IEEE Sensors Letters VOL. 10, NO. 1, JANUARY 2026

IEEE Sensors Council

Sensor applications

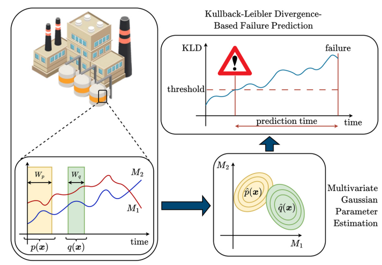
Failure Prediction in Manufacturing Processes Via Kullback–Leibler Divergence

Gianluca Tabella[1*], Mohammed Ayalew Belay[2**], Ismael Viejo[3], María Herrando[3], and Pierluigi Salvo Rossi[1,2***]

[1]Department of Gas Technology, SINTEF Energy Research, 7465 Trondheim, Norway
[2]Department of Electronic Systems, Norwegian University of Science, Technology, 7465 Trondheim, Norway
[3]Instituto Tecnológico de Aragón (ITA), 50018 Zaragoza, Spain
[*]Member, IEEE
[**]Graduate Student Member, IEEE
[***]Senior Member, IEEE

Manuscript received 24 November 2025; accepted 3 December 2025. Date of publication 5 December 2025; date of current version 22 December 2025.

Abstract—This work presents a novel algorithm for failure prediction in manufacturing processes using online unsupervised learning based on Kullback–Leibler divergence (KLD). The proposed method continuously monitors sensor data by comparing the probability distributions of a test window against those of a reference window to detect deviations that signal potential system degradation. These distributions are modeled as multivariate Gaussians to capture interdependencies between sensor signals. The algorithm is applied to real-world data from an electric arc furnace in the steel industry, demonstrating its ability to predict failures without prior offline training. Experimental results reveal that multivariate KLD analysis offers a more favorable balance between early fault detection and false alarm rates than univariate approaches. The method provides a lightweight, data-efficient, and practical solution for predictive maintenance in industrial settings where labeled failure data is limited or unavailable.



Index Terms—Sensor applications, anomaly detection, failure prediction, Kullback–Leibler divergence (KLD), manufacturing processes, predictive maintenance.

I. INTRODUCTION

Condition monitoring has become crucial within the Industry 4.0 framework, especially for manufacturing industries, to ensure reliable, safe, and effective operations. More specifically, prognostics and health management can provide timely and valuable information for effective predictive maintenance, which aims to avoid unexpected failures and reduce unexpected downtime [1].

Signal processing approaches are possible, e.g., estimating the remaining useful life of a gearbox via fusion of time-domain, frequency-domain, and envelope spectrum features was developed in [2]. However, the complexities and nonlinearities of realistic industrial scenarios make it difficult to predict system degradation accurately. Advanced Bayesian models based on enhanced particle filters have been proposed in [3]. Unsupervised anomaly detection is a more suitable paradigm for condition monitoring since typical scenarios of interest exhibit very few anomalous examples, and because monitoring systems should be effective in the presence of unseen anomalies. Support vector machines have been largely considered, e.g., for vibration analysis of jet engines [4]. Recent approaches rely on deep learning and advanced signal processing techniques, e.g., transformers and convolutional networks are combined within a multiobjective training framework in [5], while low-rank representation is exploited in [6] for mitigating the impact of noise and outliers in the training phase. Graph signal processing is exploited to model the temporal and spatial information and learn their nonlinear relation to degradation [7]. An extensive review and quantitative comparison of methods for unsupervised anomaly detection is found in [8]. Most approaches achieve good performance for sufficiently large training datasets. However, data are scarce in several practical cases due to their being expensive to collect or privacy issues that prevent sharing. In this context, anomaly detection methods have been developed based on the Kullback–Leibler divergence (KLD), to rely on small training datasets for modeling the normal behavior [9]. KLD for anomaly detection was also explored with applications in cognitive-radio systems [10].

Different from previous works, this letter contributes to exploiting a KLD-based approach for online failure prediction when a reliable training set is missing. Also, in addition to the false alarm rate, we consider the impact of different approaches on the average prediction time. Moreover, it presents a preliminary performance evaluation in a real-world manufacturing process scenario.

The rest of this letter is organized as follows. Section II provides information about the available dataset; Section III outlines the method; Section IV presents the numerical results with related discussion. Finally, Section V concludes this letter.

II. DATASET DESCRIPTION

The manufacturing process analyzed in this work involves electric arc furnaces (EAFs) in a steel production factory. The provided data include metric metering data, i.e., a time series of electricity

Corresponding author: Pierluigi Salvo Rossi (e-mail: salvorossi@ieee.org).
Associate Editor: Mahendra K. Shukla.
Digital Object Identifier 10.1109/LSENS.2025.3641051

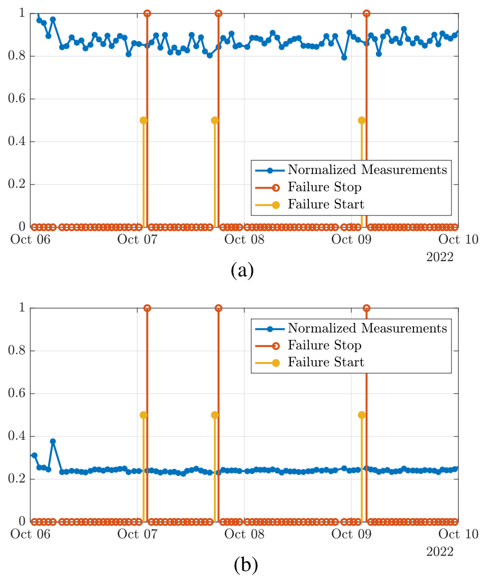


Fig. 1. Normalized measurements (blue), failure start (yellow), failure stop (red). (a) Energy. (b) Power.

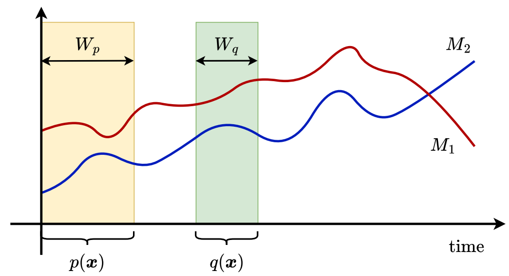


Fig. 2. Reference window (yellow) and test window (green) for PDF estimation.

consumption with a 1-min resolution collected from 30 August 2022, to 8 May 2024. It also includes the information collected in the manufacturing execution system (MES), where the main variables of each casting (i.e., temperatures, current, and power) are stored in batches. After data cleaning, some castings were neglected due to missing information or an incoherent signal for electricity consumption, reducing the total batches from 17 411 to 9047. The following criteria are established to consider the data valid: 1 MES includes the number of EAFs used; 2) MES includes the start time of the casting; 3) the metric metering data for the EAF consumption includes any electricity consumption for the batch's manufacturing period. Also, the measured data have been inspected with reference to the available labeled failure information. Failures with durations shorter than 15 min have been ignored. In this work, we discuss the results of processing data related to six variables: *energy consumption per casting* (*Energy*), *average power per casting* (*Power*), *casting duration* (*Power* ON), *steel liquid weight per casting* (*Steel*), *CH_4 consumption per casting* (*CH_4*), and *O_2 consumption per casting* (*O_2*). Without loss of generality, measurements have been individually normalized within the range [0, 1]. Fig. 1 shows an example of normalized measurements, where labeled information about failure starts and stops is also present. The overall available measurements were divided into segments for failure prediction analysis. Each segment starts from the stop of a failure to the start of the following failure. Measurements from each segment have been linearly interpolated with a time resolution of 30 min. Segments smaller than 24 h were excluded from the analysis, resulting in 132 segments suitable for the analysis.

III. KLD-BASED FAILURE PREDICTION

An information-theoretic approach for anomaly detection based on the KLD (requiring no training) has been explored for failure prediction. The KLD is a statistical distance measuring how much two distributions differ from each other [11]. The formal definition of the KLD of a continuous random variable $\boldsymbol{x} \in \mathbb{R}^k$ is the following:

$$D_{\mathrm{KL}}(p||q) = \int_{\mathbb{R}^k} p(\boldsymbol{x}) \ln\left(\frac{p(\boldsymbol{x})}{q(\boldsymbol{x})}\right) \mathrm{d}\boldsymbol{x} \tag{1}$$

where $p(\boldsymbol{x})$ and $q(\boldsymbol{x})$ are two k-dimensional probability density functions (PDFs). In the specific case of multivariate Gaussian distributions, i.e., $\boldsymbol{x} \overset{p}{\sim} \mathcal{N}(\boldsymbol{\mu}_p, \boldsymbol{\Sigma}_p)$ and $\boldsymbol{x} \overset{q}{\sim} \mathcal{N}(\boldsymbol{\mu}_q, \boldsymbol{\Sigma}_q)$, (1) translates into the following:

$$D_{\mathrm{KL}}(p||q) = \frac{1}{2}\Bigg(\mathrm{tr}(\Sigma_q^{-1}\Sigma_p) - k + (\boldsymbol{\mu}_q - \boldsymbol{\mu}_p)^{\mathrm{T}}\Sigma_q^{-1}(\boldsymbol{\mu}_q - \boldsymbol{\mu}_p) + \ln\left(\frac{\det(\Sigma_q)}{\det(\Sigma_p)}\right)\Bigg) \tag{2}$$

where $\boldsymbol{\mu}_p$ (respectively, $\boldsymbol{\mu}_q$) denotes the mean vector of $p(\boldsymbol{x})$ (respectively, $q(\boldsymbol{x})$) and $\boldsymbol{\Sigma}_p$ (respectively, $\boldsymbol{\Sigma}_q$) denotes the covariance matrix of $p(\boldsymbol{x})$ (respectively, $q(\boldsymbol{x})$).

The underlying assumption for exploring this approach is that (some) anomalous behavior leading to failures or other critical issues builds up over time because of some degradation process; thus, it could be detected early or predicted by monitoring and assessing behavioral deviations. We propose using the KLD between a distribution $p(\cdot)$ estimated from a reference window and a distribution $q(\cdot)$ estimated from a running test window as highlighted in Fig. 2 with reference to two generic measurement variables M_1 and M_2. The reference window comprises the initial W_p samples of the available measurement variables. The test window, instead, is considered to be of size W_q. To minimize the number of parameters to be estimated, the two distributions are modeled as multivariate Gaussians, with mean vectors and covariance matrices estimated via maximum likelihood[1]

$$\widehat{\boldsymbol{\mu}}_p = \frac{1}{W_p}\sum_{m=1}^{W_p} \boldsymbol{x}[m] \tag{3}$$

$$\widehat{\boldsymbol{\Sigma}}_p = \frac{1}{W_p}\sum_{m=1}^{W_p} \left(\boldsymbol{x}[m] - \widehat{\boldsymbol{\mu}}_p\right)\left(\boldsymbol{x}[m] - \widehat{\boldsymbol{\mu}}_p\right)^{\mathrm{T}} \tag{4}$$

$$\widehat{\boldsymbol{\mu}}_q[n] = \frac{1}{W_q}\sum_{m=n-W_q+1}^{n} \boldsymbol{x}[m] \tag{5}$$

$$\widehat{\boldsymbol{\Sigma}}_q[n] = \frac{1}{W_q}\sum_{m=n-W_q+1}^{n} \left(\boldsymbol{x}[m] - \widehat{\boldsymbol{\mu}}_q[n]\right)\left(\boldsymbol{x}[m] - \widehat{\boldsymbol{\mu}}_q[n]\right)^{\mathrm{T}} \tag{6}$$

where the moving nature of the test window is made explicit via the time index n. Note that the hat symbol over a parameter denotes an estimate. By computing the variations of the KLD with respect to the

[1] To ensure the invertibility of $\widehat{\boldsymbol{\Sigma}}_p$ and $\widehat{\boldsymbol{\Sigma}}_q[n]$, particularly in the presence of highly correlated signals or limited sample sizes, we recommend the following regularization: (i) convert $\widehat{\boldsymbol{\Sigma}}$ into its corresponding correlation matrix via $\widehat{\boldsymbol{R}} = \mathrm{diag}(\widehat{\boldsymbol{\Sigma}})^{-\frac{1}{2}}\, \widehat{\boldsymbol{\Sigma}}\, \mathrm{diag}(\widehat{\boldsymbol{\Sigma}})^{-\frac{1}{2}}$; (ii) upper bound the absolute value of the nondiagonal elements of $\widehat{\boldsymbol{R}}$; (iii) reconstruct the covariance matrix via $\widehat{\boldsymbol{\Sigma}} = \mathrm{diag}(\widehat{\boldsymbol{\Sigma}})^{\frac{1}{2}}\, \widehat{\boldsymbol{R}}\, \mathrm{diag}(\widehat{\boldsymbol{\Sigma}})^{\frac{1}{2}}$; (iv) apply a lower bound to the diagonal elements of $\widehat{\boldsymbol{\Sigma}}$.

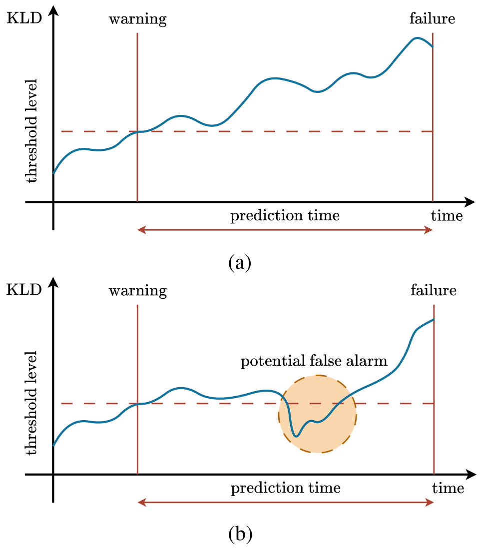


Fig. 3. Performance metrics. (a) Prediction time. (b) Potential false alarm.

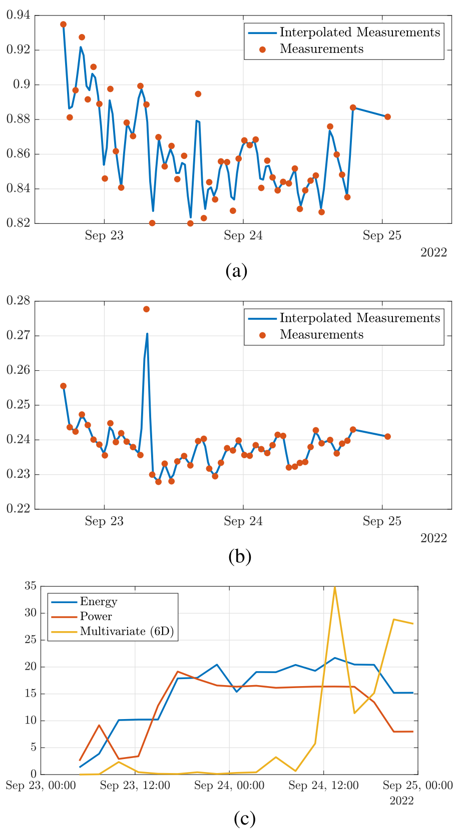


Fig. 4. KLD analysis of a segment leading to a predictable failure. (a) Energy. (b) Power. (c) KLD.

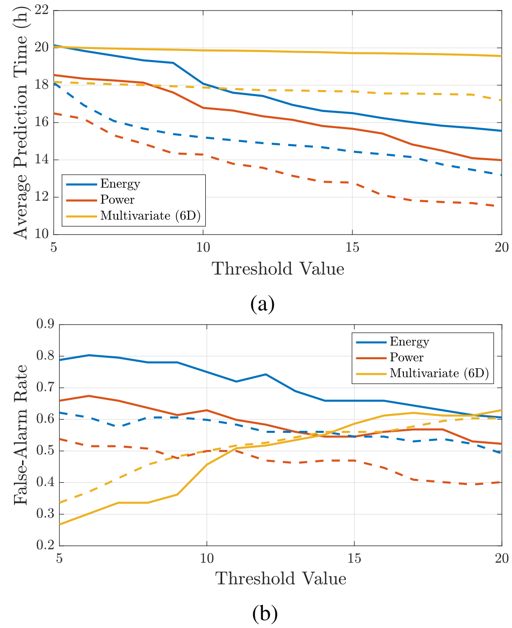


Fig. 5. Performance for 1-D and 6-D KLD for different window size: $W_p = W_q = 20$ (solid lines) and $W_p = W_q = 30$ (dashed lines). (a) Average prediction time. (b) Potential false alarm rate.

Algorithm 1: Failure Prediction Algorithm

```
Input: features' values; predefined threshold
while system is operational do
    select first W_p samples of each feature
    compute reference mean vector via (3)
    compute reference covariance matrix via (4)
    set counter as inactive
    while no failure detected do
        select most recent W_q samples of each feature
        compute test mean vector via (5)
        compute test covariance matrix via (6)
        compute KLD between test and reference via (2)
        if KLD value ≥ threshold & counter is inactive then
            raise an alarm
            start counter (active)
        end if
        if KLD value < threshold & counter is active then
            declare false alarm
            set counter as inactive
        end if
    end while
    if counter is active then
        declare correct prediction with prediction time by
        counter
    end if
end while
```

time index n, we have a mechanism for monitoring relevant statistical variations of the measurements with respect to the initial conditions. An increasing trend with time of the KLD denotes a consistent deviation from the initial operating conditions, likely leading to anomalies and potential system failure. Then, a threshold-based rule (assessing if the KLD rises above a fixed level) is suitable for early anomaly detection or as a failure prediction mechanism. Specifically, a given threshold level is considered for analyzing the potential presence of a failure, and performance is assessed with respect to *prediction time* and *potential false alarm rate*. The first moment when KLD values exceed the threshold level is considered a possibility to declare a warning. Thus, the time difference between the actual failure start (end of the segment) and this warning instant represents the prediction time (the larger, the better). In addition, if KLD values happen to be smaller than the threshold level after a warning has been identified, then the segment is counted as a potential false alarm; otherwise, if KLD values stay above the threshold level until the end of the segment, no false alarm is assumed. Fig. 3 presents a graphical description of these metrics while Algorithm 1 provides the corresponding pseudocode for their computation.

IV. RESULTS

Segments have been processed using the KLD-based approach previously described. The window sizes used are $W_p = W_q = 20$ samples (unless otherwise specified), and the test window is shifted with steps of five samples. Fig. 4 shows an example of the KLD analysis for a single segment (extracted from the measurements in September 2022). The red dots in Fig. 4(a) and (b) represent the measurements, while the blue solid lines represent the interpolated measurements. Fig. 4(c) shows the KLD values computed from the 1-D analysis (i.e., when interpolated measurements for *Energy* and *Power* are individually processed) and the KLD values computed from the 6-D analysis (i.e., when multivariate distributions are considered by pairing interpolated measurements). In all cases, an increasing trend in the KLD values is experienced when approaching the failure, even though some oscillations are present.[2] This specific segment also shows that 1-D and 6-D approaches exhibit different characteristics: both the 1-D KLD curves start the increasing trend early, while the 6-D KLD curve has a more pronounced peak value. As clarified via the following analysis, these aspects translate into different performances between the two approaches. A statistical analysis based on the available segments has been conducted to assess the potential performance of the proposed approach. Fig. 5 shows the performance in terms of average prediction time and potential false alarm rate for different levels of the threshold of three monitoring approaches based on: 1) using only the variable *Energy*, 2) using only the variable *Power*, and 3) using all six variables introduced in Section II. It is interesting to notice that, when a univariate approach is employed, the KLD approach allows a tradeoff between prediction time and false alarm rate by changing the threshold level: small threshold levels provide large prediction time (desirable) and large false alarm rates (undesirable), while large threshold levels reduce the prediction time (undesirable) and reduce the false alarm rate (desirable). This is not the case for the multivariate approach, which shows an inverse trend since it experiences an increasing false alarm rate, with a simultaneous slight decrease (nearly constant behavior) in average prediction time. This counterintuitive behavior is possible mainly due to our definition of *potential false alarm*, where multiple crossings of the threshold (due to a volatile behavior of the KLD) are counted as multiple potential false alarms [Fig. 4(c) clearly shows this behavior of the multivariate case]. However, it is apparent that moving from a 1-D approach to a 6-D approach significantly improves the performance since it allows working at a low potential false alarm rate and high average prediction times. Using only the variables *Energy* or *Power*, with $W = 20$, the average detection time spans 14 and 20 h, while the 6-D case is 20 h, approximately. Conversely, the false alarm rate in the 1-D cases exhibits unacceptable values in the [0.5, 0.8] range, while the 6-D approach provides false alarm rates in the [0.2, 0.7] range. Fig. 5 shows the impact of W on the performance, with average detection time and false alarm rate decreasing, showing a clear tradeoff between the two metrics, where a bigger value of W reduces the potential false alarm rate at the expense of a lower average prediction time. Finally, it must be noted that the complexity of the KLD-based algorithm is dominated by the matrix inversion in (2), which is $\mathcal{O}(k^3)$.

V. CONCLUSION

A KLD-based failure-prediction approach was proposed and tested on real-world data from the steel production industry. The algorithm does not require training. Performance assessment was presented and compared in the cases of univariate versus multivariate processing, with the latter approach exhibiting significantly lower false alarm rate while also reducing the average prediction time. The algorithm is suitable for other industrial applications.

ACKNOWLEDGMENT

This work was supported by the European Union under the Horizon Europe research and innovation programme through the project FLEX4FACT (https://www.flex4fact.eu flex4fact.eu) under Grant 101058657.

REFERENCES

[1] A. Sinha and D. Das, "XAI-LCS: Explainable AI-based fault diagnosis of low-cost sensors," *IEEE Sens. Lett.*, vol. 7, no. 12, Dec. 2023, Art. no. 6009304.

[2] Y. Han, M. Xu, X. Sun, X. Ding, X. Chen, and F. Gu, "Gear health monitoring and RUL prediction based on MSB analysis," *IEEE Sens. J.*, vol. 22, no. 5, pp. 4400–4409, Mar. 2022.

[3] C. K. Pang, J.-H. Zhou, and H.-C. Yan, "PDF and breakdown time prediction for unobservable wear using enhanced particle filters in precognitive maintenance," *IEEE Trans. Instrum. Meas.*, vol. 64, no. 3, pp. 649–659, Mar. 2015.

[4] P. Hayton, B. Schölkopf, L. Tarassenko, and P. Anuzis, "Support vector novelty detection applied to jet engine vibration spectra," in *Proc. Adv. Neural Inf. Process. Syst.*, Denver, CO, USA, 2000, pp. 907–913.

[5] M. A. Belay, A. Rasheed, and P. Salvo Rossi, "MTAD: Multiobjective transformer network for unsupervised multisensor anomaly detection," *IEEE Sens. J.*, vol. 24, no. 12, pp. 20254–20265, Jun. 2024.

[6] M. A. Belay, A. Rasheed, and P. Salvo Rossi, "Multivariate time series anomaly detection via low-rank and sparse decomposition," *IEEE Sens. J.*, vol. 24, no. 21, pp. 34942–34952, Nov. 2024.

[7] J. Qi, Z. Chen, Y. Song, J. Xia, and W. Li, "Remaining useful life prediction combining advanced anomaly detection and graph isomorphic network," *IEEE Sens. J.*, vol. 24, no. 22, pp. 38365–38376, Nov. 2024.

[8] M. A. Belay, S. S. Blakseth, A. Rasheed, and P. Salvo Rossi, "Unsupervised anomaly detection for IoT-based multivariate time series: Existing solutions, performance analysis and future directions," *Sensors*, vol. 23, no. 5, pp. 1–24, 2023.

[9] M. Filippone and G. Sanguinetti, "Information theoretic novelty detection," *Pattern Recognit.*, vol. 43, no. 3, pp. 805–814, 2010.

[10] A. Anderson and H. Haas, "Kullback-leibler divergence (KLD) based anomaly detection and monotonic sequence analysis," in *Proc. IEEE 74th Veh. Technol. Conf. (VTC Fall 2011)*, San Francisco, CA, USA, 2011, pp. 1–5.

[11] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York NY, USA: Wiley, 1991.

[2]Similar analyses have been conducted with different combinations of features without relevant additional insights.