machines

MDPI

Article

Deep Learning-Based Anomaly Detection in Electric Motor Production Using Mel-Spectrogram Representations and a Hybrid Neural Network

Jernej Mlinarič [1,*], Boštjan Pregelj [2] and Gregor Dolanc [2]

[1] Jozef Stefan International Postgraduate School, Jamova Cesta 39, 1000 Ljubljana, Slovenia
[2] Jozef Stefan Institute, Jamova Cesta 39, 1000 Ljubljana, Slovenia; bostjan.pregelj@ijs.si (B.P.); gregor.dolanc@ijs.si (G.D.)
* Correspondence: jernej.mlinaric@ijs.si

Abstract

End-of-Line (EoL) quality inspection of electric motors requires reliable detection of manufacturing faults before products leave the production line. However, conventional supervised deep learning approaches depend on a sufficient number of labeled instances of faulty motors, which are scarce in high-quality manufacturing environments. This limits their applicability to newly introduced products and previously unobserved fault types. This paper presents an unsupervised deep learning framework for anomaly detection in motor acoustic and vibration signals, designed for EoL quality inspection in manufacturing. The proposed method leverages Mel-frequency spectrograms (MFSs) as input features and employs a hybrid neural network combining a convolutional neural network (CNN) and bidirectional gated recurrent units (BiGRUs), effectively capturing both local spectral patterns and temporal dependencies. The method is evaluated on real industrial production data from over 2400 motors, of which approximately 4.3% were faulty, reflecting the highly imbalanced nature of high-quality manufacturing. The model was trained exclusively on healthy motor data and therefore does not require labeled faulty samples during training. Experimental results demonstrate strong discrimination between healthy and faulty motors, indicating that the proposed approach is suitable for automated EoL quality inspection in manufacturing environments where labeled faulty data are scarce.

Keywords: end-of-line quality inspection; fault detection; vibration monitoring; condition monitoring; deep learning; hybrid neural network

Check for updates

Academic Editor: Maciej Sułowicz

Received: 16 July 2026
Revised: 27 August 2026
Accepted: 3 September 2026
Published: 4 September 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.

1. Introduction

Nowadays, highly reliable electric motors play a crucial role across a wide range of applications, from industrial automation and robotics to customer applications. In modern manufacturing lines, EoL quality inspection is essential to ensure highly reliable electric motor production that meets strict performance, safety and customer requirements [1,2]. The importance of EoL inspection has been well-documented. Manufacturing tolerances and assembly variations have been demonstrated to introduce defects that must be identified and addressed before motors reach end users [3]. In this context, automated EoL inspection systems have become a critical component of zero-defect manufacturing strategies [4]. Traditional EoL inspection systems rely on signal preprocessing, feature engineering and rule-based thresholding [5,6]. Quality assessment typically combines measured electrical parameters, processed acoustic and vibrational signals and expert knowledge. Early work

on automated EoL inspection for electric motors demonstrated that a combination of acoustic signal analysis and expert knowledge can identify fault types reliably on the production line [7,8]. While these methods are well established, they often struggle to capture complex, non-linear fault signatures and are limited in their ability to generalize to new or evolving fault types [9].

Recent advances in deep learning have enabled powerful data-driven approaches for fault detection and quality control. CNNs have been recognized as a successful approach for extracting discriminative spatial features from time-frequency representations such as spectrograms, while recurrent architectures such as BiGRUs capture temporal dependencies in sequential data [10,11]. However, most existing methods are either fully supervised, requiring large labeled datasets, or they focus on predictive maintenance [12,13]. A critical challenge in industrial fault detection is the significant class imbalance inherent to real production environments. Faulty motors are rare by design, often representing less than 5% of all units produced. This characteristic makes supervised approaches impractical, as they require large quantities of labeled faulty samples [14–16].

The present study has a different objective from conventional multi-class fault diagnosis, which is primarily used for condition monitoring and maintenance. In the observed EoL production environment, the primary inspection decision is binary: whether a newly manufactured motor satisfies the expected healthy behavior or should be rejected for further inspection. Identifying the precise fault type is a secondary diagnostic task that is not required for the initial EoL accept/reject decision. Furthermore, supervised multi-class classification requires representative labelled examples of each expected fault type. In high-quality manufacturing, individual defect categories occur infrequently, and new manufacturing faults may also emerge. Therefore, a one-class anomaly detection approach is particularly well-suited to this application, as the model can be trained using large amounts of healthy production data, without the need for a predefined set of fault classes or sufficient instances of every possible defect. The proposed method therefore provides a useful complement to multi-class fault diagnosis rather than replacing it. Once an anomalous motor has been detected, further diagnostic procedures can be employed to identify the specific fault mechanism.

This work proposes a deep learning approach for unsupervised anomaly detection in EoL quality inspection of electric motors. The method uses MFSs of acoustic and vibrational signals as input features and combines CNN and BiGRU layers to model both spectral and temporal characteristics of motor behavior. This paper builds on our earlier work on supervised EoL quality inspection using deep learning and MFSs [17–20]. The primary contribution of this paper is the elimination of the requirement for labeled faulty data during training. In contrast to the previous approach [17], which relied on the use of labeled examples of faulty motors to train a discriminative classifier, the present method learns exclusively from healthy motor data. This characteristic renders the method immediately applicable to new production lines or fault types without the need for re-labeling. A dataset of over 2400 motors was collected during production, with approximately 2300 healthy motors and about 100 faulty motors. The model was trained exclusively on healthy motor data, allowing it to learn the distribution of normal operating conditions and subsequently identify anomalous deviations associated with faults. The complete production dataset reflects a realistic fault rate of around 4.3%. For the final evaluation, all available faulty motors were included in the test set, alongside a subset of previously unseen healthy motors. This resulted in a test set that was intentionally enriched with faults. This provides a sufficiently large number of faulty samples to reliably evaluate the performance of fault detection, while preserving the naturally occurring class imbalance in the complete production dataset.

The main contributions of this work are summarized as follows:

- A hybrid CNN-BiGRU architecture, capab le of detecting anomalies in acoustic and vibrational signals from motors during production.
- A data-driven one-class approach for EoL quality inspection that is trained exclusively on healthy motors, thereby avoiding the need to train a discriminative classifier on the severely underrepresented faulty class.
- A direct quantitative comparison with our prior supervised deep learning approach [17], demonstrating competitive detection performance without requiring fault labels during training.
- An experimental validation on real-world production data with realistic class imbalance (~4.3% faults), demonstrating robust separation between healthy and faulty instances and potential integration into automated manufacturing procedures.

The remainder of the paper is organized as follows. Section 2 reviews the related work. Section 3 details the data collection and preprocessing, the proposed neural network architecture, and the training architecture. Section 4 presents the experimental setup, results and discussion, while Section 5 presents the conclusions and directions for future work.

2. Related Work

EoL quality inspection systems are commonly based on signal processing methods, using the analysis of electrical, acoustic or vibration signals in the time or frequency domain [5,21]. These approaches typically rely on manually selected features such as root mean square (RMS), kurtosis and other statistical indicators, combined with rule-based thresholds or machine learning classifiers [22,23]. Despite their effectiveness, these methods often require extensive expertise and are difficult to apply to irregular or evolving defects. Automated EoL inspection has been the subject of extensive research in a variety of industrial applications, including vacuum cleaner motors [7,8], geared motors [24,25], and permanent magnet synchronous motors [26]. These studies have repeatedly demonstrated the efficacy of data-driven signal analysis in comparison to purely rule-based inspection methods. Additionally, architectures employing artificial intelligence (AI) for the purpose of testing the reliability of electric motors once they have reached the end of their usable lifespan have been proposed. These architectures integrate multiple signal sources and classification layers [27]. More recently, electric motor fault detection using Symmetrized Dot Pattern-based features was investigated [21], further demonstrating the potential of advanced signal representations for extracting discriminative fault information. However, such feature-based approaches still depend on the suitability of the selected representation for the faults under consideration.

With deep learning methods, data-driven models have become increasingly popular for fault detection. CNNs have been widely adopted for processing spectrograms and other time-frequency representations of vibration signals [10]. In [9] it was demonstrated that CNN-based feature extraction for rotating machinery achieved strong performance compared to traditional signal processing-based methods. Similarly, ref. [28] proposed a deep CNN architecture with domain adaptation capabilities to address variability in operating conditions. MFSs have also proven effective for machinery fault detection beyond their original speech recognition context [29], capturing both spectral energy distribution and temporal evolution in a compact, perceptually motivated feature space. However, most deep learning approaches for diagnosing faults are developed to identify predefined categories of faults and therefore rely on representative labelled examples of the expected faults. This requirement is particularly problematic in EoL manufacturing, where naturally occurring faulty units are scarce.

More recently, hybrid architectures that combine CNNs with recurrent layers such as long short-term memory (LSTM) or gated recurrent units (GRU) have been explored to model temporal dependencies [30]. These hybrid models have proven to be effective, reliable and robust for capturing both spectral and sequential features, leading to improved classification performance in real-world settings. Specifically, CNN-BiGRU architectures have demonstrated strong performance in bearing fault diagnosis tasks, in which identifying the temporal context of the signal is critical for classifying fault types [31,32]. Furthermore, recurrent networks operating on fault signals have demonstrated an advantage in capturing evolving patterns that pure CNN approaches miss [33]. However, these studies primarily focus on supervised fault diagnosis or machinery condition monitoring rather than one-class EoL quality inspection based on naturally occurring production defects.

2.1. Unsupervised and Reconstruction-Based Anomaly Detection

A key challenge in industrial fault detection is the limited availability of labelled fault data, which has led to the development of one-class or unsupervised learning approaches. Autoencoder-based methods, which are trained exclusively on normal data, have emerged as a leading approach. The autoencoder learns a compact representation of normal operation, and the reconstruction error serves as an anomaly score during inference [34–36]. These methods have been used for anomalous sound detection, vibration-based structural health monitoring, and industrial control systems, consistently showing that reconstruction error effectively classifies normal and abnormal operating conditions, even without labelled fault examples during training. Unsupervised transfer learning methods have also been studied. In these methods, models trained on normal data from one domain are applied to detect anomalies in related but different operating conditions [37–39]. Recent work [23] addressed the issue of scarce labels in motor fault detection through cross-sensor self-contrastive learning. This demonstrates the growing interest in learning useful fault representations when labelled data are limited. However, this approach is centered on the detection of faults in heavy locomotives rather than the EoL quality inspection of manufactured products.

Despite these advances, most unsupervised or label-scarce approaches in machinery diagnostics focus on monitoring the condition of machines during operation, using laboratory fault datasets or specific known fault mechanisms. Relatively little attention has been given to one-class anomaly detection in EoL manufacturing inspection using combined acoustic and multi-point vibration measurements acquired from naturally occurring production faults. This is an important research gap because the main EoL requirement is to distinguish between acceptable and potentially defective products reliably when samples of all possible manufacturing faults are unavailable.

2.2. Class Imbalance in Industrial Fault Detection

A commonly overlooked aspect of real-world industrial fault detection is the significant class imbalance in production data. In high-quality manufacturing environments, defective units represent a small percentage of total production, typically between 1% and 10%, since the manufacturing process is engineered to minimize fault rates [4,14]. This imbalance creates a key challenge for supervised classifiers, which tend to be biased toward the majority (healthy) class. These classifiers may achieve high accuracy yet fail to detect most faults [14]. Cost-sensitive learning, resampling strategies, and one-class classification methods have been proposed to address this problem in rotating machinery fault diagnosis [14]. Instead of directly addressing class imbalance, the unsupervised autoencoder used in this study eliminates the requirement for training a discriminative

classifier on the poorly represented faulty class, as only healthy motors are utilised during the training process. The natural class imbalance is preserved in the complete production dataset. For evaluation, the test set was intentionally enriched with faulty motors to enable a more reliable estimation of fault-detection performance while keeping all healthy samples representative of real production.

In our previous work [17], we used MFSs and deep neural networks to build a robust and reliable motor quality inspection system. However, the supervised approach required labelled faulty data for training. In contrast, the suggested approach in this paper adopts a fully unsupervised training strategy with a CNN–BiGRU architecture, designed specifically for EoL quality inspection systems. This enables the separation of faulty units from healthy ones without using fault labels during model training, making it highly suitable for industrial manufacturing environments where representative faulty samples are scarce. A detailed review of AI-enabled EoL quality control methods for electric motor manufacturing was provided in [20]. This review identified unsupervised and semi-supervised approaches as an under-explored but promising direction for the field. Accordingly, this study aims to bridge the gap between recent advances in unsupervised machinery diagnostics and their practical application in EoL inspection systems, by using real industrial acoustic and vibration measurements.

3. Methodology

This section describes the suggested methodology for unsupervised anomaly detection in the EoL quality inspection system for electric motor production. The approach consists of four main steps: data acquisition, feature representation using MFSs, training a hybrid CNN–BiGRU neural network for anomaly detection and model evaluation.

The overall workflow of the proposed EoL quality inspection approach is presented in Figure 1, covering the process from acoustic and vibration signal acquisition and MFS generation to anomaly detection using the proposed CNN–BiGRU autoencoder.

3.1. Data Acquisition and Data Pre-Processing

The presented study is associated with the Slovenian company Domel d.o.o., one of the largest manufacturers of electric motors for various purposes [40]. Acoustic (S) and vibration signals from three points (V1, V2 and V3) were collected directly on the production line from the existing EoL quality inspection system, which has already been described in detail in previous work [17] and is shown in Figure 2.

A total of 2402 motors were recorded, including 2299 healthy motors and 103 faulty motors containing various types of defects detected during manufacturing (e.g., mechanical imbalance, bearing defects, rotor eccentricities). Data were recorded with industrial-grade sensors to ensure high signal quality and to minimize the influence of background noise and production variability. It is important to note that this dataset was collected under real production conditions, without any intentional fault induction or artificial resampling. The resulting fault rate of approximately 4.3% (103 out of 2402 motors) reflects the actual defect rate encountered on the production line. This represents a severely class-imbalanced setting, which is typical of real, high-quality manufacturing environments [14]. The dataset was divided into three independent subsets, as shown in Table 1. As shown in Table 1, the training and validation datasets consist of data from healthy motors only. However, the test dataset includes all available faulty motors, together with a subset of previously unseen healthy motors, to enable a more reliable estimation of fault-detection performance.

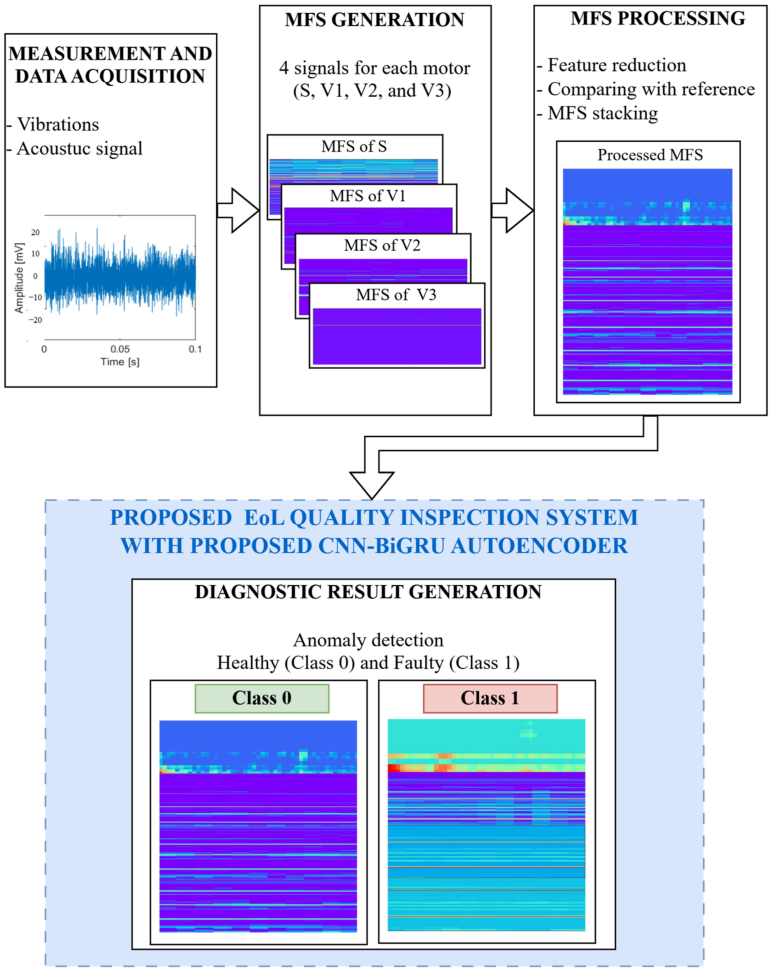


Figure 1. Overview of the proposed methodology for unsupervised EoL quality inspection of electric motors.



Figure 2. Existing EoL quality inspection system.

Table 1. Data subset presentation.

Subset Name	Description	Healthy Motors	Faulty Motors	Total
Training dataset	Used to train model.	1679	0	1679
Validation dataset	Used to validate model during training.	420	0	420
Test dataset	Used for final testing, completely unseen by model.	200	103	303

Only minimal preprocessing was applied, limited to normalizing the raw signals to remove amplitude bias and ensure compatibility between samples. No manual feature engineering or filtering was performed, preserving the raw signal characteristics and allowing the neural network to learn directly from the data. Signal preprocessing has already been described in detail in previous work [17–19]. This minimal preprocessing approach provides an advantage from an industrial integration perspective. It reduces the need for domain expertise and makes the pipeline more easily transferable to different motor types or production lines [27,41].

3.2. Feature Representation and MFSs Creation

Each preprocessed acoustic and vibration signal was transformed into an MFS representation. Mel scale was chosen because it transforms linear frequency to a scale that approximates human hearing, particularly enhancing low-frequency regions that are critical for detecting mechanical faults [17,23]. The use of MFSs for machine fault diagnosis has been validated across a range of industrial applications, including drill fault detection [29] and anomalous sound detection for predictive maintenance [37,42], which confirms its effectiveness as a compact and informative feature extraction method.

Each analyzed motor is represented by four MFSs. The four MFS representations correspond to the acoustic signal S and the vibration signals V1, V2 and V3. After feature reduction and reference-based processing, the four MFS representations were combined into a single stacked MFS for each motor. This integrated representation combines information from all four measurement signals into a single, two-dimensional input, instead of using the individual MFSs as separate input channels. Detailed MFS generation, processing and dimensionality reduction have already been explained and presented in [17]. Figure 3 illustrates the MFSs of healthy and faulty motors side by side after dimensionality reduction, as explained in [17]. The final stacked MFS served as the input to the model.

In this study, the resulting stacked spectrogram was created with a resolution of 520 Mel-frequency features $\times$ 48 time steps. To guarantee uniform input size, all spectrograms were zero-padded. This representation captures both spectral energy distribution and its temporal evolution, providing a compact feature representation for identifying subtle deviations from normal motor behavior [28,30].

Converting the original signals into MFS reduce the dimensionality of the input, which can result in the elimination of some narrow-band or high-frequency information present in the raw measurements. This is a trade-off between retaining the complete signal content and obtaining a compact time-frequency representation that is suitable for processing by a neural network. MFSs were selected because they preserve the dominant temporal and spectral energy patterns that are relevant to the mechanical deviations considered in the current EoL application, while reducing input complexity. Their effectiveness in the motor inspection system was also demonstrated in our previous work [17–19]. However, it should be noted that the proposed representation does not necessarily preserve all diagnostic information for every possible manufacturing fault.

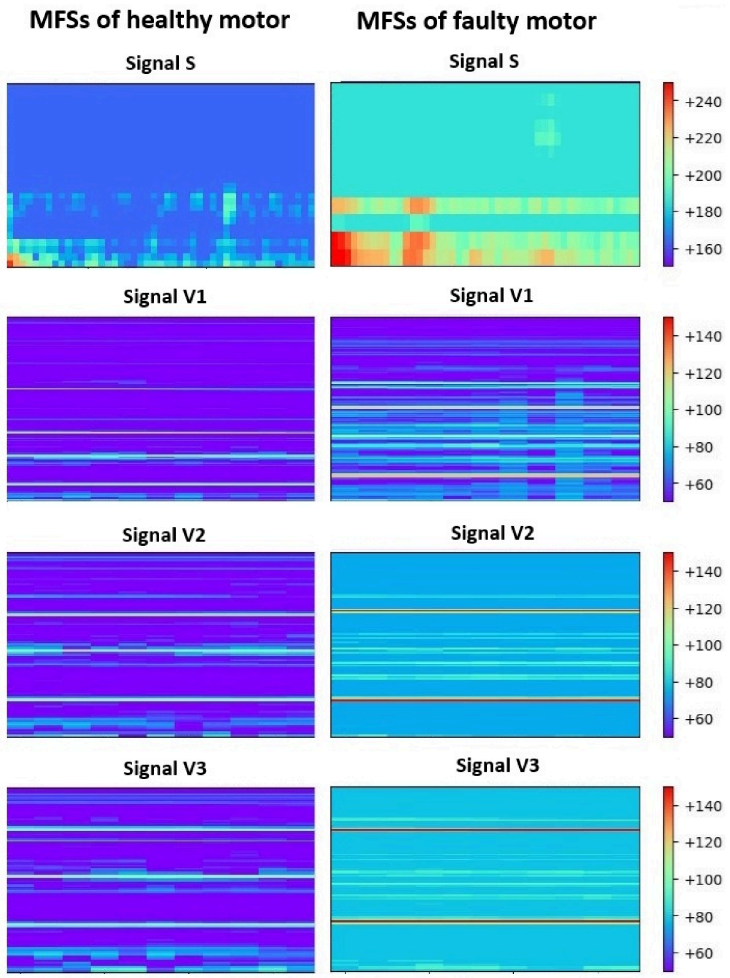


Figure 3. Comparing MFSs of healthy and faulty motors after dimensionality reduction.

3.3. Hybrid CNN-BiGRU Model and Training

For the presented task, a hybrid neural network architecture combining CNNs and BiGRUs was designed. CNN layers are implemented to extract hierarchical and localized patterns from the MFSs, identifying relevant frequency components and characteristic spectro-temporal structures. BiGRU layers are implemented to detect sequential dependencies by processing features in both forward and backward directions, enabling the model to detect evolving patterns that may indicate the occurrence or persistence of faults. Bidirectional processing is particularly important for motor signals. Mechanical faults, such as bearing defects and rotor eccentricities, produce evolving patterns over the duration of the measurement. Processing in both temporal directions enables the model to capture these dependencies more completely than a single-direction approach [31,32].

The model is constructed as an autoencoder for learning compact representations of motor signals. In the encoder, convolutional layers with batch normalization and max pooling capture spatial patterns from the MFSs. These features are then reshaped and passed through two stacked BiGRU layers to model the temporal evolution of the signals. A dense bottleneck layer compresses the information into a low-dimensional encoding. The decoder mirrors this structure, using GRU layers and time-distributed dense layers to reconstruct the original MFS, enabling the model to detect faults and anomalies. Table 2 presents the detailed architecture and hyperparameter configuration of the proposed CNN–BiGRU autoencoder.

Table 2. Neural network architecture and hyperparameter configurationof the proposed CNN-BiGRU autoencoder.

Layer (Type)	Output Shape	Details
Input	(None, 520, 48, 1)	Size of MFS
First Conv2D	(None, 520, 48, 32)	First 2D convolutional layer
First Batch Normalization	(None, 520, 48, 32)	Normalizes activations, improves training stability
First Max Pooling 2D	(None, 260, 24, 32)	Down-sampling
First Dropout	(None, 260, 24, 32)	Prevents overfitting, dropout value 0.5
Second Conv2D	(None, 260, 24, 128)	Second 2D convolutional layer
Second Batch Normalization	(None, 260, 24, 128)	Further normalization after convolution
Second Max Pooling 2D	(None, 130, 12, 128)	Down-sampling
Second Dropout	(None, 130, 12, 128)	Additional dropout, dropout value 0.5
Reshape	(None, 130, 1536)	Converting convolutional feature maps into a sequence for processing by the BiGRU.
First Bidirectional BiGRU	(None, 130, 128)	First BiGRU layer capturing bidirectional context
Second Bidirectional BiGRU	(None, 256)	Second BiGRU layer, flattening to a vector
First Dense Bottleneck	(None, 80)	Fully connected layer
Repeat Vector	(None, 48, 80)	Repeating the latent representation over 48 time steps in order to initialise the decoder
First Decoder BiGRU	(None, 48, 256)	Decoder BiGRU
Second Decoder BiGRU	(None, 48, 512)	Decoder BiGRU
Time Distributed Dense	(None, 48, 520)	Linear reconstruction of 520 spectral features at each time step
Output Reshape	(None, 520, 48, 1)	Reconstructs the original MFS dimensions

The model is trained end-to-end using the Adam optimizer and mean squared error (MSE) as the reconstruction loss. Training was performed on the training dataset, while model validation was performed on validation dataset. Both datasets contained only healthy motors, allowing the model to learn the distribution of normal operating conditions.

In order to avoid overfitting, an early stopping criterion with a patience of 20 epochs was applied. After training, reconstruction errors were computed for all samples. The decision threshold was defined as $(\mu + 3\sigma)$, where μ and σ are the mean and the standard deviation of the reconstruction errors for healthy validation motors. This criterion was selected because the distribution of healthy reconstruction errors was approximately Gaussian, providing a simple, conservative rule for identifying observations that substantially deviate from normal behavior. Motors whose reconstruction error exceeds this threshold are marked as faulty.

3.4. Evaluation

The model was evaluated using a dedicated test set, separate from both the training and validation datasets to ensure an unbiased assessment of performance. This set contained 103 faulty motors and 200 healthy motors, as shown in Table 1. The model's performance was assessed by:

- Reconstruction error distributions;
- Confusion matrix;

- Precision, recall, F1-score and accuracy;
- Macro- and weighted-average classification metrics;
- ROC–AUC;
- t-SNE visualization;
- Comparison with the previous supervised method [17].

The evaluation provides a comprehensive understanding of the model's ability to separate faulty motors from healthy ones in a real production environment. Faulty motors were treated as the positive class when calculating true positive rate (TPR) and false positive rate (FPR). The test set class distribution was selected to provide a sufficiently large number of faulty samples for evaluation, but it does not reflect the natural production fault prevalence of 4.3%. Therefore, prevalence-dependent metrics should be interpreted in the context of the selected test composition.

4. Experimental Results

The evaluation focuses on the model's ability to separate healthy from faulty motors using reconstruction errors from the hybrid CNN–BiGRU autoencoder.

4.1. Model Training and Convergence

Figure 4 shows the training and validation reconstruction losses of the proposed CNN–BiGRU autoencoder. Both losses decreased substantially during training, following similar trends and indicating stable learning with no clear evidence of overfitting. Throughout most of the training process, the validation loss remained comparable to or lower than the training loss. Near the end of training, a slight increase in validation loss was observed, followed by a subsequent decrease to its minimum value in the final epoch.

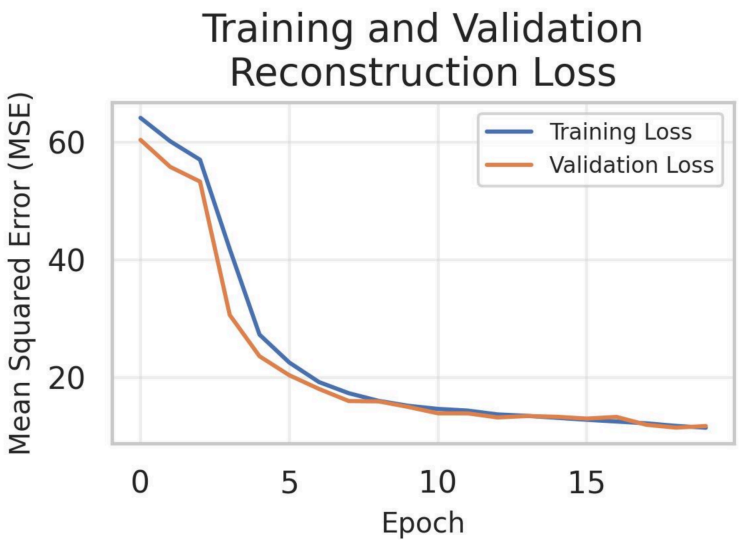


Figure 4. Training and validation reconstruction loss.

4.2. Reconstruction Error Analysis

After training, reconstruction errors (MSE between input and reconstructed MFSs) were computed for all test motors in the test dataset. Figure 5 illustrates the distribution of reconstruction errors for healthy (blue) and faulty (red) motors. In general, healthy motors showed low reconstruction errors, while faulty motors showed a noticeably higher reconstruction errors, confirming that the autoencoder effectively learned the normal operating patterns.

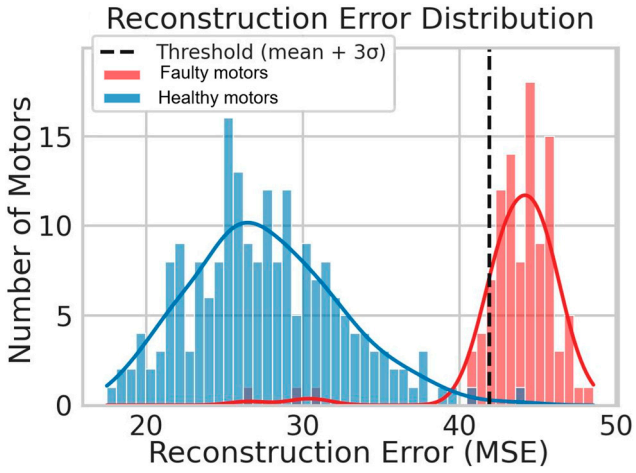


Figure 5. Distribution of reconstruction errors for healthy and faulty motors.

Following common industrial anomaly detection practice, a threshold was set at mean + 3σ because reconstruction errors for healthy samples approximately follow a Gaussian-like distribution. Motors exceeding this threshold were marked as faulty. With this criterion, the model achieved high separation between healthy and faulty motors, demonstrating the robustness of unsupervised training in detecting deviations from normal operating conditions.

4.3. Confusion Matrix

Figure 6 represents the confusion matrix which summarizes the model predictions for the test dataset.

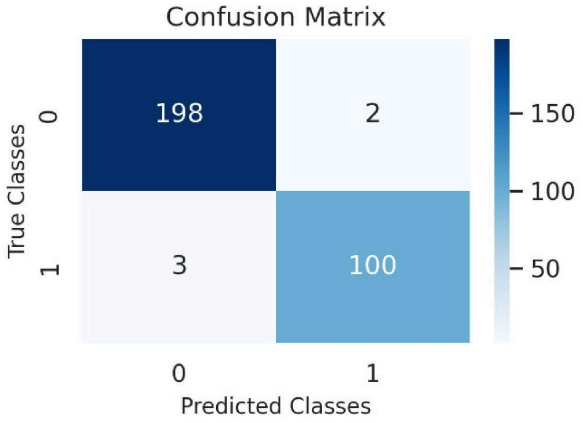


Figure 6. Confusion matrix of test dataset of healthy and faulty motors.

From this confusion matrix, the true positive rate (TPR) was computed as (1):

$$TPR = \frac{TP}{TP + FN} = \frac{100}{100 + 3} \approx 0.971 \quad (1)$$

where TP denotes true positive and FN false negative. Similarly, the false positive rate (FPR) was computed as (2):

$$FPR = \frac{FP}{FP + TN} = \frac{2}{2 + 198} = 0.01 \quad (2)$$

where FP denotes false positive and TN true negative. The low FPR of 1% is especially important in the context of industrial deployment. False alarms result in the unnecessary removal of healthy motors from the production line, which increases inspection costs

and reduces production. Therefore, the combination of a *TPR* greater than 97% and an *FPR* of 1% represents a practically useful optimal operating point for an EoL quality inspection [27,41,43].

4.4. Receiver Operating Characteristic (ROC) and Area Under the Curve (AUC)

The ROC curve was plotted using the reconstruction errors of all test motors and the AUC was computed. The resulting AUC value of 0.98 indicates excellent discriminative capability between healthy and faulty motors. Figure 7 presents the ROC curve, showing strong separation. An AUC of 0.98 is particularly noteworthy given that the model was trained exclusively on healthy data and was not exposed to any fault signatures. This demonstrates that the reconstruction error derived from the CNN-BiGRU autoencoder provides an effective anomaly score for the considered dataset, even in the presence of diverse fault types, such as mechanical imbalance, bearing defects, and rotor eccentricities.

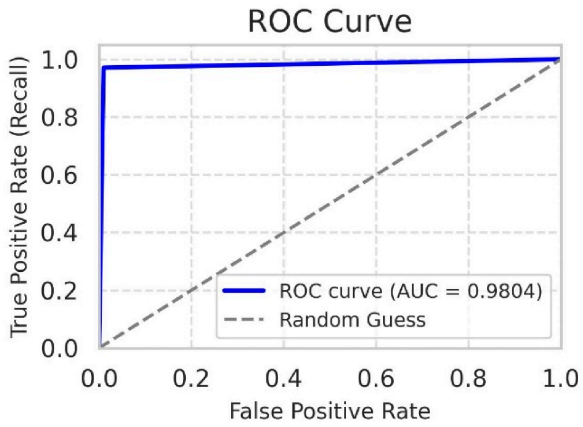


Figure 7. ROC curve for healthy and faulty motors.

4.5. t-Distributed Stochastic Neighbour Embedding (t-SNE) Visualization

To visualize the high-dimensional feature space learned by the model, t-SNE was applied to the latent representations extracted from the bottleneck layer of the autoencoder. Figure 8 shows that healthy motors (Class 0, blue) cluster tightly together, while faulty motors (Class 1, red) form distinct clusters or lie apart from the healthy cluster. This visualization further confirms the model's ability to capture meaningful variations in motor operating conditions. This representation could potentially be used to classify fault types in the future.

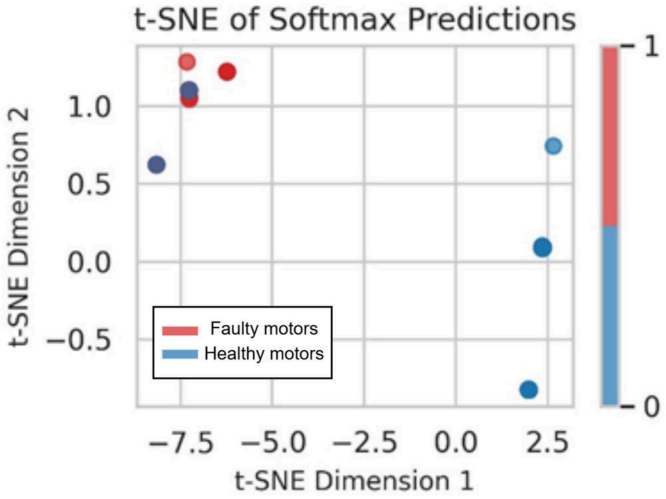


Figure 8. t-SNE visualization.

4.6. Comparison with Supervised Approach

To evaluate the impact of eliminating the need for labeled fault data on performance, we compared the proposed unsupervised CNN-BiGRU autoencoder against our previous supervised deep learning approach for the same EoL inspection system [17]. The previous approach used the same EoL inspection system and the same MFS-based feature representation but trained a supervised CNN-BiGRU classifier with labeled examples of healthy and faulty motors. Unlike the supervised method, which requires labeled examples of healthy and faulty motors during training, the proposed approach is trained exclusively on healthy motor data and detects anomalies based on reconstruction error. The reported performance metrics were obtained from separate experimental evaluations presented in this work and in [17]. Therefore, the comparison should be interpreted as indicative rather than as a direct benchmark on an identical test set. Table 3 summarizes the comparison.

Table 3. A comparison of the proposed unsupervised approach and the previous supervised method.

Metric	Supervised CNN–BiGRU [17]	Proposed CNN–BiGRU
Precision (Healthy motor)	1	0.985
Recall (Healthy motor)	0.987	0.99
F1-score (Healthy motor)	0.994	0.988
Precision (Faulty motor)	0.818	0.98
Recall (Faulty motor, also TPR)	1	0.971
F1-score (Faulty motor)	0.9	0.976
Accuracy	0.988	0.983
Macro Precision	0.909	0.983
Macro Recall	0.994	0.981
Macro F1-score	0.947	0.982
Weighted Precision	0.99	0.983
Weighted Recall	0.988	0.983
Weighted F1-score	0.989	0.984
AUC	0.99	0.98
Fault labels required for training	Yes	No
Applicable to previously unseen fault types *	No	Potentially

* Potentially, since the model learns only the distribution of healthy motor data.

The supervised model achieved slightly higher overall accuracy and a higher weighted F1 score. However, the proposed unsupervised approach demonstrated substantially higher precision for the fault class. This suggests that motors identified as faulty by the proposed method are much less likely to be false alarms while maintaining a high fault detection recall of 97.1%. Therefore, the F1 score for the fault class increased from 0.900 to 0.976, showing a better balance between false positives and false negatives.

From an industrial perspective, the main benefit of the proposed approach is that it eliminates the need for labeled fault samples during training. Since faulty motors are rare in high-quality manufacturing environments, collecting large, representative labeled datasets is costly and time-consuming. The proposed CNN–BiGRU autoencoder can be deployed immediately on new production lines and adapted to evolving fault conditions by learning only from healthy motor data. This makes it particularly suitable for practical EoL quality inspection applications, as it does not require additional fault labeling.

4.7. Discussion

The results demonstrate that, despite being trained exclusively on healthy motor data, the proposed CNN–BiGRU autoencoder can effectively distinguish between healthy and faulty motors in an industrial EoL quality inspection system. The model achieved an AUC of 0.98, a fault recall of 97.1%, and a false positive rate of 1%. This low false positive rate suggests that the chosen threshold (mean + 3σ) provides a favorable balance between identifying faulty motors and preventing the unnecessary rejection of healthy units. In practice, this trade-off is important because missed faults may result in defective products proceeding further in the production process, whereas excessive false alarms increase inspection costs and reduce production efficiency.

The complete production dataset contained only around 4.3% faulty motors, highlighting the significant class imbalance typically encountered in high-quality manufacturing. Since the proposed model is trained exclusively on healthy motors, the training procedure is independent of the availability of the minority fault class [14,44]. This is a significant practical advantage over supervised approaches, especially when introducing new products or production lines, for which representative faulty samples are unavailable. Comparing the new approach with a previous supervised approach [17] indicates that removing the requirement for labelled faulty training samples is possible while maintaining competitive detection performance. However, as the two approaches were evaluated in separate experiments, this comparison should be considered indicative rather than a direct benchmark.

The t-SNE visualization provides further support for these results by showing a separation between healthy and faulty motors in latent space. This suggests that the model has learned to distinguish between normal and abnormal motor behavior. As the model was trained without predefined fault classes, this property could also enable the detection of previously unseen fault types. However, this capability requires further experimental validation.

Minimal preprocessing was applied to the raw data, thereby reducing the need for extensive feature engineering or manual expertise. This makes it easier to integrate the proposed method into existing EoL inspection systems.

Several limitations should be considered. While all 103 available faulty motors were included in the evaluation, they do not represent every possible manufacturing fault. Furthermore, the test set was intentionally enriched with faulty motors and therefore does not reflect the natural 4.3% fault prevalence of the full production dataset. The current method employs a fixed anomaly threshold, which may need adjusting if signal distributions change due to sensor ageing, environmental conditions or variations in the manufacturing process. Therefore, adaptive thresholding based on streaming statistics could improve long-term robustness [45]. Finally, a systematic comparison with alternative unsupervised methods, including variational autoencoders (VAEs), generative adversarial networks (GANs) and contrastive learning, is an important area for future research. These methods might provide better uncertainty estimation or improved generalization to novel fault types and increase reliability in industrial applications.

5. Conclusions

This study presented an unsupervised CNN–BiGRU autoencoder for EoL quality inspection of electric motors, based on acoustic and vibration measurements taken in an industrial production environment. Training exclusively on healthy motors eliminates the need for labelled faulty samples during model development.

When evaluated on 303 motors that were previously unseen, including 103 faulty units, the approach achieved an area under the curve (AUC) of 0.98, a fault recall of 97.1%, a false positive rate of 1%, and a faulty-class F1-score of 97.6%. These results show that one-class

learning can effectively detect faults in the considered EoL application, while maintaining performance that is competitive with that of the previous supervised approach.

The main practical advantage is that the anomaly detector can be developed using healthy production data, eliminating the need to wait for sufficient examples of individual manufacturing faults to accumulate. Future work should evaluate the approach using larger datasets and additional production conditions, investigate adaptive anomaly thresholds and compare alternative unsupervised architectures systematically. Integrating the approach into real-time EoL inspection would represent a further step towards industrial deployment.

Author Contributions: Conceptualization, J.M. and G.D.; methodology, J.M. and G.D.; software, J.M. and B.P.; validation, B.P. and G.D.; formal analysis, G.D.; investigation, J.M.; resources, J.M. and B.P.; data curation, J.M. and B.P.; writing—original draft preparation, J.M.; writing—review and editing, G.D.; supervision, G.D. All authors have read and agreed to the published version of the manuscript.

Funding: Authors are grateful for the financial support of Slovenian Research and Innovation Agency in the scope of grant P2-0001 and L2-4454.

Data Availability Statement: The data that support the findings of this study are not publicly available due to confidentiality agreements with the manufacturer but are available from the corresponding author upon reasonable request.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial Intelligence
AUC	Area Under the Curve
BiGRU	Bidirectional Gated Recurrent Unit
CNN	Convolutional Neural Network
EoL	End-of-Line
FN	False Negative
FP	False Positive
FPR	False Positive Rate
GAN	Generative Adversarial Networks
GRU	Gated Recurrent Unit
LSTM	Long Short-Term Memory
MFS	Mel-Frequency Spectrogram
MSE	Mean Squared Error
ROC	Receiver Operating Characteristic
RMS	Root Mean Square
TN	True Negative
TP	True Positive
TPR	True Positive Rate
t-SNE	t-distributed Stochastic Neighbour Embedding
VAE	Variational Autoencoder

References

1. Folz, K.J.; Gomes, H.M. An Investigation of Machine Learning Strategies for Electric Motor Anomaly Detection Using Vibration and Audio Signals. *Eng. Comput.* **2024**, *42*, 465–487. [CrossRef]
2. Tang, M.; Liang, L.; Zheng, H.; Chen, J.; Chen, D. Anomaly Detection of Permanent Magnet Synchronous Motor Based on Improved DWT-CNN Multi-Current Fusion. *Sensors* **2024**, *24*, 2553. [CrossRef]
3. Popsi, N.R.S.; Anik, A.; Verma, R.; Viana, C.; Iyer, K.L.V.; Kar, N.C. Influence of Electric Motor Manufacturing Tolerances on End-of-Line Testing: A Review. *Energies* **2024**, *17*, 1913. [CrossRef]

4. Azamfirei, V.; Psarommatis, F.; Lagrosen, Y. Application of Automation for In-Line Quality Inspection, a Zero-Defect Manufacturing Approach. *J. Manuf. Syst.* **2023**, *67*, 1–22. [CrossRef]
5. Baldini, G.; Giuliani, R.; Geib, F. On the Application of Time Frequency Convolutional Neural Networks to Road Anomalies' Identification with Accelerometers and Gyroscopes. *Sensors* **2020**, *20*, 6425. [CrossRef]
6. Zhang, W.; Peng, G.; Li, C.; Chen, Y.; Zhang, Z. A New Deep Learning Model for Fault Diagnosis with Good Anti-Noise and Domain Adaptation Ability on Raw Vibration Signals. *Sensors* **2017**, *17*, 425. [CrossRef]
7. Benko, U.; Petrovčič, J.; Mussiza, B.; Juričić, D. A System for Automated Final Quality Assessment in the Manufacturing of Vacuum Cleaner Motors. *IFAC Proc. Vol.* **2008**, *41*, 7399–7404. [CrossRef]
8. Boškoski, P.; Petrovčič, J.; Musizza, B.; Juričić, A. An End-Quality Assessment System for Electronically Commutated Motors Based on Evidential Reasoning. *Expert. Syst. Appl.* **2011**, *38*, 13816–13826. [CrossRef]
9. Junior, R.F.R.; Areias, I.A.D.S.; Campos, M.M.; Teixeira, C.E.; Da Silva, L.E.B.; Gomes, G.F. Fault Detection and Diagnosis in Electric Motors Using 1d Convolutional Neural Networks with Multi-Channel Vibration Signals. *Measurement* **2022**, *190*, 110759. [CrossRef]
10. Janssens, O.; Slavkovikj, V.; Vervisch, B.; Stockman, K.; Loccufier, M.; Verstockt, S.; Van De Walle, R.; Van Hoecke, S. Convolutional Neural Network Based Fault Detection for Rotating Machinery. *J. Sound Vib.* **2016**, *377*, 331–345. [CrossRef]
11. Ye, L.; Ma, X.; Wen, C. Rotating Machinery Fault Diagnosis Method by Combining Time-Frequency Domain Features and CNN Knowledge Transfer. *Sensors* **2021**, *21*, 8168. [CrossRef]
12. Wang, L.; Zhang, X. Anomaly Detection for Automated Vehicles Integrating Continuous Wavelet Transform and Convolutional Neural Network. *Appl. Sci.* **2023**, *13*, 5525. [CrossRef]
13. Yang, H.; Liu, Z.; Ma, N.; Wang, X.; Liu, W.; Wang, H.; Zhan, D.; Hu, Z. CSRM-MIM: A Self-Supervised Pretraining Method for Detecting Catenary Support Components in Electrified Railways. *IEEE Trans. Transp. Electrif.* **2025**, *11*, 10025–10037. [CrossRef]
14. Xu, Q.; Lu, S.; Jia, W.; Jiang, C. Imbalanced Fault Diagnosis of Rotating Machinery via Multi-Domain Feature Extraction and Cost-Sensitive Learning. *J. Intell. Manuf.* **2020**, *31*, 1467–1481. [CrossRef]
15. Surucu, O.; Gadsden, S.A.; Yawney, J. Condition Monitoring Using Machine Learning: A Review of Theory, Applications, and Recent Advances. *Expert Syst. Appl.* **2023**, *221*, 119738. [CrossRef]
16. Tama, B.A.; Vania, M.; Lee, S.; Lim, S. Recent Advances in the Application of Deep Learning for Fault Diagnosis of Rotating Machinery Using Vibration Signals. *Artif. Intell. Rev.* **2023**, *56*, 4667–4709. [CrossRef]
17. Mlinarič, J.; Pregelj, B.; Dolanc, G. End-of-Line Quality Control Based on Mel-Frequency Spectrogram Analysis and Deep Learning. *Machines* **2025**, *13*, 626. [CrossRef]
18. Mlinarič, J.; Pregelj, B.; Boskoski, P.; Petrovcic, J.; Dolanc, G. A Transfer Learning Approach to Machine Learning-Based End-of-Line Quality Inspection. *Adv. Prod. Eng. Manag.* **2025**, *20*, 277–290. [CrossRef]
19. Mlinaric, J.; Pregelj, B.; Boskoski, P.; Dolanc, G.; Petrovcic, J. Optimization of Reliability and Speed of the End-of-Line Quality Inspection of Electric Motors Using Machine Learning. *Adv. Prod. Eng. Manag.* **2024**, *19*, 182–196. [CrossRef]
20. Mlinarič, J.; Dolanc, G. AI-Enabled End-of-Line Quality Control in Electric Motor Manufacturing: Methods, Challenges, and Future Directions. *Machines* **2026**, *14*, 149. [CrossRef]
21. Spirto, M.; Malfi, P.; Nicolella, A.; Melluso, F.; Cosenza, C.; Savino, S.; Niola, V. Fault Detection of Electric Motors via Symmetrized Dot Pattern-Based Features. *Int. J. Mech. Syst. Dyn.* **2026**. [CrossRef]
22. Defard, T.; Setkov, A.; Loesch, A.; Audigier, R. PaDiM: A Patch Distribution Modeling Framework for Anomaly Detection and Localization. In *Proceedings of the International Conference on Pattern Recognition*; Springer International Publishing: Cham, Switzerland, 2020.
23. Xu, Y.; He, D.; Luo, X.; Jin, Z.; Sun, H. CSSCL: A Cross-Sensor Self-Contrastive Learning for Nondestructive Fault Detection of Heavy Locomotive Motors in Label-Scarce Scenarios. *Nondestruct. Test. Eval.* **2026**, 1–34. [CrossRef]
24. Wissbrock, P.; Pelkmann, D.; Tölle, B. Automate Quality Prediction in an End-of-Line Test of a Highly Variant Production of Geared Motors—Discussion of a Full Concept. *PHM Soc. Eur. Conf.* **2021**, *6*, 10. [CrossRef]
25. Wißbrock, P.; Koschek, L.; Ren, Z.; Nejdl, W. Enhancing Quality Inspection of Highly Variant Geared Motors. *Appl. Acoust.* **2025**, *235*, 110687. [CrossRef]
26. Garcia-Calva, T.; Morinigo-Sotelo, D.; Fernandez-Cavero, V.; Romero-Troncoso, R. Early Detection of Faults in Induction Motors—A Review. *Energies* **2022**, *15*, 7855. [CrossRef]
27. Soyturk, M.; Coskun, K.; Izmitlioglu, O.; Tumer, B.; Gunes, D.; Saracoglu, S.; Bulut, B.; Ketmen, H.B.; Hanedar, I.; Asan, T.; et al. An AI-Based Architecture Framework for Improving End-of-Line Reliability Tests of Electric Motors. In *Proceedings of the IECON 2022—48th Annual Conference of the IEEE Industrial Electronics Society*; IEEE: Brussels, Belgium, 2022; pp. 1–6.
28. Gao, Y.; Kim, C.H.; Kim, J.-M. A Novel Hybrid Deep Learning Method for Fault Diagnosis of Rotating Machinery Based on Extended WDCNN and Long Short-Term Memory. *Sensors* **2021**, *21*, 6614. [CrossRef]
29. Tran, T.; Lundgren, J. Drill Fault Diagnosis Based on the Scalogram and Mel Spectrogram of Sound Signals Using Artificial Intelligence. *IEEE Access* **2020**, *8*, 203655–203666. [CrossRef]

30. Lee, H.; Yu, J. A Fault Detection Framework for Rotating Machinery with a Spectrogram and Convolutional Autoencoder. *Appl. Sci.* **2025**, *15*, 7698. [CrossRef]
31. Li, C.; Dong, C.; Fang, T. Bearing Fault Diagnosis Based on FFT-CNN-BiGRU-Attention. In *Proceedings of the 4th International Conference on Computer, Artificial Intelligence and Control Engineering*; ACM: Heifei, China, 2025; pp. 880–886.
32. Yasenjiang, J.; Zhao, Y.; Xiao, Y.; Hao, H.; Gong, Z.; Han, S. Bearing Fault Diagnosis Based on Time–Frequency Dual Domains and Feature Fusion of ResNet-CACNN-BiGRU-SDPA. *Sensors* **2025**, *25*, 3871. [CrossRef]
33. Zhang, Y.; Zhou, T.; Huang, X.; Cao, L.; Zhou, Q. Fault Diagnosis of Rotating Machinery Based on Recurrent Neural Networks. *Measurement* **2021**, *171*, 108774. [CrossRef]
34. Thill, M.; Konen, W.; Wang, H.; Bäck, T. Temporal Convolutional Autoencoder for Unsupervised Anomaly Detection in Time Series. *Appl. Soft Comput.* **2021**, *112*, 107751. [CrossRef]
35. Holly, S.; Heel, R.; Katic, D.; Schoeffl, L.; Stiftinger, A.; Holzner, P.; Kaufmann, T.; Haslhofer, B.; Schall, D.; Heitzinger, C.; et al. Autoencoder Based Anomaly Detection and Explained Fault Localization in Industrial Cooling Systems. *arXiv* **2022**, arXiv:2210.08011. [CrossRef]
36. Zhang, Y.; Li, X.; Gao, L.; Chen, W.; Li, P. Intelligent Fault Diagnosis of Rotating Machinery Using a New Ensemble Deep Auto-Encoder Method. *Measurement* **2020**, *151*, 107232. [CrossRef]
37. Michau, G.; Fink, O. Unsupervised Transfer Learning for Anomaly Detection: Application to Complementary Operating Condition Transfer. *Knowl.-Based Syst.* **2021**, *216*, 106816. [CrossRef]
38. Suefusa, K.; Nishida, T.; Purohit, H.; Tanabe, R.; Endo, T.; Kawaguchi, Y. Anomalous Sound Detection Based on Interpolation Deep Neural Network. In *Proceedings of the ICASSP 2020—2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*; IEEE: Barcelona, Spain, 2020; pp. 271–275.
39. Yan, P.; Abdulkadir, A.; Luley, P.-P.; Rosenthal, M.; Schatte, G.A.; Grewe, B.F.; Stadelmann, T. A Comprehensive Survey of Deep Transfer Learning for Anomaly Detection in Industrial Time Series: Methods, Applications, and Directions. *IEEE Access* **2024**, *12*, 3768–3789. [CrossRef]
40. Domel. About Us. Available online: https://www.domel.com/about-us (accessed on 8 July 2026).
41. Sundaram, S.; Zeid, A. Artificial Intelligence-Based Smart Quality Inspection for Manufacturing. *Micromachines* **2023**, *14*, 570. [CrossRef]
42. Di Fiore, E.; Ferraro, A.; Galli, A.; Moscato, V.; Sperlì, G. An Anomalous Sound Detection Methodology for Predictive Maintenance. *Expert Syst. Appl.* **2022**, *209*, 118324. [CrossRef]
43. Nunes, C.; Nunes, R.; Pires, E.J.S.; Barroso, J.; Reis, A. A Machine Learning Tool to Monitor and Forecast Results from Testing Products in End-of-Line Systems. *Appl. Sci.* **2023**, *13*, 2263. [CrossRef]
44. Choi, W.-H.; Kim, J. Unsupervised Learning Approach for Anomaly Detection in Industrial Control Systems. *Appl. Syst. Innov.* **2024**, *7*, 18. [CrossRef]
45. Wang, Y.; Li, J.; Yang, B.; Li, H. Stream-Data-Clustering Based Adaptive Alarm Threshold Setting Approaches for Industrial Processes with Multiple Operating Conditions. *ISA Trans.* **2022**, *129*, 594–608. [CrossRef]

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