der Walt, S.J.; Gommers, R.; Virtanen, P.; Cournapeau, D.; Wieser, E.; Taylor, J.; Berg, S.; Smith, N.J.; et al. Array Programming with NumPy. *Nature* **2020**, *585*, 357–362. [[CrossRef](#)] [[PubMed](#)]
53. McKinney, W. Data Structures for Statistical Computing in Python. In *Proceedings of the 9th Python in Science Conference (SciPy 2010), Austin, TX, USA, 28 June–3 July 2010*; Volume 1, pp. 56–61. [[CrossRef](#)]
54. Meng, Q.; Wen, W.; Bai, Y.; Liu, Y. A Fault Detection Method for Electrohydraulic Switch Machine Based on Oil-Pressure-Signal-Sectionalized Feature Extraction. *Entropy* **2022**, *24*, 848. [[CrossRef](#)] [[PubMed](#)]
55. Lv, X.-L.; Chiang, H.-D.; Dong, N. Automatic DNN Architecture Design Using CPSOTJUTT for Power System Inspection. *J. Big Data* **2023**, *10*, 150. [[CrossRef](#)]
56. Sculley, D. Web-Scale k-Means Clustering. In *Proceedings of the 19th International Conference on World Wide Web, Raleigh, NC, USA, 26–30 April 2010*; ACM: New York, NY, USA, 2010; pp. 1177–1178.
57. Van Der Maaten, L.; Hinton, G. Visualizing Data Using T-SNE. *J. Mach. Learn. Res.* **2008**, *9*, 2579–2605.
58. Wang, T. A Photovoltaic Power Ultra Short-Term Prediction Method Integrating Hungarian Clustering and PSO Algorithm. *Energy Inform.* **2025**, *8*, 5. [[CrossRef](#)]
59. Selle, M.; Kircher, M.; Schwennen, C.; Visscher, C.; Jung, K. Dimension Reduction and Outlier Detection of 3-D Shapes Derived from Multi-Organ CT Images. *BMC Med. Inform. Decis. Mak.* **2024**, *24*, 49. [[CrossRef](#)] [[PubMed](#)]
60. Feng, K.; Borghesani, P.; Chang, H.; Chin, Z.Y.; Peng, Z.; Smith, W. *Gear Wear Run-to-Failure Dataset, V2*; Mendeley Data. 2021. [[CrossRef](#)]

61. Feng, K.; Smith, W.A.; Borghesani, P.; Randall, R.B.; Peng, Z. Use of Cyclostationary Properties of Vibration Signals to Identify Gear Wear Mechanisms and Track Wear Evolution. *Mech. Syst. Signal Process.* **2021**, *150*, 107258. [[CrossRef](#)]
62. Montgomery, D.C.; Peck, E.A.; Vining, G.G. *Introduction to Linear Regression Analysis*, 5th ed.; John Wiley & Sons: Hoboken, NJ, USA, 2012; ISBN 9780470542811.

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.