2025 23rd European Control Conference (ECC)
June 24-27, 2025. Thessaloniki, Greece

Real World Comparative Analysis of Unsupervised Machine Learning Techniques for Anomaly Detection in Washing Machine Production

Federico Vesentini[1], Francesco Giuseppe Cordoni[2],
Gianluca Bacchiega[3], Robert Radu[3] and Riccardo Muradore[1]

Abstract— **In modern industrial production lines, ensuring product quality, customer satisfaction and minimizing production costs is of primary importance. Many learning techniques to address the issue of fault detection require training on labelled databases with a large number of anomalous audio samples that, however, are difficult or impossible to obtain. Furthermore, understanding which audio features are really crucial for anomaly detection is non-trivial. The article presents a comparative analysis of three unsupervised machine learning techniques based on the analysis of audio files/features, suitable to the case where a significant number of anomalies is not available; and a strategy for isolating audio features that are really important for anomaly detection. Experimental results show that the technique based on the isolation of the correct audio features is better than brute-force techniques.**

I. INTRODUCTION

The production processes that lead to the manufacturing of goods, in modern industry, are usually very complex and with many variables involved, that might cause significant changes into the final products' characteristics. To guarantee quality, customer' satisfaction and to reduce production costs, every step of the production process needs to be constantly monitored in such a way to detect possible anomalies as soon as possible. However, an efficient monitoring of the production process can be not so straightforward for many reasons, such as e.g., high instrumentation costs. The problem of fault detection in production lines can be tackled with several approaches, e.g. via in-depth analysis of (i) audio files as time-series data, or (ii) time-domain and frequency-domain features.

In a recent work [1], authors analyzed time, frequency and time-frequency audio features with several machine learning techniques such as k-Nearest Neighbor (kNN), Support Vector Machine (SVM), Kernel Liner Discriminant Analysis (KLDA) and Sparse Discriminant Analysis (SDA) for the fault detection among rotating bearings. Similarly, in [2] a SVM is trained on specific audio features for the fault diagnosis, via noise classification, of train plug doors. A survey on deep learning fault diagnosis methods in railways maintenance based on audio analysis is presented in [3]. Another work, based on audio features extracted via Empirical Mode Decomposition (EMD), is presented in [4]. A recent survey on fault diagnosis based on audio signal analysis can be found in [5]. Working with supervised learning techniques requires labeled training databases, such as e.g., AudioSet [6], which are often difficult or even impossible to obtain. Additionally, instances of anomalous products in production lines are usually very few compared to instances of healthy products, yielding to collecting unbalanced datasets. These limitations, along with the challenge of identifying which audio features are truly significative, can hinder the application of such methods in real-world scenarios. To overcome these limitations, many works [7], [8], [9] rely on unsupervised machine learning techniques such as Multi-Layer Perceptrons (MLPs), Long-Short Time Memory (LSTM) networks and pre-trained audio neural networks (PANNs). Furthermore, in [10] an evaluation of machine learning methodologies for the anomaly detection in cyber-security is presented.

The aim of this article is not only to propose a comparative analysis of three different unsupervised machine learning techniques, based on an Isolation Forest (IF) and a Sparse Auto-Encoder (SAE), for the identification of defective machines in a washing machine production line, but also to illustrate a possible strategy, based on k-Means (kM), for separating significant audio features from insignificant ones in this context. The databases used for training and validating the proposed methods were obtained by recording washing machines directly on the production line, while the anomalies used as ground-truth were isolated by expert personnel. Anomaly classification performance of each method is evaluated via confusion matrix.

The paper is organized as follows. In Section II we describe the k-Means audio feature selection procedure that we propose for the Isolation Forest, and the Sparse Auto-Encoder we use on raw audio files and spectrograms. In Section III we describe the audio recording setup, the datasets on which we tested the methodologies and discuss the obtained results. Finally, in Section IV we draw some conclusions.

II. ANOMALY DETECTION METHODOLOGIES

Malfunctions in washing machines can stem from issues like mechanical assembly errors or pre-existing defects in their components. Anomalies can manifest as unusual sounds and noise generated during testing, such as changes in pitch or ticking. Experts can detect anomalies through listening,

*This work was partly supported by the Fondo Europeo di Sviluppo Regionale; Regione Lombardia; POR FESR 2014-2020 azione 1.1.b.1.2 - bando "Tech Fast Lombardia", progetto SEQUAD.

[1]Federico Vesentini and Riccardo Muradore are with the Department of Engineering for Innovation Medicine of the University of Verona, Italy. federico.vesentini,riccardo.muradore@univr.it

[2]Francesco Giuseppe Cordoni is with the Department of Civil Environmental and Mechanical Engineering of the University of Trento, Italy. francesco.cordoni@unitn.it

[3]Gianluca Bacchiega and Robert Radu are with I.R.S. srl, Padova, Italy. bacchiega,radu@irsweb.it

but production environments tend to be noisy due to the presence of workers and machinery, making it possible for small defects to go unnoticed. Furthermore, quality control staff members have subjective hearing perceptions that necessitate extensive training.

The models that we consider in this comparative analysis are an IF trained on specific time-domain and frequency-domain features, extracted directly from every audio file of the training datasets by using a procedure based on kM (i), a SAE trained directly on the time-series of the audio files (ii) and a SAE trained on the spectrograms extracted from audio files contained in the training dataset (iii).

A. Audio Features Selection via k-Means

The features that are usually considered in audio analysis are *Amplitude Envelope* (AE), *Root-Mean-Square Energy* (RMS), *Zero-Crossing-Rate* (ZCR), *Fluctuation Strength* (FS), *Spectral Flux* (SF), *Spectral Centroid* (SC) and *Band Power* (BP), see e.g., [11], [12], [13], [14]. However, it is not certain that training the IF on all the features necessarily leads to the optimal value for the score-threshold S_{thr}: training the model by including all features could introduce noise, thereby potentially biased classification performance.

The procedure for isolating only the most significant audio features via kM is the following

(i) true anomalies[1] are inserted into a dataset of regular audio files[2];
(ii) audio features are tested one-by-one, by forcing kM to identify two data clusters and the misclassification percentage is calculated relative to the true anomalies;
(iii) once the feature that yields the best results is identified, all the other are added and tested one at a time, until the best classification results are achieved.

What emerged by applying this procedure is that time-domain features such as RMS energy and AE perform very well in classifying true anomalies, with only 0.10% misclassification, but with a large percentage of regular files classified as anomalous, 13.60%. Frequency-domain features like FS, SF and BP badly detect true anomalies (1.90% of misclassifications), but perform much better in detecting false anomalies (0.20% of misclassifications). ZCR and SC are practically irrelevant for anomaly detection, and therefore discarded.

The best combination of features is given by RMS energy and BP, allowing kM to cluster the true anomalies with a good grade of precision (0.25% of anomalies misclassified as regularities), but at the same time with a small number of false anomalies (0.20%). Fig. 1 shows the kM clusters related to this result: the small dense group of blue dots represents the cluster of regular cases (1908 samples), while the large sparse group of red dots is representing the anomalous cases (92 samples). The cluster centroids are represented by the yellow rhombuses. From these clusters it is possible to notice that anomalous audio files are characterized by a large value of both RMS energy and BP. These features are described in detail in Section II-B.

[1]Anomalies identified by the production line personnel
[2]Recorded from perfectly functioning washing machines

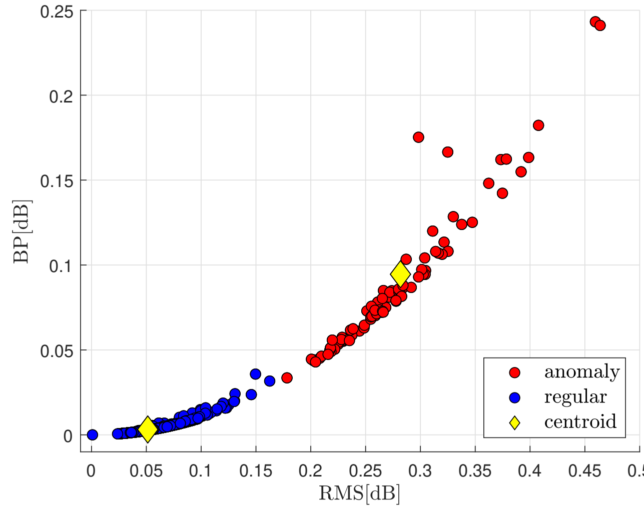

Fig. 1. RMS energy and BP clusters, calculated by kM.

B. Machine Learning via Isolation Forest

IF is trained on the time-domain and the frequency-domain features isolated via kM in Section II-A. Audio files are divided into equal overlapping subintervals called *frames*, denoted with F, and features are extracted on every frame and aggregated by calculating their mean among frames:

- *Root-Mean-Squared Energy* (RMS): RMS of frame F is a measure of its loudness, given by
$$RMS_F = \sqrt{\frac{1}{|F|}\sum_{i=0}^{n} s^2(k_i)},$$
where $|F|$ is the cardinality of the frame F and $s(k_i)$ is the k_i-th sample of the signal s.
- *Band Power* (BP): BP of frame F quantifies the energy distribution across different frequency components; it is calculated from the *Power Spectrum Density* (PSD) via Welch's method, see [15].

IF assigns a score ranging from 0 to 1 to every sample and calculates a score-threshold, namely S_{thr}, over which the samples are considered to be anomalous. The threshold is tuned to maximize the F_1-score given by

$$F_1 = \frac{2TR}{2TR + FA + FR}, \tag{1}$$

where TR is the number of true regular, FR the number of false regular and FA the number of false anomalous audio datas, detected for a fixed value of $S_{thr} \in [0, 1]$.

C. Sparse Auto-Encoder

SAE is designed to compress (encode) the input data and then reconstruct it (decode) so that the decoded output is as close as possible to the input data. We use the trained unsupervised neural network to reconstruct another dataset acquired from both healthy and malfunctioning washing machines. These anomalies have been isolated by expert personnel who works in the production line.

The neural network we trained for anomaly detection on audio time-series is a fully-connected SAE. In particular, the activation function g_{enc} for the encoder is the positive saturating linear function, given by

$$g_{enc}(z) = \begin{cases} 0, & \text{if } z \leq 0 \\ z, & \text{if } 0 \leq z \leq 1 \\ 1, & \text{if } z \geq 1 \end{cases}$$

while the activation function of the decoder g_{dec} is the linear function, i.e., $g_{dec}(z) = z$. The loss function used for the training phase is the *Sparse Mean Squared Error* given by

$$SMSE = \frac{1}{N}\sum_{k=1}^{K}\sum_{i=1}^{N}\left(X_i^{(k)} - \hat{X}_i^{(k)}\right)^2 + \lambda\Omega_w + \beta\Omega_s \quad (2)$$

where K is the number of time-series, λ is the coefficient of the L^2-regularization term Ω_w and β is the coefficient for the sparsity regularization term Ω_s.

The L^2-regularization term is defined by

$$\Omega_w = \frac{1}{2}\sum_{i=1}^{L}\sum_{j=1}^{n_l}\sum_{h=1}^{h_l}(w_{ijh})^2, \quad (3)$$

where L is the number of hidden layers, n_l is the output size of layer l, h_l is the input size of layer l and w_{ijh} are the weights for each layer.

There exist many types of functions that can be adopted as sparsity regularization term Ω_s, the one we have adopted is the *Kullback-Leibler divergence* function [16], that provides a measure of how different two distributions are.

Time-series belonging to anomalous audio files should exhibit a large reconstruction mean squared error R_{err}

$$R_{err} = MSE(X - \hat{X}) = \frac{1}{N}\sum_{i=1}^{N}\left(X_i - \hat{X}_i\right)^2, \quad (4)$$

where N is the number of samples of every time-series, X_i is the i-th sample of the original time-series X, and $\hat{X}_i$ is the i-th sample of the reconstructed series $\hat{X}$.

D. Sparse Auto-Encoder on Audio Spectrograms

A spectrogram is a way of representing an audio signal in the (time,frequency)-domain: it is obtained via Short-Time Fourier Transform (STFT) and displays the spectrum of frequencies of a signal at different time intervals. SAE trained on the spectrograms is of the same type described in Section II-C. The only difference is the activation function of both the encoder and the decoder, which is the logistic sigmoid for both, given by

$$g_{enc}(z) = g_{dec}(z) = \frac{1}{1 + e^{-z}} \quad (5)$$

where $z \in \mathbb{R}$.

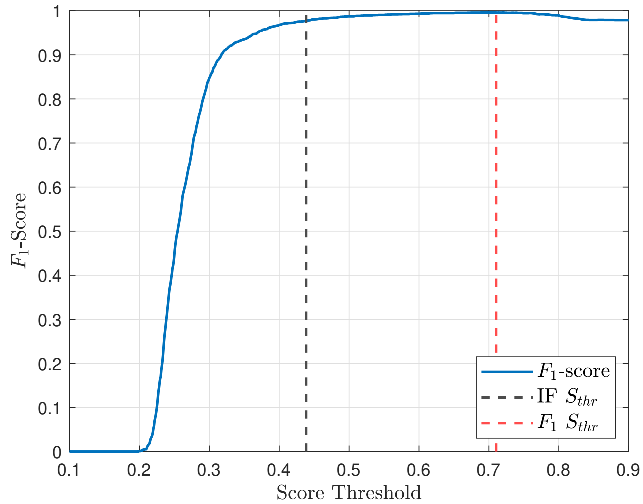

Fig. 2. F_1-score curve (blue continuous line) plotted against the score-threshold calculate by the IF (black dashed line) and the score-threshold maximizing the curve (red dashed line).

III. EXPERIMENTAL RESULTS

In this section we describe the results of our anomaly detection comparative analysis. We firstly describe the audio acquisition setup, then we point out some details about the training and test datasets, we describe the procedure to choose the best suited audio features for the IF training and finally we show, compare and comment the obtained results. IF and SAE training are done on a PC with an AMD Ryzen 7 4570U processor (8 cores, 16 threads) operating at 1.7 GHz and 16 GB DDR4 of RAM. The accuracy of the results are evaluated using a *confusion matrix* [17], whose coefficients depend on the value of the reconstruction error R_{err} and the score threshold S_{thr}, whose optimal positioning is achieved by maximizing the F_1-score (1).

Fig. 2 shows the F_1-score curve of IF: the black dashed line is the S_{thr} value set by IF, while the red dashed is the one maximizing the F_1-score.

A. Audio Recording Setup Description

The audio files are recorded with a single channel microphone positioned at $0.5\,\mathrm{m}$, perpendicular to the washing machine drum, see Fig. 3, for $1.5\,\mathrm{s}$ during the spinning test phase, when the machine drum reaches its steady-state velocity. The microphone is connected directly to the test PC audio sound card and acquisition is controlled with a software implemented in NI LabVIEW, using the Sound and Vibration Toolkit. Each test generates a 32-bit .wav file, sampled at $F_s = 22.050\,\mathrm{Hz}$.

Audio files may be influenced by many types of background noise, generated by all the equipment inside the manufacturing facility, e.g., disturbances due to pneumatic actuators installed on the test line.

B. Data Sets

The training dataset, D_{train}, is composed by $K = 2.000$ time-series extracted from audio files; every time-series con-

TABLE I
WASHING MACHINE TYPE PERCENTAGE WITHIN D_{test}.

WM Type	number of models	percentage
type 1	282	14.10%
type 2	174	8.70%
type 3	188	9.40%
type 4	284	14.20%
type 5	282	14.10%
type 6	188	9.40%
type 7	201	10.05%
type 8	189	9.45%
type 9	85	4.25%
type 10	127	6.35%

sists of $N = 33.073$ points, acquired with a frequency equal to F_s. Since there are many different types of washing machines, each with different characteristics, in order to avoid bad or unbalanced training, D_{train} contains 10 different types of washing machines for a total of 200 models per type. The testing dataset, D_{test}, contains other K samples the same types of washing machines, but in different proportions, see Table I, and with a total of $K_a = 83$ anomalies identified by the production line personnel.

In order to better understand the relevance of regular data misclassified as anomalies, we analyzed an additional testing dataset, D_{add}, of the same type and size as D_{test}, but consisting solely of regular audio files.

C. Isolation Forest Results

We extracted the audio features described in Section II-B by firstly subdividing all time-series into overlapping frames whose length is equal to

- $K_F = 128$ samples of which $K_{over} = 64$ (50%) are overlapping for RMS energy; and
- $K_F = 20.000$ samples of which $K_{over} = 10.000$ (50%) are overlapping for BP.

Framing in time-domain is tipically done with relatively small and half-overlapping windows, whose number of samples is proportional to the power of 2, see e.g. [11]. Framing in frequency-domain requires large windows in order to obtain a reasonable frequency representation of audio signals.

We point out that in every frequency-domain feature that requires to calculate the PSD, we considered the frequency interval [0, 2.000] Hz since most of the signal power is

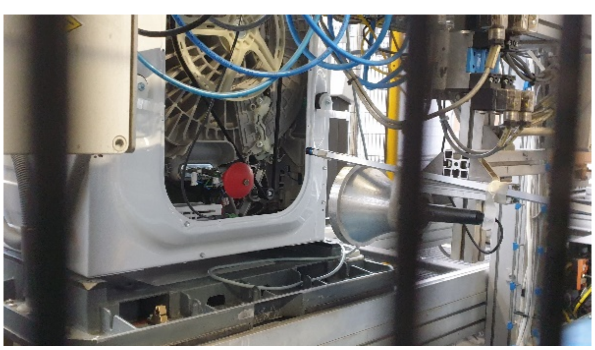

Fig. 3. A washing machine during the data acquisition phase.

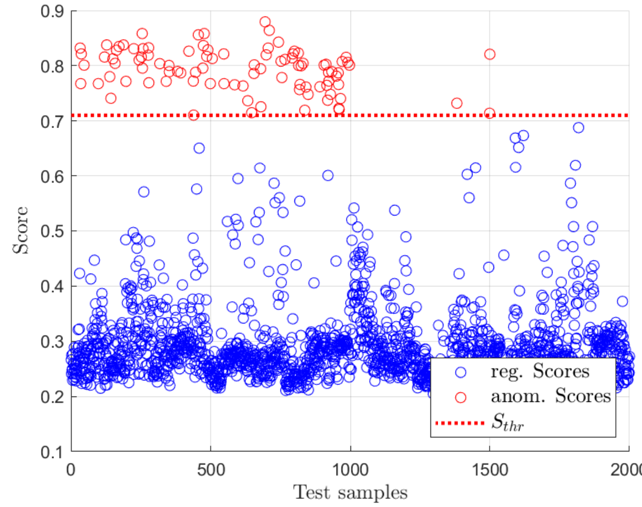

Fig. 4. Time-series of D_{test} clustered by the isolation forest. All the sample above S_{thr} (dashed red line) are anomalous (empty red circles) while those below are regular (empty blue circles).

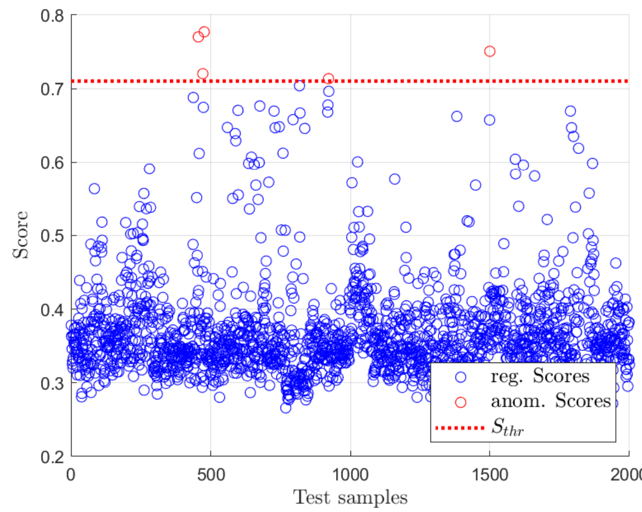

Fig. 5. Time-series of D_{add} clustered by the isolation forest. All the sample (empty red circles) above S_{thr} (dashed red line) are anomalous while those below (empty blue circles) are regular.

concentrated in such frequency band. This excludes the resonance peaks related to the metal cone into which the microphone was inserted, see Fig. 3.

Fig. 4 shows how IF clustered the time-series in D_{test}: all samples above the score threshold $S_{thr} = 0.7103$ (empty red circles) are considered anomalous, while those below (empty blue circle) are regular. IF found a total of 94 anomalies of which 81 over 83 are correctly classified, 2 are misclassified as regular and 13 regular time-series are misclassified as anomalous. In D_{add}, the model found only 5 anomalies over 2.000 time-series (0.25%). Fig. 5 shows the clustering with respect to a score threshold set to $S_{thr} = 0.7103$.

D. Sparse Auto-Encoder on Time-Series Results

The SAE has a hidden layer with 25 neurons and, for the SMSE (2), we have chosen an L^2-regularization coefficient

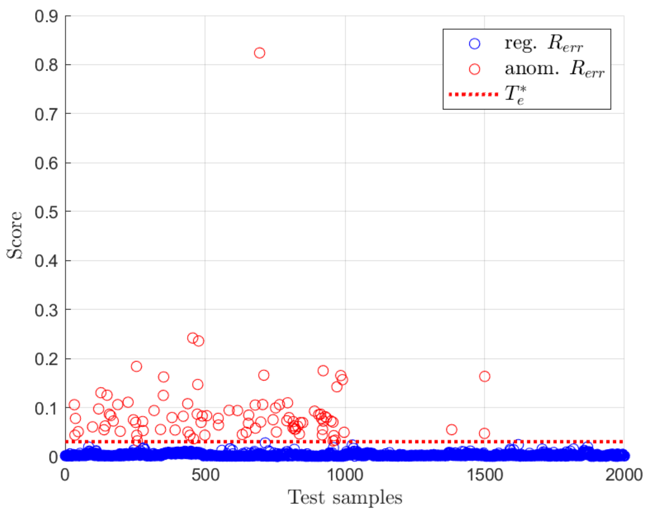

Fig. 6. R_{err} of the time-series stored in D_{test} (empty blue circles) plotted against the R_{err}-threshold (dashed red line). SAE neural network found 94 anomalies (empty red circles).

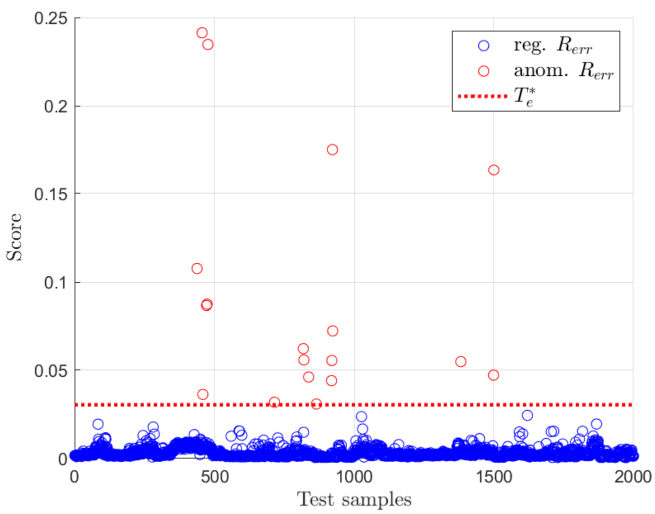

Fig. 7. R_{err} of D_{add} (empty blue circles) plotted against the R_{err}-threshold (dashed red line). SAE neural network found 18 anomalies (empty red circles).

equal to $\lambda = 0.01$ and a sparsity coefficients $\beta = 4$. Fig. 6 shows the reconstruction error, R_{err}, of every time-series in D_{test} plotted against the reconstruction error optimal threshold $T_e^* = 0.30 \cdot 10^{-1}$ defined via F_1-score (red dashed line). R_{err} of D_{train} is equal to $T_e = 3.5 \cdot 10^{-3}$. The blue circles are reconstruction errors that can be considered regular, while the red ones are those classified as anomalous, i.e. their R_{err} is above T_e^*. SAE found 94 anomalies of which 79 over 83 have been correctly identified, therefore 4 of them have been misclassified as regular and 15 regular samples have been classified as anomalous. The 83 anomalies, previously identified by the production line personnel, have been inserted into D_{test} within the first 1.000 time-series and indeed, as you can see in Fig. 6, between samples 0 and 1.000 there is a clear cluster of large-R_{err} samples.

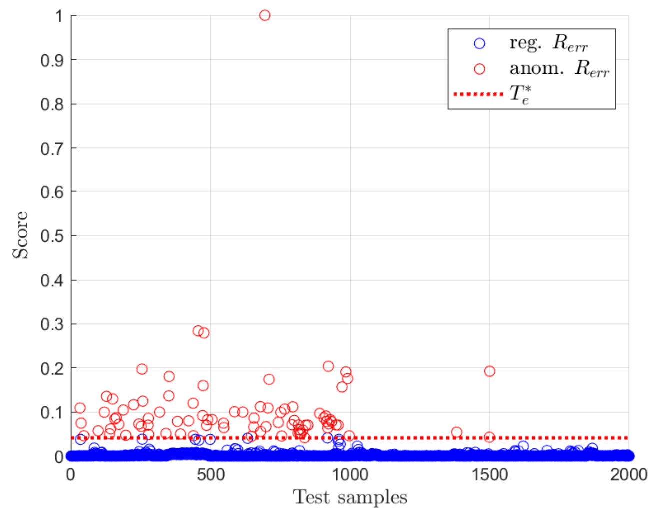

Fig. 8. R_{err} of the spectrgrams derived from D_{test} (empty blue circles), plotted against the T_e^* (dashed red line). SAE neural network found 85 anomalies (empty red circles).

Finally, by considering the same optimal threshold T_e^* of the previous case, the neural network found a 0.90% of time-series that can be considered as anomalous, i.e. 18 over 2.000 samples, in D_{add}, see Fig. 7.

E. Sparse Auto-Encoder on Spectrograms Results

We computed the spectrograms of audio files in D_{train} and D_{test} datasets by assuming frames of length

- $K_F = 1.024$ samples of which $K_{over} = 820$ (80%) are overlapping; and
- $K_{FFT} = 1.024$ samples for the Fast Fourier Transform.

Each spectrogram is a 513×157 matrix of complex coefficients. We trained the same type of SAE described in Section II-C on the absolute values of these coefficients, by assuming a hidden layer made of 10 neurons, a maximum number of epochs equal to 150 and, for the SMSE, a L^2-regularization coefficient equal to $\lambda = 0.001$ and a sparsity coefficient equal to $\beta = 0.01$.

The SAE reconstructs the training dataset with an average reconstruction error (4) equal to $T_e = 0.0136$. However, setting a threshold on this value does not lead to the best possible anomaly detection results. The optimal threshold $T_e^* = 0.0420$, above which every reconstruction error is considered to be anomalous, is obtained by maximizing the F_1-score (1), as discussed in Sections III-C and III-D.

Fig. 8 shows the distribution of reconstruction errors R_{err} of the spectrograms obtained from D_{test} (blue circles) plotted against the optimal reconstruction error T_e^* (red dashed line). The autoencoder has identified a total amount of 85 anomalies (blue circles) of which 72 out of 83 are correctly classified, therefore 11 of them are misclassified as regular ones and 15 regular spectrograms are misclassified as anomalous. We have used the SAE also to reconstruct D_{add} described in Section III-B: the model found 35 anomalies over 2.000 spectrograms (1.75%).

Isolation Forest

	Anomalous	Regular
Anomalous	99.90%	0.10%
Regular	0.65%	99.35%

SAE on Time-Series

	Anomalous	Regular
Anomalous	99.80%	0.20%
Regular	0.75%	99.25%

SAE on Spectrograms

	Anomalous	Regular
Anomalous	99.45%	0.55%
Regular	0.65%	99.35%

Fig. 9. Confusion matrices of the anomaly detection results.

F. Discussion of the Anomaly Detection Results

As it is possible to notice from the extra-diagonal elements (highlighted in blue) of the confusion matrices in Fig. 9, both IF and SAE on time-series detect the true anomalies of D_{test} very well. This is very important for production quality and customer's satisfaction, since the percentage of anomalous washing machines misclassified as regular ones are, in both cases, is very small: 0.10% for IF and 0.20% for SAE. Spectrograms perform slightly worse than the other two models, with a percentage of anomalous washing machines classified as regular equal to 0.55%, while false negative cases are consistent with other two models.

However, the percentage of regular washing machines misclassified as anomalous is higher than the percentage of anomalous washing machines misclassified as regular ones in every approach we have exploited, in particular for the SAE. This may be due to the fact that the recording phase of the audio files in D_{train} and D_{test} are affected by a high background noise of the production line, caused by the staff at work and by other machineries in operation.

This fact does not constitute a relevant problem, since it is definitely preferable to inspect again regular washing machines misclassified as anomalous before they leave the production line to be sold, then selling defective washing machines. It is worth pointing out that a relatively high percentage of regular machines incorrectly classified as possibly defective could affect production efficiency by clogging the production lines; therefore the adoption of the IF is preferable over the other two approaches.

IV. CONCLUSION

In this article, we proposed a comparative analysis of three methodologies based on two unsupervised learning models for fault detection in a washing machine production line, namely an Isolation Forest trained on properly selected features extracted from audio datasets (i) and a Sparse Auto-Encoder designed to reconstruct the audio time-series (ii) and the related spectrograms (iii). In particular, for the Isolation Forest, we proposed a strategy for selecting the most appropriate audio features for training based on k-means.

What emerged is that the Isolation Forest is the best model, having the lowest overall misclassification rates, although not too far from those of the Auto-Encoder trained on time series data. The Auto-Encoder trained on spectrograms shows a relatively high percentage of false positive cases, even in the reconstruction of the additional database containing only audio of perfectly functioning washing machines. The cause of this behavior could be due to a high sensitivity of the spectrograms to the background noise of the production/assembly line.

REFERENCES

[1] M. Altaf, M. Uzair, M. Naeem, A. Ahmad, S. Badshah, J. A. Shah, and A. Anjum, "Automatic and efficient fault detection in rotating machinery using sound signals," *Acoustics Australia*, vol. 47, pp. 125–139, 2019.

[2] Y. Sun, Y. Cao, and L. Ma, "A fault diagnosis method for train plug doors via sound signals," *IEEE Intelligent Transportation Systems Magazine*, vol. 13, no. 3, pp. 107–117, 2020.

[3] L. De Donato, F. Flammini, S. Marrone, C. Mazzariello, R. Nardone, C. Sansone, and V. Vittorini, "A survey on audio-video based defect detection through deep learning in railway maintenance," *IEEE Access*, vol. 10, pp. 65 376–65 400, 2022.

[4] T. S. Mohammed, M. Rasheed, M. Al-Ani, Q. Al-Shayea, and F. Alnaimi, "Fault diagnosis of rotating machine based on audio signal recognition system: an efficient approach," *Int. J. Simul. Syst. Sci. Technol*, vol. 21, pp. 8–1, 2020.

[5] L. Tang, H. Tian, H. Huang, S. Shi, and Q. Ji, "A survey of mechanical fault diagnosis based on audio signal analysis," *Measurement*, p. 113294, 2023.

[6] J. F. Gemmeke, D. P. Ellis, D. Freedman, A. Jansen, W. Lawrence, R. C. Moore, M. Plakal, and M. Ritter, "Audio set: An ontology and human-labeled dataset for audio events," in *2017 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, pp. 776–780. IEEE, 2017.

[7] H. Lee, P. Pham, Y. Largman, and A. Ng, "Unsupervised feature learning for audio classification using convolutional deep belief networks," *Advances in neural information processing systems*, vol. 22, 2009.

[8] A. A. T. Anvar and H. Mohammadi, "A novel application of deep transfer learning with audio pre-trained models in pump audio fault detection," *Computers in Industry*, vol. 147, p. 103872, 2023.

[9] M. Divya, C. Narayanappa, and S. Gangadharaiah, "Modeling a fault detection predictor in compressor using machine learning approach based on acoustic sensor data," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 9, 2021.

[10] N. Elmrabit, F. Zhou, F. Li, and H. Zhou, "Evaluation of machine learning algorithms for anomaly detection," in *2020 international conference on cyber security and protection of digital services (cyber security)*, pp. 1–8. IEEE, 2020.

[11] J. Breebaart and M. F. McKinney, "Features for audio classification," in *Algorithms in Ambient Intelligence*, pp. 113–129. Springer, 2004.

[12] C.-C. Lin, S.-H. Chen, T.-K. Truong, and Y. Chang, "Audio classification and categorization based on wavelets and support vector machine," *IEEE Transactions on Speech and Audio Processing*, vol. 13, no. 5, pp. 644–651, 2005.

[13] L. Bahatti, O. Bouattane, M. E. Echhibat, and H. Z. Mohamed, "An efficient audio classification approach based on support vector machines," *International journal of advanced computer science and applications*, vol. 7, no. 5, 2016.

[14] W. Shuiping, T. Zhenming, and L. Shiqiang, "Design and implementation of an audio classification system based on svm," *Procedia Engineering*, vol. 15, pp. 4031–4035, 2011.

[15] M. H. Gruber, *Statistical digital signal processing and modeling*. Taylor & Francis, 1997.

[16] B. A. Olshausen and D. J. Field, "Sparse coding with an overcomplete basis set: A strategy employed by v1," *Vision research*, vol. 37, no. 23, pp. 3311–3325, 1997.

[17] C. M. Bishop, *Pattern Recognition and Machine Learning*. Berlin, Heidelberg: Springer-Verlag, 2006.