Proc. of the European Safety and Reliability Conference (ESREL 2026)
Edited by José C. Matos, Paulo B. Lourenço, Michael Beer, Daniel Oliveira, Edoardo Patelli, Hélder S. Sousa, Rui A. Silva and Monica Santamaria
 Published by Research Publishing, Singapore.
doi: 10.3850/ESREL2026061419_esrel26-p26195-cd

The Role of Reliability Knowledge in Unsupervised Fault Detection of Wind Turbines

Renan Lopes Ricardo

Department of Mechatronics and Mechanical Systems Engineering, University of São Paulo, Av. Prof. Mello Moraes, 2231, São Paulo, SP, Brazil. E-mail: renan.lopesrica@usp.br

Welker Facchini Nogueira

Department of Mechatronics and Mechanical Systems Engineering, University of São Paulo, Av. Prof. Mello Moraes, 2231, São Paulo, SP, Brazil. E-mail: welkerfacchini@usp.br

Miguel Angelo de Carvalho Michalski

Department of Mechatronics and Mechanical Systems Engineering, University of São Paulo, Av. Prof. Mello Moraes, 2231, São Paulo, SP, Brazil. E-mail: michalski@usp.br

Gilberto Francisco Martha de Souza

Department of Mechatronics and Mechanical Systems Engineering, University of São Paulo, Av. Prof. Mello Moraes, 2231, São Paulo, SP, Brazil. E-mail: gfmsouza@usp.br

Unsupervised learning techniques have become essential tools for data-driven fault detection in complex industrial systems. However, their effectiveness depends strongly on how the input space represents the underlying physical and causal relationships of degradation. This study investigates the role of incorporating reliability knowledge, derived from analyses such as Failure Mode and Effects Analysis, Failure Mode and Symptoms Analysis, or Failure Mode and Observability Analysis, in shaping the data representation underlying the effectiveness and interpretability of anomaly detection models. The analysis focuses on the failure-mode level, comparing generalist models, trained with all available monitoring variables, to knowledge-guided specialist models, trained only with variables causally associated with specific failure mechanisms. Two unsupervised algorithms, including One-Class Support Vector Machine and Isolation Forest, are employed as neutral demonstration vectors, so that the observed effects can be attributed primarily to the structure of the feature space rather than to algorithmic characteristics. Performance is assessed using open SCADA data from wind turbines through metrics such as false alarm rate, early warning time, and alarm persistence. Additionally, geometric distances between the subspaces learned by generalist and specialist models are analysed to quantify structural divergence in feature representation. The study aims to assess whether knowledge-guided specialization can facilitate the emergence of fault-related deviations in unsupervised settings and support more interpretable model behaviour. By examining the consistency of these effects across different learning algorithms, the study provides quantitative insights into how reliability knowledge shapes feature spaces toward physically meaningful dimensions of system behaviour.

Keywords: Unsupervised fault detection, reliability knowledge, feature selection, failure modes, nominal space representation, anomaly detection, SCADA data, wind turbines.

1. Introduction

Unsupervised machine learning techniques have become central to data-driven fault detection in complex industrial systems, particularly in applications where labelled fault data are scarce or unavailable. In this context, models are trained to characterize nominal system behaviour and identify deviations that may indicate incipient faults. Despite their widespread adoption, the performance of these approaches remains highly variable across applications, and their limitations are often attributed to algorithmic sensitivity, lack of robustness, or data-related issues (Sikorska et al., 2011; Zio, 2022; Yan et al., 2025).

However, these limitations cannot be fully explained at the algorithmic level alone (Souza et al., 2021; Zio, 2022). The way system behaviour is represented in the input space plays a decisive role in determining whether degradation phenomena can emerge in a meaningful and interpretable manner. When monitored variables are weakly connected to the physical mechanisms underlying failure, fault-related deviations (symptoms) may remain geometrically embedded within the nominal (health) space, regardless of the learning technique employed.

From a reliability engineering perspective, system degradation is inherently structured around failure modes, each associated with specific physical mechanisms and observable symptoms. Reliability knowledge, typically formalized through analyses such as Failure Mode and Effects Analysis (FMEA), Failure Mode and Symptoms Analysis (FMSA), or Failure Mode and Observability Analysis (FMOA), provides a systematic framework to relate failure mechanisms to measurable variables that act as indicators of abnormal behaviour (Souza et al., 2021; Silva et al., 2022; Nogueira et al., 2025). While this knowledge is routinely used to support diagnosis and maintenance decision-making, its role in structuring data representations for unsupervised fault detection remains comparatively underexplored.

Most existing efforts to incorporate domain knowledge into data-driven PHM focus on modifying learning algorithms, for instance, through hybrid models, constrained optimization, or *post-hoc* interpretability techniques (von Hahn and Mechefske, 2022; Yan et al., 2025). These approaches implicitly assume that fault-relevant information is already encoded in the selected variables. In contrast, the more fundamental question of whether degradation can geometrically manifest in the nominal space under a given choice of descriptors remains largely unaddressed.

The present study addresses this gap by investigating the role of reliability knowledge at the level of data representation rather than algorithm design. Using open SCADA data from wind turbines provided by the CARE to Compare benchmark dataset (Gück et al., 2024), generalist (GEN) representations built from all available monitoring variables are contrasted with specialist (ESP) representations defined by variables causally associated with a specific failure mode. Nominal behaviour is learned exclusively from healthy data using Principal Component Analysis (PCA) as a geometric decomposition tool, and fault-induced effects are analysed through subspace similarity measures, geometric deviation metrics, and unsupervised anomaly detection results obtained with Isolation Forest (IF) and One-Class Support Vector Machine (OCSVM) models.

By analysing representation-driven effects across algorithms, this study shows that reliability knowledge structurally shapes data representations and influences unsupervised fault detection effectiveness.

This paper is organized as follows. Section 2 presents the methodology, Section 3 reports and discusses the obtained results, and Section 4 concludes the paper.

2. Materials and Methods

This section describes the materials and methods used to investigate how reliability-knowledge-guided descriptor selection influences the geometry of nominal representations and the manifestation of fault-related deviations in unsupervised fault detection. The overall methodological workflow is summarized in Fig. 1 and detailed in the following subsections.

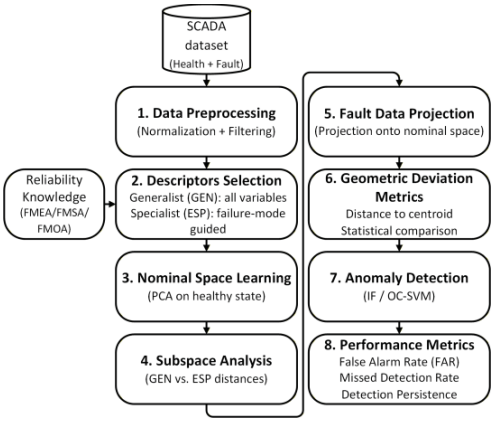

Fig. 1.Methodological workflow

2.1.*Data Description and Preprocessing*

The study uses wind turbine SCADA data from the open CARE to Compare benchmark datasets, where each dataset corresponds to a single turbine and comprises a historical *train* segment followed

by a *prediction* segment, with measurements recorded at 10-minute resolution (Gück et al., 2024). In addition to SCADA variables, the datasets include timestamps, asset identifiers, a train/prediction flag, and an operational status identifier.

The current analysis focuses on turbine 14 and event 34, corresponding to a failure characterized by high temperature in the transformer cell. The operational status identifier (*status_type_id*) encodes turbine operating regimes, ranging from normal operation to downtime and service states. Although both normal operation (*status_type_id* = 0) and idling (*status_type_id* = 2) are considered healthy conditions in the benchmark, only records corresponding to normal operation within the *train* segment are used to define a conservative nominal baseline.

SCADA variables are provided as summary statistics over each acquisition interval. In this study, only average values are retained, as they sufficiently characterize turbine operation while avoiding redundancy and unnecessary dimensionality. Preprocessing includes the removal of constant variables and filtering to an operational region defined by wind speed (5–30 m/s) and active power (0.1–1.5 kW), following benchmark recommendations (Gück et al., 2024) to ensure operational consistency and mitigate artefacts related to missing or zero-encoded values.

The resulting nominal dataset is randomly split into training and test partitions using a 60/40 ratio. All variables are standardized using z-score normalization, with parameters estimated from the nominal training data and applied consistently to the test set as well as to the full *train* and *prediction* segments, ensuring comparable scaling and preventing data leakage.

2.2.*Reliability-Guided Descriptors Selection*

A descriptor is defined as a monitored variable whose behaviour is causally related to the physical mechanisms underlying a given failure mode (Souza et al., 2021). Accordingly, descriptor selection is treated here as a reliability-guided process grounded in engineering knowledge of system degradation.

Relevant reliability knowledge can be formalized through structured analyses such as FMEA, FMSA, and FMOA (Souza et al., 2021; Silva et al., 2022; Nogueira et al., 2025). These analyses establish explicit relationships between failure mechanisms, observable symptoms, and the variables capable of capturing fault-related manifestations. Based on this mapping, descriptors are selected according to their relevance to the targeted failure mode, which in this study corresponds to a high-temperature event in the transformer cell.

Two multivariate representations are constructed from the same dataset. The generalist representation (GEN) includes all available input features retained after preprocessing, providing a broad description of the turbine operating state across electrical, mechanical, and operational domains. In contrast, the specialist representation (ESP) is defined by a reduced subset of descriptors selected according to reliability knowledge. For the considered transformer-related failure mode, ESP includes only variables directly associated with thermal, electrical, and load-related phenomena, namely rotor speed, apparent power, current, transformer core temperature, and transformer cell temperature.

By construction, the ESP representation defines a subspace of the GEN representation, emphasizing variables expected to exhibit coherent changes under the targeted degradation process. The sets of variables defining the GEN and ESP representations are fixed *a priori*, based solely on SCADA descriptions and reliability analyses, so that any subsequent differences can be attributed exclusively to reliability-guided descriptor selection.

2.3.*Nominal Space Learning and Subspace Analysis*

Nominal system behaviour is learned exclusively from healthy data using Principal Component Analysis (PCA) (Jolliffe, 2002), employed here as a geometric decomposition tool. The procedure is applied independently to the GEN and ESP representations, enabling a direct comparison between the nominal subspaces induced by each descriptor set.

Let $\mathbf{X}_H \in \mathbb{R}^{n_H \times p}$ and $\mathbf{Z}_H \in \mathbb{R}^{n_H \times q}$ denote the standardized healthy training datasets for the GEN and ESP representations, respectively, where n_H is the number of healthy samples and p and q are the numbers of descriptors. PCA is computed via Singular Value Decomposition (Golub and Reinsch, 1970), such that $\mathbf{X}_H =$

$\mathbf{U}_X\mathbf{\Sigma}_X\mathbf{V}_X^{\mathrm{T}}$ and $\mathbf{Z}_H = \mathbf{U}_Z\mathbf{\Sigma}_Z\mathbf{V}_Z^{\mathrm{T}}$, where $\mathbf{V}$ contain the principal directions in the space of variables, and $\mathbf{U}$ spans the corresponding subspaces in the observation space.

The nominal subspace dimensionality is defined as the number of retained principal components (PCs), k, based on the cumulative explained variance criterion. For each case, the smallest k satisfying Eq. (1) is selected:

$$\frac{\sum_{i=1}^{k}\lambda_i}{\sum_{i=1}^{m}\lambda_i} \geq \eta \tag{1}$$

where λ_i denotes the eigenvalues of the covariance matrix sorted in descending order, m is the total number of variables in the representation ($m = p$ for GEN or $m = q$ for ESP), and $\eta \in \{0.80, 0.90, 0.95\}$ is the dimensionless cumulative explained-variance threshold. For each η, the effective dimensionality used in the comparison is defined as the minimum number of retained components between the two representations.

Projections of healthy data onto the nominal subspaces are given by the score matrices $\mathbf{Y}_H = \mathbf{X}_H\mathbf{V}_{X,k}$ and $\mathbf{S}_H = \mathbf{Z}_H\mathbf{V}_{Z,k}$. Structural differences between the GEN and ESP nominal subspaces are quantified using the truncated orthonormal bases $\mathbf{U}_{X,k}$ and $\mathbf{U}_{Z,k}$. The principal angles θ_i between the subspaces are computed from the singular values of $\mathbf{U}_{X,k}^{\mathrm{T}}\mathbf{U}_{Z,k}$, and the corresponding Chordal distance is defined by Eq. (2).

$$d_{\mathrm{chord}} = \left(\sum_{i=1}^{k}\sin^2\theta_i\right)^{1/2} \tag{2}$$

Additionally, the Frobenius distance between the associated projection operators $\mathbf{P}_X = \mathbf{U}_{X,k}\mathbf{U}_{X,k}^{\mathrm{T}}$ and $\mathbf{P}_Z = \mathbf{U}_{Z,k}\mathbf{U}_{Z,k}^{\mathrm{T}}$ is defined by Eq. (3), where $\|\cdot\|_{\mathrm{F}}$ denotes the Frobenius norm.

$$d_{\mathrm{F}} = \| \mathbf{P}_X - \mathbf{P}_Z \|_{\mathrm{F}} \tag{3}$$

2.4. *Fault Data Projection and Geometric Deviation Metrics*

After learning the nominal subspaces from healthy data, observations from the prediction segment containing the fault event are projected onto the previously defined GEN and ESP nominal spaces. Let $\mathbf{X}_P \in \mathbb{R}^{n_P\times p}$ and $\mathbf{Z}_P \in \mathbb{R}^{n_P\times q}$ denote the standardized *prediction* datasets for the GEN and ESP representations, respectively, where n_P is the number of samples in the *prediction* period.

Using the loading matrices obtained from the PCA of healthy data (Section 2.3), prediction samples are projected onto the nominal subspaces as $\mathbf{Y}_P = \mathbf{X}_P\mathbf{V}_{X,k}$ and $\mathbf{S}_P = \mathbf{Z}_P\mathbf{V}_{Z,k}$ where $\mathbf{V}_{X,k}$ and $\mathbf{V}_{Z,k}$ define the truncated bases of the GEN and ESP nominal subspaces.

For each representation, the nominal centroid is defined as the mean of the healthy training scores, denoted by $\bar{\mathbf{y}}_X$ and $\bar{\mathbf{s}}_Z$. Geometric deviation from nominal behaviour is quantified as the Euclidean distance between each projected observation and the corresponding centroid. For the i-th observation, the distances are given by $d_X^{(i)} = \| \mathbf{Y}^{(i)} - \bar{\mathbf{y}}_X \|_2$ and $d_Z^{(i)} = \| \mathbf{S}^{(i)} - \bar{\mathbf{s}}_Z \|_2$, where $\|\cdot\|_2$ denotes the Euclidean norm, with analogous definitions for healthy training and prediction data.

To summarize fault-induced geometric deviations, distance distributions from healthy training data and from the prediction segment are compared using median-based statistics. The relative displacement is quantified through the ratio of medians, presented in Eq. (4),

$$r_\Theta = \frac{\mathrm{med}(d_{\Theta,\mathrm{pred}})}{\mathrm{med}(d_{\Theta,\mathrm{train}})}\,,\ \Theta \in \{X, Z\} \tag{4}$$

and through the median-centred displacement measures, defined by Eq. (5),

$$\Delta d_\Theta = d_{\Theta,\mathrm{pred}} - \mathrm{med}\left(d_{\Theta,\mathrm{train}}\right)\,,\ \Theta \in \{X, Z\} \tag{5}$$

Median-based metrics are used for robustness to asymmetric distributions and outliers, and to assess whether reliability-guided descriptor selection yields a clearer separation between nominal and degraded behaviour in an algorithm-independent manner.

2.5. *Anomaly Detection and Model Performance Metrics*

To complement the geometric analysis, unsupervised anomaly detection models are employed to assess whether representation choices (GEN vs. ESP) lead to consistent differences in detection behaviour. Two widely used techniques are considered: Isolation Forest (IF) (Liu et al., 2008) and One-Class Support Vector Machine (OCSVM) (Schölkopf et al., 2001). Their distinct operating principles allow representation-driven effects to be analysed

independently of the specific algorithmic assumptions.

Models are trained independently for each representation, resulting in four detectors: OCSVM_GEN, OCSVM_ESP, IF_GEN, and IF_ESP. For OCSVM, a radial basis function kernel is adopted, with hyperparameters tuned separately for GEN and ESP via grid search and cross-validation. For IF, anomaly decisions are obtained using a contamination-based quantile threshold derived from training scores.

Performance evaluation is conducted in three stages. First, an objective baseline assessment is performed on the nominal training and test partitions, where only normal-operation samples are available. In this case, evaluation focuses on nominal behaviour preservation, quantified by the False Positive Rate (FPR), allowing stability and potential overfitting to be assessed.

Second, the models are applied to the complete *train* segment after operational filtering. A weak operational ground-truth proxy is defined from the status identifier, treating codes {0, 2} (normal and idle) as normal and {1, 3, 4, 5} as non-normal. Under this proxy, confusion matrices and derived metrics are computed to evaluate discrimination between operating regimes while limiting spurious alarms. Performance is further stratified by *status_type_id*, reporting recall for normal regimes and specificity for non-normal regimes.

Finally, model outputs are analysed along the time axis for both training and prediction segments. Binary anomaly flags are compared with operational status and selected SCADA variables to assess temporal consistency and alignment with fault-related signal evolution. To mitigate sensitivity to isolated outliers, a persistence criterion is applied to the prediction segment, whereby a fault is considered confirmed only after at least n_{persist} consecutive anomalous samples.

3. Results and Discussion

Results follow the workflow described in Section 2, starting from the analysis of nominal space structure and proceeding to fault-induced geometric deviations and anomaly detection performance.

Fig. 2 compares the cumulative explained variance of the PCA models learned from healthy data under the GEN and ESP representations. The specialist representation concentrates variance into substantially fewer components, resulting in a more compact nominal subspace for the same explained-variance threshold.

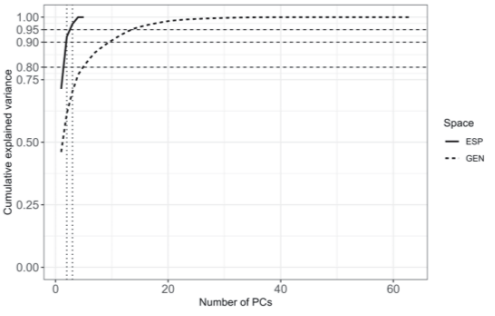

Fig. 2.Cumulative explained variance of the nominal PCA spaces for the GEN and ESP representations.

To quantify structural differences between the nominal spaces induced by GEN and ESP, subspace similarity metrics were computed using the truncated PCA bases, as defined in Section 2.3. For each variance threshold η, distances were evaluated using a common dimensionality $k_{\text{used}} = min(k_{\text{GEN}}, k_{\text{ESP}})$. As reported in Table 1, both Chordal and Frobenius distances indicate a clear structural divergence between the two representations, which persists across variance thresholds.

Table 1. Chordal and Frobenius distances between the GEN and ESP nominal subspaces for different cumulative explained-variance thresholds.

η	k_{GEN}	k_{ESP}	k_{used}	d_{chordal}	d_{F}
80%	5	2	2	0.7867	1.1126
90%	10	2	2	0.7867	1.1126
95%	14	3	3	1.1386	1.6102

Fig. 3 provides a geometric visualization of the nominal spaces and the projection of prediction data for $\eta = 90\%$ ($k = 2$). Under nominal (health) conditions, the GEN representation exhibits a broad dispersion in the PC space, reflecting the aggregation of variables associated with multiple subsystems. In contrast, the ESP representation yields a more compact and structured nominal cluster. When prediction data are projected onto the nominal subspaces, fault-related samples

remain largely embedded within the GEN space, whereas a pronounced displacement relative to the nominal cluster is observed in the ESP space.

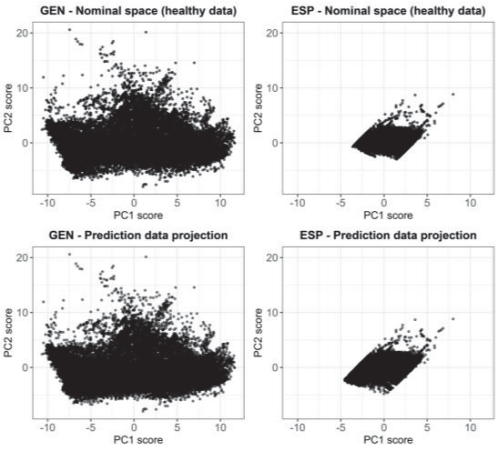

Fig. 3.PCA score plots of GEN and ESP nominal spaces (healthy data) and projection of *prediction* segment data, shown on the first two principal components (η = 90%, k = 2).

Fault-induced deviations were quantified using the distance-to-centroid metrics defined in Section 2.4. For the GEN representation, no statistically significant difference was observed between the training and prediction distance distributions (p-value = 0.78, two-sided Wilcoxon test), indicating limited geometric separation. In contrast, the ESP representation exhibits a statistically significant displacement of prediction data relative to the nominal centroid (p-value < 10^{-15}, two-sided Wilcoxon test). Median-centred displacement measures, summarized in Table 2, further confirm that fault-induced deviations are substantially larger in the ESP representation than in GEN.

Table 2. Median distance to the nominal centroid for healthy training and prediction data, and corresponding median ratios, for GEN and ESP representations (η = 90%, k = 2).

Representation	Median (health)	Median (prediction)	Median Ratio
GEN	5.67	5.52	0.97
ESP	1.92	2.63	1.37

A direct statistical comparison confirms that fault-induced deviations are significantly larger in the ESP representation than in GEN (p-value < 10^{-15}, one-sided Wilcoxon test), demonstrating that reliability-guided descriptor selection enhances geometric separability independently of any anomaly detection model.

The next step is to assess whether this structural advantage translates into differences in anomaly detection performance. To this end, IF and OCSVM models were applied to the GEN and ESP representations under identical training and evaluation protocols. Results are reported for the four detectors defined in Section 2.5 (OCSVM_GEN, OCSVM_ESP, IF_GEN, and IF_ESP), with all models trained exclusively on nominal health data.

For OCSVM, both representations share the same regularization parameter (ν = 0.01), while the kernel width differs between representations (γ = 0.0625 for ESP and γ = 0.03125 for GEN). This difference reflects the distinct geometric scales of the two feature spaces and results in a lower model complexity for ESP, which relies on 194 support vectors compared with 392 for GEN.

For IF, the ESP configuration consists of 128 trees operating on 5 descriptors, with splits over 3 variables at each node. In turn, the GEN configuration requires 256 trees operating on 63 descriptors, with splits over 47 variables. All the remaining hyperparameters (contamination = 0.05, subsampling ratio = 0.5, and feature fraction per split = 0.75) are equal across representations, ensuring a consistent comparison.

On the nominal training and test sets, which contain exclusively healthy samples, performance is characterized by the FPR. IF exhibits similar FPRs for GEN and ESP, with 955 false alarms out of 19,098 samples (FPR ≈ 5.0%). On the other hand, OCSVM consistently produces fewer false alarms under the ESP representation (190 false positives, FPR ≈ 1.0%), whereas the GEN representation produced 230 false positives (FPR ≈ 1.2%).

Similar trends were observed on the test set. The IF models produced comparable false positive rates for GEN and ESP (687 and 669 false alarms, respectively), while OCSVM again exhibited a clearer representation effect, with 93 false positives for ESP and 275 for GEN. These results indicate a consistent reduction in spurious alarms under the specialist representation and a more stable nominal behaviour.

The analysis was extended to the full operational training segment using a weak operational proxy derived from the status codes. Confusion matrices for both algorithms and representations are reported in Fig. 4.

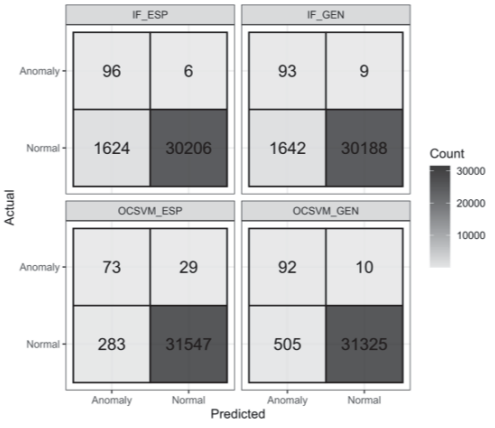

Fig. 4.Confusion matrices of IF and OCSVM for GEN and ESP representations on the operational *train* segment dataset.

For IF, the ESP representation yields a strictly improved operating point, with higher True Positives (TP) and True Negatives (TN), and fewer False Positives (FP) and False Negatives (FN). For OCSVM, ESP leads to a more conservative decision boundary, reducing false positives and increasing specificity at the expense of sensitivity. Overall, both algorithms benefit from the ESP representation in terms of alarm selectivity, although the sensitivity-specificity trade-off remains algorithm-dependent.

Temporal behaviour was analysed along the prediction segment by comparing anomaly flags with the evolution of transformer cell temperature (Fig. 5). ESP-based models exhibit earlier and more temporally consistent anomaly activation aligned with the sustained temperature increase, whereas GEN-based models show delayed or intermittent responses.

To formalize this observation, a persistence criterion of $n_{\text{persist}} = 20$ consecutive anomalous samples was applied. Since the SCADA sampling interval is 10 minutes, $n_{\text{persist}} = 20$ corresponds to a sustained anomaly lasting 200 minutes (3 h 20 min), which provides a physically interpretable confirmation window and reduces spurious confirmations due to short-lived fluctuations, particularly for thermal-related events.

Fault confirmation times are reported in Table 3. Models trained on the ESP representation confirm the presence of the fault substantially earlier than their GEN counterparts, particularly for IF. Although the exact physical onset of the fault is unknown, these results indicate that reliability-guided descriptor selection improves temporal coherence and persistence behaviour, leading to earlier and more consistent fault confirmation.

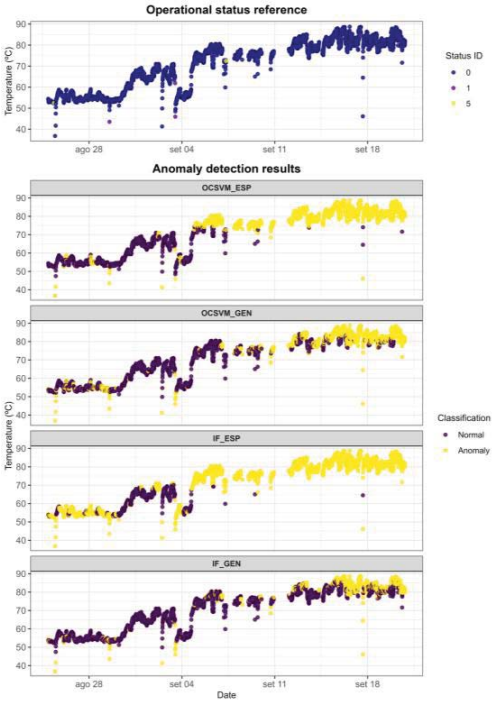

Fig. 5.Time evolution of transformer cell temperature during the *prediction* segment, with anomaly detection results for OCSVM and IF under GEN and ESP representations.

Table 3. Fault confirmation index (Conf. Index) and time (Conf. Time) under a persistence criterion of $n_{\text{persist}} = 20$.

Model	Conf. Index	Conf. Time
OCSVM_ESP	468	2023-08-28 9:10
OCSVM_GEN	579	2023-08-29 3:40
IF_ESP	39	2023-08-25 4:50
IF_GEN	2570	2023-09-16 5:40

The results demonstrate that incorporating reliability knowledge at the descriptor selection stage consistently enhances the geometric

separability of fault-related behaviour and its manifestation in unsupervised anomaly detection, independently of the specific detection algorithm.

4. Conclusions

This work analysed the role of reliability knowledge in unsupervised fault detection by isolating the impact of descriptor selection on nominal space geometry, fault-induced deviations, and anomaly detection performance. Rather than introducing algorithmic modifications, the study focused on how representation design conditioned detection outcomes in the analysed case study.

Results showed that reliability-guided descriptor selection led to more compact and structurally distinct nominal spaces, in which fault-related deviations emerged more clearly at a geometric level. These representation effects translated into more selective and temporally coherent anomaly detection behaviour across both IF and OCSVM models, reducing spurious alarms while enabling earlier and more stable fault confirmation under persistence-based decision rules. Although sensitivity-selectivity trade-offs remained algorithm-dependent, specialist representations consistently improved the alignment between geometric deviation and detection outcome in this application.

Nonetheless, the present study is limited to a single turbine and fault mode and relies on SCADA-based descriptors and weak operational proxies. As such, the reported improvements should be interpreted as case-dependent rather than general. Future work should therefore examine how descriptor relevance varies across failure mechanisms, how persistence criteria can be adaptively defined to balance early warning and false alarms, and how reliability-guided representations interact with more complex detection architectures.

Overall, the results indicate that unsupervised fault detection effectiveness and interpretability depend primarily on feature space representation rather than algorithm choice. Reliability-guided descriptor selection thus improves robustness and early warning capability in the studied case.

References

Golub, G. H., and C. Reinsch. (1970). Singular value decomposition and least squares solutions. Numerische Mathematik 14, 403–420.

Gück, S., J. Klemke, and J. A. Schmid. (2024). CARE to Compare: A real-world benchmark dataset for early fault detection in wind turbine data. *Data* 9, 138.

Jolliffe, I. T. (2002). *Principal Component Analysis* (2nd ed.). Springer.

Liu, F. T., K. M. Ting, and Z.-H. Zhou. (2008). Isolation Forest. In Proceedings of the 2008 IEEE International Conference on Data Mining, pp. 413–422. IEEE.

Nogueira, W. F., A. H. A. Melani, and G. F. M. de Souza. (2025). Wind Turbine Fault Detection Through Autoencoder-Based Neural Network and FMSA. *Sensors* 25, 4499.

Schölkopf, B., J. C. Platt, J. Shawe-Taylor, A. J. Smola, and R. C. Williamson. (2001). Estimating the support of a high-dimensional distribution. *Neural Computation* 13, 7, 1443–1471.

Sikorska, J. Z., M. Hodkiewicz, and L. Ma. (2011). Prognostic modelling options for remaining useful life estimation by industry. *Mechanical Systems and Signal Processing* 25, 1803–1836.

Silva, R. F., A. H. A. Melani, M. A. C. Michalski, and G. F. M. de Souza. (2022). Failure Mode and Observability Analysis (FMOA): An FMEA-based method to support fault detection and diagnosis. In M. C. Leva, E. Patelli, L. Podofillini, and S. Wilson (Eds.), Proceedings of the 32nd European Safety and Reliability Conference (ESREL 2022), pp. 1220–1227. Research Publishing.

Souza, G. F. M., A. Caminada Netto, A. H. A. Melani, M. A. C. Michalski, and R. F. Silva. (2021). *Reliability Analysis and Asset Management of Engineering Systems*. Elsevier.

Von Hahn, T. and C. K. Mechefske. (2022). Knowledge Informed Machine Learning using a Weibull-based Loss Function. *Journal of Prognostics and Health Management* 2, 1.

Yan, R., Z. Zhou, Z. Shang, Z. Wang, C. Hu, Y. Li, Y. Yang, X. Chen, and R. X. Gao. (2025). Knowledge driven machine learning towards interpretable intelligent prognostics and health management: Review and case study. *Chinese Journal of Mechanical Engineering* 38, 5.

Zio, E. (2022). Prognostics and Health Management (PHM): Where are we and where do we (need to) go in theory and practice. *Reliability Engineering & System Safety* 218, 108119.

Acknowledges

The authors Gilberto F. M. Souza and Renan L. Ricardo gratefully wish to acknowledge the support of the Brazilian National Council for Scientific and Technological Development / Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) by research grants 303986/2022-0 and research assistant grant.