

# Fault Detection in Industrial Pumps Using Cyclic Frequencies with Autoencoder

Daniela M. da Silva

*Postgraduate in Electrical and Computer Engineering*  
*Federal University of Rio Grande do Norte*  
Natal-RN, Brazil  
daniela.maia.064@ufrn.edu.br

Luiz F. Q. Silveira

*Department of Computer Engineering and Automation*  
*Federal University of Rio Grande do Norte*  
Natal-RN, Brazil  
felipe.silveira@ufrn.br

**Abstract**—Industrial machines play a fundamental role in the productive sector, being responsible for maintaining efficiency and operational continuity across various industrial segments. Early fault detection in such equipment, through machine learning techniques, is essential to increase operational reliability, reduce corrective maintenance costs, and avoid unplanned downtimes. In this work, we propose an approach for anomaly detection in industrial pumps based on acoustic signal analysis, using Spectral Correlation Density (SCD) combined with unsupervised learning models. SCD is employed to capture cyclostationary spectral patterns present in the signals, allowing the extraction of maximum-value vectors per cyclic frequency ( $\alpha$ -profile). Subsequently, the Random Forest algorithm is used to select the most relevant cyclic frequencies, resulting in a lower-dimensional yet highly discriminative representation. The extracted descriptors are used as input to autoencoders trained exclusively on normal data, with anomaly detection performed based on reconstruction error. The methodology was evaluated using real data from the MIMII Dataset, covering sections 00, 02, and 04, and showed consistent performance, with AUCs above 0.90 and an average F1-score of 0.80, outperforming previous spectrogram-based approaches.

**Index Terms**—Fault detection, Industrial pumps, Cyclostationary signals, Spectral correlation density, Machine learning.

## I. INTRODUCTION

Automated monitoring of industrial equipment plays a key role in enhancing operational reliability, reducing corrective maintenance costs, and preventing unexpected downtime. In this context, acoustic signals have gained prominence due to their non-invasive monitoring capabilities and ease of acquisition in industrial environments, making them a viable alternative to traditionally used vibration signals. [2]

In industrial processes, especially in continuously operating rotating machinery, acoustic and vibratory signals often exhibit non-stationary characteristics. These signals frequently present cyclostationary behavior, evidenced by statistical patterns that repeat over time as a result of rotational or oscillatory motion of mechanical components. Analytical methods that explore such time-varying spectral properties tend to achieve better results in anomaly detection, particularly under high-noise conditions. The Cyclic Spectral Density (SCD) stands out in this context as an effective technique for capturing cyclically modulated spectral components, revealing patterns that conventional spectral approaches fail to detect [1].

Recent studies have explored various cyclostationary analysis strategies for fault diagnosis. [17] proposed a method using acoustic signals collected along railway tracks, applying cyclic analysis to identify bearing faults. On the other hand, most works in the literature rely on vibration signals. [10] employed cyclostationary analysis combined with statistical tests to detect faults in electric motor bearings. [7] modeled vibration signals as bi-periodically correlated processes in wind turbines, while [15] combined Ensemble Empirical Mode Decomposition (EEMD) with cyclic analysis to isolate modulated patterns in noisy environments.

Other studies have also sought to extract relevant cyclic information in challenging scenarios. [16] proposed a condition monitoring method for subway ventilation fans subjected to piston wind effects, which degrade the quality of vibration signals. Their strategy combined a similarity-based detection technique with second-order cyclostationary analysis and was experimentally validated on axial and jet fans with bolt-loosening faults. [4] investigated an SCD-based approach to detect bearing faults in noisy signals, demonstrating that cyclostationary analysis can identify modulated components even under low signal-to-noise conditions.

More recently, [9] proposed a complete methodology for acoustic anomaly detection using the MIMII dataset, leveraging MFCC features, Deep Denoising Autoencoders, PCA with MAP-based selection, and traditional unsupervised models like Isolation Forest and LOF. Despite not using cyclostationary analysis, their approach shares conceptual similarities with ours in terms of autoencoder-based reconstruction and dimensionality reduction.

Additionally, [11] explored the use of SCD for fault diagnosis in induction motor bearings, identifying alarm regions in cyclic spectra via statistical metrics. However, their approach does not incorporate machine learning or deep models. In contrast, the present work extends the application of SCD by integrating it with Random Forest-based feature selection and unsupervised learning through autoencoders, offering improved adaptability and robustness for industrial anomaly detection under noisy and variable conditions.

Despite advances in the field, the application of SCD techniques to acoustic signals in industrial environments remains underexplored. Moreover, studies combining cyclic spectral

descriptors with computationally efficient classifiers, such as Random Forest, or unsupervised learning models, such as autoencoders, are still scarce. This gap is especially relevant in scenarios that demand high accuracy and fast response times using only easily acquired acoustic signals.

This work proposes a fault detection approach for industrial pumps based on the extraction of spectral descriptors using SCD, followed by feature selection via *Random Forest* and detection through unsupervised learning models, with emphasis on the use of autoencoders. The proposed method is evaluated using real data from the *MIMII Dataset*, considering multiple operational sections. The results demonstrate the effectiveness of the approach under different operating conditions and its superiority compared to conventional spectrogram-based methods.

The remainder of this paper is organized as follows: Section II details the methodology, including feature extraction and classification; Section III describes the dataset; Section IV presents and discusses the results; and Section V provides the conclusions and future directions.

## II. PROPOSED METHODOLOGY

This section presents the proposed methodology for fault detection in industrial pumps, which integrates the spectral analysis of acoustic signals through SCD with machine learning techniques. First, acoustic signals are collected and the SCD is computed to extract cyclic features. Then, the Random Forest algorithm is employed to perform feature selection based on feature importance. Subsequently, anomaly detection is carried out using an Autoencoder model in an unsupervised learning setting. The motivations and theoretical foundations underlying this approach are also discussed. The overall workflow is illustrated in Figure 1.

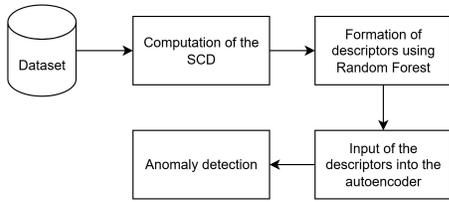


Fig. 1. Method workflow.

### A. Model Overview

The proposed method aims at the automatic and robust detection of faults in industrial pumps through cyclostationary analysis of acquired acoustic signals. The adopted approach combines the extraction of spectral features using SCD with a classifier based on *Random Forest*. The processing flow consists of the following main steps:

- 1) Acquisition of acoustic signals from machines under normal and faulty conditions;
- 2) Computation of the SCD of the acquired signals;
- 3) Careful selection of the most relevant cyclic frequencies;

- 4) Automatic classification using *Random Forest* to extract the most significant  $\alpha$  values;
- 5) Fault detection using an unsupervised learning network, the Autoencoder.

### B. SCD Computation

In digital signal processing, it is common to assume that signals are stationary, which allows the application of classical and computationally efficient techniques such as the Fast Fourier Transform (FFT). However, many signals encountered in real-world applications, including those generated by industrial mechanical systems, are in practice non-stationary. Among these, the class of cyclostationary signals stands out, whose main characteristic is the periodicity of their statistical moments. Formally, a signal  $x(t)$  is said to be second-order cyclostationary if its mean  $m_x(t)$  and autocorrelation function  $R_x(t, \tau)$  are periodic functions in time  $t$ , with period  $T$ , which implies:

$$m_x(t) = E[x(t)] = m_x(t + T) \quad (1)$$

$$R_x(t, \tau) = E \left[ x \left( t + \frac{\tau}{2} \right) x \left( t - \frac{\tau}{2} \right) \right] = R_x(t + T, \tau) \quad (2)$$

These properties allow  $R_x(t, \tau)$  to be represented as a Fourier series in  $t$ , whose coefficients correspond to the cyclic autocorrelation of the signal for different cyclic frequencies  $\alpha$ .

In industrial contexts, acoustic signals from rotating machines and other equipment with regular operational cycles often exhibit cyclostationary behavior, reflecting the intrinsic repetitiveness of mechanical operations. In such cases, conventional spectral analysis techniques may be insufficient to highlight subtle periodic components, especially in noisy environments. Cyclostationary analysis, particularly the computation of the SCD, has proven to be especially effective in revealing these hidden periodicities and correlating different spectral bands, thus enabling a more robust and informative signal characterization [5].

Formally, the SCD is defined as the Fourier transform of the Cyclic Autocorrelation Function (CAF), given by:

$$S_x^\alpha(f) = \int_{-\infty}^{+\infty} R_x^\alpha(\tau) e^{-j2\pi f\tau} d\tau \quad (3)$$

where  $R_x^\alpha(\tau)$  is the Fourier coefficient of the periodic autocorrelation function, calculated as:

$$R_x(t, \tau) = \sum_{\alpha} R_x^\alpha(\tau) e^{-j2\pi\alpha t} \quad (4)$$

In  $S_x^\alpha(f)$ , the variable  $f$  represents the spectral frequency, while  $\alpha$  is the cyclic frequency associated with the statistical periodicity of the signal. The SCD quantifies the correlation between spectral components separated by  $\alpha$ , i.e., located at  $f + \alpha/2$  and  $f - \alpha/2$ , revealing modulation or periodic structures in the signal that would not be captured by conventional spectral methods [6]. This property is particularly useful for identifying cyclically modulated components in industrial signals, where periodicities are associated with the rotational or oscillatory behavior of machinery.

In practice, SCD is computed over a discrete grid of spectral frequencies  $f$  and cyclic frequencies  $\alpha$ , resulting in a two-dimensional complex matrix:

$$S_x^\alpha(f) \in \mathbb{C}^{N_\alpha \times N_f} \quad (5)$$

where  $N_\alpha$  represents the number of discrete samples of the cyclic frequency  $\alpha$ , and  $N_f$  is the number of samples of the spectral frequency  $f$ . Each row of the matrix corresponds to a fixed value of  $\alpha$ , while each column represents a frequency  $f$ . To make its estimation computationally feasible, SCD is implemented using successive Fast Fourier Transforms (FFTs), following well-established approaches in the literature [5].

### C. Alpha-Profile

After computing the SCD matrices, each audio clip is represented by a one-dimensional vector that characterizes the magnitude profile as a function of the cyclic frequency, referred to as the *alpha-profile*. This vector, denoted as  $v(\alpha)$ , is constructed by selecting, for each positive value of  $\alpha$ , the maximum spectral magnitude across  $f$ , that is:

$$v(\alpha) = \max_f |S_x^\alpha(f)| \quad (6)$$

Since only the positive values of  $\alpha$  are considered, the vector  $v(\alpha) \in \mathbb{R}^{N_\alpha/2}$  represents the cyclic spectral signature of the audio clip. The *alpha-profile* results in a significant dimensionality reduction by condensing the two-dimensional complex matrix  $S_x^\alpha(f)$  into a one-dimensional real vector  $v(\alpha)$ . Considering only positive  $\alpha$  values further contributes to this reduction.

### D. Selection of Relevant Cyclic Frequencies and Descriptor Construction

In the feature extraction stage, the input signal, represented by cyclic frequencies obtained through the SCD, undergoes a selection process in which the components with the highest magnitudes are identified and stored. These predominant frequencies, due to their relevance in representing signal characteristics, are stored for each analyzed audio clip. Simultaneously, each clip is labeled as class 0 for normal conditions and class 1 for anomalous conditions. This process results in a structured dataframe that will be utilized in the classification stage.

This step is essential for dimensionality reduction, as each audio sample, after computing the SCD, generates a two-dimensional matrix with dimensions of 501x501. To make the analysis more efficient, only the values corresponding to positive alphas were considered, effectively halving the data volume. Consequently, each audio sample is represented as a 1x251 vector, and the resulting dataframe encompassing all audio samples has dimensions of 1,149x251 for section 00, 1,116x251 for section 02, and 802x251 for section 04. These represent the descriptors used for classification.

### E. Random Forest Classifier

The Random Forest algorithm was adopted in this study with the goal of classifying anomalies in acoustic signals while simultaneously identifying which cyclic frequency indices (*alphas*) are most relevant for this task. In addition to its effectiveness in classification problems, Random Forest provides the Feature Importance mechanism, which estimates the contribution of each variable to the model's decision-making process. This analysis is based on the reduction of node impurity (such as the Gini index) or the decrease in performance when variable values are permuted [3].

The implementation used in this study was the `RandomForestClassifier` from the `scikit-learn` library, which offers a robust and flexible interface for training and evaluating ensemble-based models. Based on the feature importance evaluation, it is possible to select only the most significant *alphas* and their respective frequencies, thereby reducing the dimensionality of the data without compromising the anomaly detection capability. This step is essential for optimizing the performance of the subsequent model based on unsupervised learning.

### F. Autoencoder architecture

For the anomaly classification task using an unsupervised approach, an autoencoder network is employed. This model is characterized by its unsupervised learning nature and is trained exclusively on data considered normal. In its basic structure, the autoencoder consists of three main components: the encoder, the latent space, and the decoder.

The encoder is responsible for receiving the input data and extracting the most relevant features, projecting this information into a lower-dimensional latent space. This latent space acts as a compressed representation of the original data.

Subsequently, the decoder reconstructs the data from the latent representation, aiming to reproduce the original inputs with the smallest possible error. During this process, the reconstruction error is calculated, representing the difference between the original and reconstructed data. Since the model is trained only on normal data, it is expected to yield low reconstruction errors for such inputs. In contrast, when anomalous data are presented to the model, the reconstruction error tends to be significantly higher. Based on this behavior, a threshold can be established to distinguish between normal and anomalous patterns.

The autoencoder was implemented using the TensorFlow and Keras libraries, specifically employing the following components: `from tensorflow.keras.models import Model`, `from tensorflow.keras.layers import Input, Dense, BatchNormalization, Activation`, and `from tensorflow.keras.optimizers import Adam`. These tools provided the necessary flexibility and performance for designing and training the neural network architecture used in this study.

The autoencoders used in this study are implemented as fully connected neural networks (Multilayer Perceptrons –

MLPs) with different architectural configurations depending on the machine section analyzed. Specifically, for sections 02 and 04, an asymmetric architecture was adopted, consisting of three dense layers in the encoder and two in the decoder, connected through a latent representation of size 64. For section 00, a deeper and symmetric structure was used, with four encoding layers and four decoding layers, centered on a compact latent space of dimension 8.

In both cases, each layer is followed by a non-linear activation function (Leaky ReLU for sections 02 and 04; ReLU for section 00) and batch normalization, which helps stabilize training. The models were trained using the Adam optimizer with a learning rate of  $10^{-3}$ , employing early stopping based on validation loss. The loss functions varied according to the section: Mean Absolute Error (MAE) was used for sections 02 and 04, while Mean Squared Error (MSE) was adopted for section 00. These design decisions were defined empirically through iterative experimentation, aiming to achieve high reconstruction quality and robustness to noisy industrial data.

### III. DATASET

#### A. Description

The data used in this study comes from the MIMII Dataset [14]. This dataset comprises 32,157 audio clips, of which 26,092 are labeled as normal, each with a duration of 10 seconds and sampled at 16 kHz. The audio clips were recorded from four types of industrial machines: fans, valves, pumps, and slide rails. Each machine type includes 7 different sections, numbered from 00 to 06, containing both normal and abnormal sounds. The sections represent different machine models while maintaining consistent types of anomalous conditions.

The original study introducing this dataset [13] employed Autoencoders for anomaly detection, using Mel-spectrograms as input descriptors. Reported accuracy results were: 84% for section 00, 45% for section 02, and 99% for section 04, indicating significant performance variations across different sections.

#### B. Types of Faults Considered

In this study, audio clips from pumps with a signal-to-noise ratio (SNR) of +6 dB were used, specifically from sections 00, 02, and 04. These pumps operate under discharge suction conditions for a water reservoir, and the anomalous conditions present in the data include leaks, contamination, clogging, and other representative faults typically found in industrial environments. It is important to note that the dataset does not include different degrees or severities of anomalies — each fault is treated uniformly without gradation [13].

#### C. Evaluation Metrics

To assess the performance of the proposed model, the following metrics were used: confusion matrix, accuracy, precision, recall, and F1-score. The confusion matrix measures the effectiveness of the model on a test dataset by comparing its predictions against previously established true values [8].

Precision and recall metrics provide insights into the accuracy of the model's predictions. Precision considers false positives, whereas recall accounts for false negatives. The F1-score is computed as the harmonic mean of precision and recall, balancing both metrics into a single measure. Accuracy, in turn, evaluates the model's overall performance, representing the proportion of correct predictions relative to the total number of analyzed samples [12]. These metrics can be formally defined by the following equations:

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \quad (7)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (8)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (9)$$

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (10)$$

Where TP represents the correctly classified positive cases, TN represents the correctly classified negative cases, FP refers to negative cases classified as positive, and FN indicates positive cases classified as negative.

### IV. RESULTS AND DISCUSSION

This section presents the experimental results of the proposed methodology, including examples of the generated spectral profiles and the performance evaluation of the Random Forest and Autoencoder models. The most relevant features identified by Random Forest were used as descriptors in the anomaly detection step performed by the Autoencoder.

#### A. Data Splitting and Preprocessing

The experiments were conducted using acoustic data from industrial pumps from sections 00, 02, and 04 of the MIMII Dataset under +6 dB SNR condition. Table I summarizes the number of normal and anomalous segments per section.

TABLE I  
DATASET COMPOSITION BY SECTION.

Type of machine	Model ID	Segments in normal condition	Segments in anomalous condition
Pump	00	1006	143
	02	1005	111
	04	702	100
Total		2713	354

The audio signals were randomly and stratified split into training (80%) and testing (20%) sets to ensure proper validation of the Random Forest classifier, preserving the ratio between normal and anomalous samples.

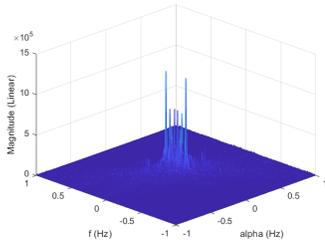


Fig. 2. SCD function of the normal acoustic signal (section 00).

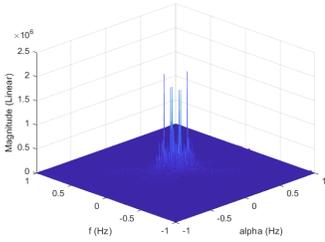


Fig. 3. SCD function of the anomalous acoustic signal (section 00).

### B. Descriptor Selection with Random Forest

Figures 2 and 3 illustrate the SCD for examples of acoustic signals captured under normal and anomalous conditions for section 00, while Figures 4 and 5 show the corresponding alpha-profiles,  $v(\alpha)$ , of these signals, highlighting differences in amplitudes associated with cyclic frequencies.

The Random Forest classifier was configured with 100 estimators, `log_loss` as the split criterion, and the maximum number of features evaluated per node set to `log2` for all sections (00, 02, and 04). These hyperparameters were defined based on empirical testing, where different configurations were evaluated through cross-validation to identify the combination that provided the best trade-off between accuracy, generalization capacity, and robustness. The choice of `log_loss` aimed to improve the probabilistic calibration of the model, while limiting the number of features per split to `log2` helped control overfitting.

The data were organized into a structured *dataframe*, associating each instance with its respective label (normal or anomalous), as illustrated in Table II.

Table IV summarizes the accuracies obtained for the models

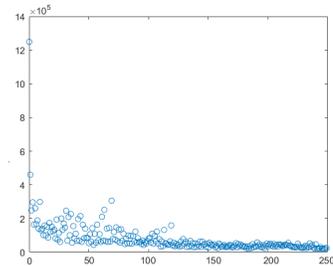


Fig. 4. Alpha-profile for the normal acoustic signal (section 00).

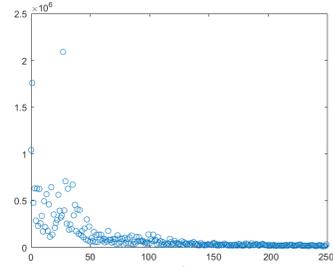


Fig. 5. Alpha-profile for the anomalous acoustic signal (section 00).

TABLE II  
EXAMPLE OF THE DATAFRAME STRUCTURE.

Audio	$\alpha_1$	$\alpha_2$	...	$\alpha_{N_\alpha/2}$	Class
1	value	value	...	value	0
2	value	value	...	value	1
⋮	⋮	⋮	...	⋮	⋮
N	value	value	...	value	0

trained on the analyzed sections, including the corresponding average values. Each section represents a distinct machine model, and the variations in performance among them can be attributed to differences in the spectral and operational characteristics of the respective acoustic signals.

TABLE III  
AUC PER EVALUATED MACHINE.

Type of machine	ID of machine	AUC
	00	0.96
Pump	02	0.96
	04	0.97
Avg.		0.96

During the construction of the decision trees, certain cyclic frequency indices proved to be especially relevant for the classification process. Based on this, a subset composed of the 30 most significant indices was selected using the *feature importances* method from the Random Forest algorithm. From these *alpha* indices, the corresponding columns of the SCD were extracted, where each column represents a frequency vector associated with its respective *alpha* value. These vectors were then used as input descriptors for the Autoencoder model in the anomaly detection stage.

### C. Autoencoder Training

With the extracted and organized descriptors, the data were split into 70% for training and 30% for testing. It is important to emphasize that only normal samples were used during training, while the test set consisted of a combination of normal and anomalous data, aiming to evaluate the model's ability to detect deviations based on reconstruction error.

Sections 02 and 04 employed an autoencoder architecture composed of asymmetric encoding and decoding layers. The encoder network  $E()$  consists of the following layers: *Flatten*(Input), *FC*(256, *Leaky ReLU*), *Batch Nor-*

malization;  $FC(128, Leaky ReLU)$ ,  $Batch Normalization$ ; and  $FC(64, None)$ , representing the latent space. The decoder network  $D()$  is composed of:  $FC(128, Leaky ReLU)$ ,  $Batch Normalization$ ;  $FC(256, Leaky ReLU)$ ,  $Batch Normalization$ ; and  $FC(output, sigmoid)$ , followed by a *Reshape* operation to restore the original input shape. In this context,  $FC(a, b, f)$  denotes a fully connected layer with  $a$  input neurons,  $b$  output neurons, and activation function  $f$ , while *Batch Normalization* refers to batch normalization. The *Leaky ReLU* activation function is used to mitigate the dead neuron problem. The model was trained using the Adam optimizer, with a learning rate of  $10^{-3}$ , for up to 100 epochs, employing *early stopping* with a patience of 10 epochs, monitoring validation loss. The loss function used was the *Mean Absolute Error* (MAE).

For section 00, a different architecture was adopted, characterized by a deeper and symmetric structure. The input was first flattened, resulting in a dimension of  $d = 30 \times 64$ . The encoder  $E()$  was composed of four dense layers:  $FC(128, ReLU)$ ,  $BN$ ;  $FC(128, ReLU)$ ,  $BN$ ;  $FC(128, ReLU)$ ,  $BN$ ; and  $FC(128, ReLU)$ ,  $BN$ . The latent space was represented by a  $FC(8, ReLU)$ ,  $BN$  layer. The decoder  $D()$  symmetrically mirrored the structure with:  $FC(128, ReLU)$ ,  $BN$ ;  $FC(128, ReLU)$ ,  $BN$ ;  $FC(128, ReLU)$ ,  $BN$ ; and  $FC(128, ReLU)$ ,  $BN$ , ending with an output layer  $FC(d, None)$ . In this case, all activation functions used were ReLU (Rectified Linear Unit). Training was carried out with the same optimization settings, using Adam with a learning rate of  $10^{-3}$ , and early stopping based on validation loss. The loss function employed was the *Mean Squared Error* (MSE).

All architectural choices and hyperparameters—such as the number of layers, size of latent space, activation functions, learning rate, and loss functions—were defined empirically. These choices were guided by a grid search and iterative experimentation, aiming to optimize the reconstruction capability of the models, minimize overfitting, and ensure good generalization performance when exposed to anomalous data in the test set.

#### D. Model Evaluation

The model’s performance was evaluated based on the confusion matrix and the AUC, F1-score, precision, and recall metrics, as summarized in Table IV. For section 00, the classifier achieved an AUC of 0.94, F1-score of 0.80, precision of 0.73, and recall of 0.88, demonstrating a good balance between anomaly detection capability and false alarm rate. Figure 6 shows the corresponding confusion matrix for this model, in which it can be observed that, out of 445 samples, 126 anomalies were correctly detected, while only 17 were mistakenly classified as normal (false negatives). On the other hand, 256 normal signals were correctly recognized and 46 were incorrectly classified as anomalies (false positives).

In sections 02 and 04, the model maintained stable performance, with AUCs of 0.91 and 0.93, F1-scores of 0.80 for both, and recalls of 0.92 and 0.84, respectively. The variations in results can be attributed to operational and spectral differences among the analyzed machines. The overall average of

the metrics across the three sections was: AUC of 0.92, F1-score of 0.80, precision of 0.72, and recall of 0.88, highlighting the effectiveness and robustness of the proposed method for fault detection under varying operational conditions.

From a practical standpoint, false positives (normal samples incorrectly flagged as anomalies) may lead to unnecessary inspections, downtime, and increased operational costs. While such occurrences are undesirable, they are often acceptable in critical systems when weighed against the risks of undetected faults. Conversely, false negatives (anomalies wrongly classified as normal) pose a more serious risk in industrial environments, as they can result in unanticipated equipment failure, safety hazards, production losses, and higher maintenance costs. Therefore, in high-risk applications, optimizing recall is often prioritized to ensure the early identification of potential faults, even at the expense of precision. The results achieved by the proposed method reflect a favorable trade-off, offering high recall with acceptable levels of false positives, making it suitable for real-world deployment in industrial condition monitoring systems.

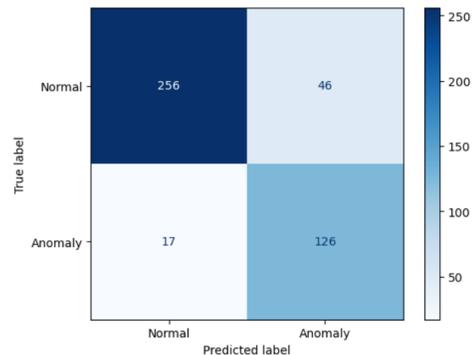


Fig. 6. Confusion matrix for section 00.

Table IV presents a summary of the performance metrics for the models trained on each analyzed section, including the corresponding average values. These results enable a direct comparison of the classifier’s effectiveness across different subsets of the dataset. Each section corresponds to a distinct machine model, and the performance variations observed can be attributed to the spectral and operational characteristics specific to the acoustic signals associated with each piece of equipment.

TABLE IV  
PERFORMANCE METRICS BY SECTION.

Type of machine	Model ID	AUC	F1-Score	Precision	Recall
Pump	00	0.94	0.80	0.73	0.88
	02	0.91	0.80	0.70	0.92
	04	0.93	0.80	0.75	0.84
Avg.		0.92	0.80	0.72	0.88

#### E. Comparative Analysis

In the original study that introduced the *MIMII Dataset*, autoencoders were employed using Mel-scaled spectrograms

as input descriptors. The results showed significant variation across sections: 84% accuracy for section 00, 45% for section 02, and 99% for section 04 [13].

The approach proposed in this work, based on Spectral Correlation Density (SCD) combined with dense autoencoders, demonstrated more stable and competitive performance. As shown in Table IV, all sections achieved AUCs above 0.90, with an average F1-score of 0.80 and an average recall of 0.88. Section 02 is particularly noteworthy, where the model outperformed the original study by more than 35 percentage points, indicating a higher ability of the proposed method to detect anomalies in noisy environments with spectral complexity.

This improvement can be attributed to the higher sensitivity of SCD in capturing cyclic components present in faulty signals, providing more discriminative descriptors. Additionally, the use of the *Random Forest* classifier to select the most relevant cyclic frequency indices contributed to dimensionality reduction and improved the robustness of the autoencoder, resulting in better model generalization across different sections and operational conditions.

Furthermore, the proposed architecture presents promising potential for real-time (online) deployment. Both the SCD feature extraction and the autoencoder inference can be optimized to run with low latency, and the dimensionality reduction step reduces computational overhead during execution. Preliminary tests showed that inference times remained within acceptable limits for online monitoring, and the overall model footprint is compatible with deployment in embedded systems equipped with mid-range processing units. This indicates that the methodology is not only accurate but also viable for practical industrial applications requiring continuous and autonomous condition monitoring.

## V. CONCLUSIONS

The approach based on Spectral Correlation Density (SCD) has proven to be a promising alternative for the analysis of industrial acoustic signals due to its ability to capture cyclic patterns characteristic of the cyclostationary behavior of machines in operation. By extracting spectral representations associated with different  $\alpha$  indices, it was possible to generate descriptors sensitive to variations in the operational states of the equipment. The feature selection step using the *Random Forest* algorithm allowed the identification of the most relevant subsets of cyclic frequencies for the task, reducing data dimensionality and optimizing model performance.

The autoencoders trained with these descriptors yielded consistent results, outperforming the baseline from the original *MIMII Dataset* study, especially in section 02, where a significant performance gain was observed. The obtained results indicate that the combination of SCD, feature selection, and unsupervised learning constitutes a robust strategy for fault detection in noisy industrial environments.

As future work, we intend to explore the use of cepstral coefficients extracted from the subset of cyclic frequencies selected by *Random Forest*, combining these new representations with unsupervised learning approaches. The goal is to

enhance the model's ability to detect subtle or rare anomalies, expanding the applicability of the proposed method to industrial scenarios where fault labeling is not feasible.

## VI. ACKNOWLEDGMENTS

The authors gratefully acknowledge the National Council for Scientific and Technological Development (CNPq) for the financial support of this work, provided through the project funded under process No. 445158/2024-7.

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

## REFERENCES

- [1] Zrar Khalid Abdul, Abdulbasit K. Al-Talabani, and Dlair O. Ramadan. A hybrid temporal feature for gear fault diagnosis using the long short term memory. *IEEE Sensors Journal*, 20(23):14444–14452, December 2020.
- [2] Xiaoyang Bi, Shuqian Cao, and Daming Zhang. A variety of engine faults detection based on optimized variational mode decomposition-robust independent component analysis and fuzzy *c*-mean clustering. *IEEE Access*, 7:27756–27768, 2019.
- [3] Leo Breiman. Random forests. *Machine Learning*, 45(1):5–32, October 2001.
- [4] Zhipeng Feng and Fulei Chu. Cyclostationary analysis for gearbox and bearing fault diagnosis. *Shock and Vibration*, 2015:1–12, 2015.
- [5] W. Gardner. Measurement of spectral correlation. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 34(5):1111–1123, October 1986.
- [6] Steven Hong, Eric Like, Zhiqiang Wu, and Cem Tekin. Multi-user signal classification via spectral correlation. In *2010 7th IEEE Consumer Communications and Networking Conference*, page 1–5. IEEE, January 2010.
- [7] Ihor Javorskyj, Ivan Matsko, Roman Yuzevovych, Oleh Lychak, and Roman Lys. Methods of hidden periodicity discovering for gearbox fault detection. *Sensors*, 21(18):6138, September 2021.
- [8] Keerthi Kethineni and Pradeepini Gera. Iot-based privacy-preserving anomaly detection model for smart agriculture. *Systems*, 11(6):304, June 2023.
- [9] Seong-Mok Kim and Yong Soo Kim. Enhancing sound-based anomaly detection using deep denoising autoencoder. *IEEE Access*, 12:84323–84332, 2024.
- [10] Anna Michalak, Justyna Hebda-Sobkowicz, Jacek Wodecki, Krzysztof Szabat, Marcin Wolkiewicz, Sebastian Weisse, Jerome Valire, Radosław Zimroz, and Agnieszka Wylomańska. Enhancement of cyclic spectral coherence map by statistical testing approach—application to bearing faults diagnosis in electric motors. *Measurement Science and Technology*, 36(1):016169, November 2024.
- [11] Fabian Perez-Sanjines, Cédric Peeters, Timothy Verstraeten, Jérôme Antoni, Ann Nowé, and Jan Helsen. Fleet-based early fault detection of wind turbine gearboxes using physics-informed deep learning based on cyclic spectral coherence. *Mechanical Systems and Signal Processing*, 185:109760, February 2023.
- [12] Rahul Gowtham Poola, Lahari Pl, and Siva Sankar Y. Covid-19 diagnosis: A comprehensive review of pre-trained deep learning models based on feature extraction algorithm. *Results in Engineering*, 18:101020, June 2023.
- [13] Harsh Purohit, Ryo Tanabe, Kenji Ichige, Takashi Endo, Yuki Nikaido, Kaori Suefusa, and Yohei Kawaguchi. MIMII dataset: Sound dataset for malfunctioning industrial machine investigation and inspection, 2019.
- [14] Harsh Purohit, Ryo Tanabe, Takeshi Ichige, Takashi Endo, Yuki Nikaido, Kaori Suefusa, and Yohei Kawaguchi. MIMII Dataset: Sound dataset for malfunctioning industrial machine investigation and inspection. In *Proceedings of the Detection and Classification of Acoustic Scenes and Events 2019 Workshop (DCASE2019)*, pages 209–213, November 2019.
- [15] Sanaz Roshanmanesh, Farzad Hayati, and Mayorkinos Papaalias. Utilisation of ensemble empirical mode decomposition in conjunction with cyclostationary technique for wind turbine gearbox fault detection. *Applied Sciences*, 10(9):3334, May 2020.

- [16] Y. Song, J. Liu, and Y. Wang. The cmm-pwda feature extraction method of subway fan under piston wind effect. *Revista Internacional de Métodos Numéricos para Cálculo y Diseño en Ingeniería*, 36, July 2020.
- [17] Xing Zhao, Yiming Lu, Baoxian Chang, and Liqun Chen. Study on the extraction method for track-side acoustic features based on cyclic stationary analysis. *Machines*, 11(10):957, October 2023.