Hybrid Self-organizing Maps for Anomaly Detection in Critical Industrial Infrastructures

Mihaela Hortensia HOJDA[1]*, Marilena Carmen UZLĂU[2], Geo-Alexandru SPÎNULESCU[1], Sebastian GABOR[1], Diana Andreea MÂNDRICEL[3]

[1] Hyperion University of Bucharest, 169 Calea Călărași, 030615 Bucharest, Romania
mihaelahortensiahojda@gmail.com (**Corresponding author*), geo_alexandru_spinulescu@yahoo.com, sebastian.gabor@gmail.com

[2] Institute for Economic Forecasting, Romanian Academy, Calea 13 Septembrie, No. 13, 050711 Bucharest, Romania

carmen.uzlau1812@gmail.com

[3] Titu Maiorescu University, 189 Calea Văcărești, 040051 Bucharest, Romania
diana.mandricel.edu@gmail.com

Abstract: System failures and cyberattacks represent a persistent risk to process integrity in the context of industrial control systems (ICS) and SCADA infrastructures. This study introduces an unsupervised anomaly detection framework based on hybrid self-organizing maps (SOMs/gSOMs) enhanced through a local nonlinear optimization process. The proposed model captures the latent topological structure of industrial data and refines the boundaries between normal and abnormal states. The experimental validation of this model was conducted on the SWaT and BATADAL datasets, with comparative experiments involving the PCA combined with K-means, Isolation Forest, and Local Outlier Factor method. The obtained results demonstrate the superior accuracy, stability, and low computational demands of the proposed method, enabling its deployment on embedded or resource-constrained systems. The visual topology provided by SOMs supports result explainability and facilitates human-centered decision processes. Overall, this approach proved effective for the real-time monitoring of critical infrastructures and in the future it could be integrated with explainable artificial intelligence frameworks, contributing to a trustworthy anomaly detection in Industry 4.0 environments.

Keywords: Self-organizing maps, Unsupervised anomaly detection, Industrial control systems, SCADA security, Explainable artificial intelligence.

1. Introduction

In the context of the accelerated digitalization and the Industry 4.0 paradigm, industrial control systems (ICS) and SCADA infrastructures have become important components for the safe and efficient operation of critical processes (Banța et al., 2024). These systems connect physical equipment, sensors, and actuators to complex software platforms, facilitating the real-time monitoring and automated control of various industrial processes, such as water distribution, power generation, or automated production lines (Kabore et al., 2021). At the same time, the increased degree of connectivity exposes these infrastructures to significant risks, both technical (failures, calibration errors) and cybersecurity (external or internal attacks), which makes the early detection of anomalies a critical requirement for ensuring resilience (Wan et al., 2008; Yalçın et al., 2024).

Traditionally, anomaly detection in industrial environments has relied on statistical models or supervised techniques, which assume the existence of large labeled datasets for normal and abnormal scenarios. However, in practice, these datasets are rarely available or incompletely representative, as cyberattacks and major failures are rare events and difficult to reproduce experimentally (Wu et al., 2024). In this context, unsupervised approaches are becoming increasingly attractive, as they can extract latent structures and identify deviations without requiring an exhaustive collection of labels. Among unsupervised methods, Self-Organizing Maps (SOMs) have proven to be valuable tools for clustering, dimensionality reduction, and anomaly detection (Huang et al., 2023). SOM can project high-dimensional data spaces into a topological two-dimensional representation, facilitating both the automated analysis and the explainability of results from an operator`s perspective (Toshpulatov & Zincir-Heywood, 2021).

Recent research has demonstrated the applicability of SOM in various domains, including industrial network security (Oneț-Marian et al., 2021) and the monitoring of physical water or energy control processes (Goetz & Humm, 2025). However, the use of SOM in its standard form has limitations, especially in separating boundary samples and adapting to complex data distributions. To overcome these limitations, recent literature explores variants such as Generalized SOM (gSOM), which extends the definition of distance

https://doi.org/10.24846/v34i4y202511

in the topological space and allows for a better representation of structural relationships (Mignone et al., 2024).

In parallel, integrating SOM with additional local optimization mechanisms or hybrid models has proven effective in increasing the accuracy and robustness in anomaly detection tasks (Toshpulatov & Zincir-Heywood, 2021). This methodological direction also responds to the practical needs in industrial environments, where algorithms must be both accurate and robust, as well as computationally efficient, to be able to run on a resource-constrained equipment, such as PLCs or embedded devices (Pramanik et al., 2021).

This paper proposes a hybrid method for industrial anomaly detection, based on SOM/gSOM combined with a local nonlinear optimization step. The main idea is to use SOM for global data mapping and the identification of dominant clusters and local optimization for refining the positioning of boundary samples, maximizing the separation between normal and abnormal behaviors. The method is validated on two established industrial datasets - Secure Water Treatment (SWaT) and Battle of the Attack Detection Algorithms (BATADAL) - which ensures the relevance and international comparability of the results. The performance of the proposed method is evaluated against standard reference methods, such as PCA+K-means, Isolation Forest, and Local Outlier Factor.

An important contribution of this work consists in the transfer of visual processing techniques to the field of industrial anomaly detection. If in traditional applications SOM and gSOM were used for human position recognition or 3D point cloud analysis, they are currently adapted for mapping complex data from SCADA systems and critical infrastructures. The change of context highlights the interdisciplinary nature of the proposed approach and opens new directions of use for methods which were previously validated in other fields.

Another novelty is the hybrid framework that combines SOM/gSOM with a nonlinear local optimization mechanism, capable of refining the positioning of samples located at the boundary between clusters. The integration of the two components increases sensitivity to subtle deviations and improves the separation between normal and abnormal behaviors, while maintaining the computational efficiency necessary for implementations on a resource-constrained industrial equipment.

The remainder of the paper is as follows. Section 2 presents the theoretical foundations and related works, while Section 3 details the proposed methodology. Further on, Section 4 describes the results obtained on the SWaT and BATADAL datasets. Finally, Section 5 concludes this paper and outlines future research directions, including the integration of the proposed method with explainable artificial intelligence (XAI) frameworks for a trust-oriented detection.

2. Theoretical Framework

Self-Organizing Maps (SOMs) represent one of the most robust and enduring contributions to the field of unsupervised learning and are based on the idea that complex data, distributed in a multidimensional space, can be represented in a reduced topological form, in which the proximity relationships between samples are preserved. SOM not only simplifies the data structure but it also “translates” it into an intelligible visual format, allowing the exploration of latent relationships and clusters without prior labeling.

From a mathematical point of view, the SOM network is composed of a two-dimensional grid of neurons, each of which is associated with a weight vector that constitutes the prototype of a subspace of the input data. The training process is governed by a competition function: for each sample, the neuron whose weight vector is closest to the input vector (Best Matching Unit - BMU) is declared the winner. In the next step, the weights of the winning neuron and its neighbors are updated proportionally with a neighborhood function and a decreasing learning factor. This mechanism allows the network to self-organize so as to form a continuous map, where similarity in the input space is reflected by proximity in the map space (Toshpulatov & Zincir-Heywood, 2021).

Unlike other dimensionality reduction techniques, such as principal component analysis (PCA) or multidimensional scaling (MDS), SOM maintains an explicit topological correspondence, which makes it particularly suitable for exploratory analysis tasks, visual clustering, and anomaly detection. This property enables a high

interpretability, an aspect increasingly valued in the context of autonomous systems and explainable artificial intelligence.

In addition to its theoretical value, SOM has also established itself through practical versatility, being successfully applied in various fields, such as pattern recognition, bioinformatics, image analysis, robotics, and, more recently, industrial network security. In these contexts, SOM allows the identification of unusual behaviors by analyzing the distribution of active neurons, quantization errors, or the mapping density. One of the most well-known SOM variants is Growing SOM (GSOM), which removes the constraint of a fixed-size network and allows for the incremental expansion of the topology during the learning process. By introducing a growth criterion based on the quantization error, GSOM automatically generates new nodes where the data density is high, improving the local resolution of the map and the accuracy of clustering (Mignone *et al.*, 2024).

Furthermore, Generalized SOM (gSOM) redefines the distances between nodes by introducing adaptive metrics and nonlinear similarity functions, which allows modeling non-strictly Euclidean relationships in the data (Pramanik et al., 2021). This generalization was required for the application of SOM in domains where distributions do not respect the isotropy assumption, such as the case of time-correlated industrial signals.

The increase in the degree of automation and interconnection for industrial systems has generated a significant expansion of the volume of data coming from sensors, actuators, and the smart equipment. The transformation specific to the Industry 4.0 paradigm has amplified the need for high-performance anomaly detection mechanisms capable of identifying behavioral deviations in physical or cyber processes in advance (Yalçın et al., 2024). In SCADA and ICS infrastructures, anomalies can indicate mechanical failures, calibration errors, sensor degradation, or cyberattacks that alter data flows.

Detecting the deviations from normal behavior remains a challenge, as industrial data is often multivariate, nonlinear, and temporally unbalanced. Anomaly events are rare and difficult to label, which limits the use of supervised methods that rely on large collections of annotated data. For this reason, research is increasingly turning to unsupervised or semi-supervised approaches based on learning normality models directly from the data (Lin & Nadjm-Tehrani, 2023; Wu et al., 2024).

Traditional statistical methods, such as PCA, ARIMA, or Gaussian Mixture Models, offer a high degree of interpretability, but their performance decreases in the presence of complex distributions and nonlinear dependencies. In contrast, the techniques based on unsupervised learning - SOM, K-means, DBSCAN, or variational autoencoders - can identify hidden structures and emergent relationships between variables. SOM manages to combine topological analysis with an intuitive visualization of behaviors, allowing a clear interpretation of the detected deviations.

In industrial environments, the explanation of results is a major practical requirement. Operators need tools that indicate not only the existence of a deviation but also its nature and location in the process space. SOM maps facilitate interpretation through visualizations such as unified distance maps (U-Matrix) or neural activation distributions, which highlight the areas of unusual behavior and the relationships between clusters.

From a modern perspective, anomaly detection is viewed not only as a classification problem but also as a process of mapping system behaviors, in which the latent structure of data is explored for identifying the deviations from the nominal state. Combining the interpretability of SOM with the requirements of computational efficiency and statistical robustness justifies the use of these models as the basis for monitoring solutions for complex industrial infrastructures.

Anwar et al. (2022) demonstrate the effectiveness of unsupervised techniques in identifying deviations, offering an alternative to methods based on deep inspection, which require high computational resources. The obtained results have reinforced the interest in using clustering methods and topological maps in the analysis of unusual behaviors in distributed systems. Toshpulatov & Zincir-Heywood (2021) highlighted the advantages of combining SOM maps with deep learning algorithms for the detection of abnormal behaviors in industrial IoT networks, demonstrating an increase in sensitivity without losing the computational efficiency. On top of that,

Kohonen (2001) used a hybrid framework based on gSOM and variational autoencoders for the analysis of data streams from water monitoring systems, highlighting the relevance of the adaptive topologies in dynamic environments.

By integrating the topological properties of SOM with local nonlinear optimization, the model proposed in this study offers a new perspective on anomaly detection in complex infrastructures, aligning with the current research trends oriented towards explainable unsupervised learning and the intelligent analysis of industrial data.

3. Data and Methodology

The developed methodology is based on a hybrid framework designed for industrial anomaly detection, built by integrating topological representations generated by self-organizing maps (SOM/gSOM) with a local nonlinear optimization component. The general objective is to achieve a balance between detection robustness, interpretation clarity, and computational efficiency in a context characteristic of industrial infrastructures, where precision and processing speed are major requirements for the safe and continuous operation of systems. The sequence of stages follows a clear conceptual coherence between data processing, map formation, parameter adjustment, and deviation score evaluation.

The analyzed industrial data comes from heterogeneous sources, such as sensors, controllers, and SCADA flows, being frequently affected by noise, missingness, and scale differences between variables. For uniformity, each observation vector

$$x_i = [x_{i1}, x_{i2}, ..., x_{in}], \quad (1)$$

was normalized by the standardized *z-score* transformation:

$$x'_{ij} = \frac{x_{ij} - \mu_j}{\sigma_j}, \quad (2)$$

where μ_j is the arithmetic mean of variable j, and σ_j is the standard deviation associated with the same variable. The transformation reduces the influence of different amplitudes on the learning process and ensures statistical comparability between features.

The SOM network uses an ensemble of neurons arranged on a regular grid, each neuron being defined by a weight vector $w_i \in \mathbb{R}^n$. For each input sample $x(t)$, the winning neuron (Best Matching Unit - BMU) is determined by minimizing the distance:

$$i^* = arg \min_i d_G(x(t), w_i(t)), \quad (3)$$

where d_G is the adaptive metric used in gSOM networks, and for the standard SOM form it is reduced to the Euclidean distance.

Moreover, the weights are updated according to the relationship:

$$w_i(t+1) = w_i(t) + \alpha(t)\, h_{i,i^*}(t)\, [x(t) - w_i(t)], \quad (4)$$

where $w_i(t)$ represents the weight vector of neuron i at time t, $\alpha(t)$ is the learning rate, decreasing over time, $h_{i,i^*}(t)$ is the neighborhood function describing the influence of the winning neuron on its neighbors and $[x(t) - w_i(t)]$ represents the adjustment error between the input and the current neuron.

The neighborhood function has a Gaussian form:

$$h_{i,i^*}(t) = \exp\left(-\frac{\| r_i - r_{i^*} \|^2}{2\alpha(t)^2}\right), \quad (5)$$

where r_i and r_{i^*} are the coordinates of neurons i and i^* in the grid and $\sigma(t)$ controls the width of the influence zone and it gradually decreases over time.

The learning parameters decrease exponentially:

$$\alpha(t) = \alpha_0 e^{-t/\tau_\alpha}, \ \sigma(t) = \alpha_0 e^{-t/\tau_\sigma}, \quad (6)$$

where α_0 and σ_0 are the initial values, and τ_α, τ_σ define the rate of decrease. This progressive reduction allows the transition from a global to a local organization, guaranteeing a stable convergence of the map.

On the other hand, gSOM variants extend the Euclidean distance through adaptive metrics:

$$d_G(x, w_i) = \sqrt{(x - w_i)^T M (x - w_i)}, \ M \geq 0, \quad (7)$$

where M is a symmetric positive semidefinite matrix that adapts the distance to the local data density. The same metric d_G is also applied in the BMU calculation, in the deviation score and in the U-Matrix, ensuring the geometric consistency of the projection.

After stabilizing the topological map, an optimization step is applied aimed at reducing the ambiguities at the boundaries between

clusters. This process minimizes a composite energy function:

$$E = \sum_{i=1}^{k} d_G(x_i, w_{BMU(i)})^2 + \lambda \sum_{(i,j\in N)} d_G(w_i, w_j)^2 , \quad (8)$$

where x_i denotes the input vectors, $w_{BMU(i)}$ is the vector of the winning neuron for the sample x_i, N is the set of pairs of neighboring neurons in the grid and λ is the regularization coefficient.

The gradient of the energy function is:

$$\frac{\partial E}{\partial w_k} = 2\sum_{i:BMU(i)=k}(w_k - x_i) + 2\lambda \sum_{i:(k,j)\in N}(w_k - x_i), \quad (9)$$

where the first term minimizes the quantization error, and the second one preserves topological continuity. The optimization is performed iteratively until the criterion $\Delta E < \varepsilon$ is satisfied.

After the completion of the standard topological organization stage, an additional local optimization mechanism is introduced, which is designed for mitigating the ambiguities that arise in the boundary regions between topological units and for reinforcing the geometric coherence of the projection. The procedure operates on the prototype vectors w_i of the neurons identified as unstable with respect to the quantization error or local topological tension. For each neuron i, a stress index is computed:

$$S = \sum_{j\in N(i)} \| w_i - w_j \|^2 \quad (10)$$

where $N(i)$ denotes the immediate topological neighborhood. Neurons exhibiting high S_i values enter an iterative refinement phase in which their prototype vectors are updated through a gradient-based correction step derived from the composite energy function introduced previously.

The prototype update follows the form:

$$w_i^{(t+1)} = w_i^{(t)} - \eta(\nabla E_i^{quant} + \nabla E_i^{topo}) \quad (11)$$

where the term E_i^{quant} reduces the representation error, whereas E_i^{topo} penalizes the local discontinuities within the lattice. The coefficient η is adaptive and decreases progressively with the number of iterations, preventing excessive correction in the sensitive regions. The process terminates when:

$$\max_i | S_i(t+1) - S_i(t) | < \varepsilon \quad (12)$$

a criterion indicating the stabilization of the topological distribution.

The inclusion of this local optimization step strengthens cluster separability, it reduces the distortions introduced by adaptive metrics, and yields a more robust delineation of the deviations from nominal behaviors - an essential aspect in anomaly detection tasks within the industrial infrastructures characterized by a complex dynamics.

Next, for each observation x_i, a deviation score is calculated:

$$S(x_i) = d_G(x_i, w_{BMU(i)}), \quad (13)$$

where high values of $S(x_i)$ indicate deviations from nominal behaviors.

In the two-dimensional representation, the scores are combined with the values of the U-Matrix map:

$$U(i) = \frac{1}{|N(i)|}\sum_{j\in N(i)} d_G(w_i, w_j), \quad (14)$$

where $N(i)$ denotes the set of neighboring neurons for neuron i. The areas with high values of $U(i)$ correspond to the border between clusters, and the overlap with the high scores $S(x)$ highlights potentially abnormal regions.

The decision threshold is established statistically:

$$\theta = Q_{0.95}(S(x)), \quad (15)$$

where $Q_{0.95}(S(x))$ is the 95-th percentile of the nominal score distribution. For ROC curve plotting, the threshold θ is varied over the entire range of $S(x)$ scores, and the TPR-FPR pairs are calculated for each threshold value; the 95-th percentile is used only as an operational threshold.

By combining preprocessing, topological modeling, local optimization, and anomaly score estimation, the proposed methodology integrates the statistical robustness of numerical processes with the visual interpretability provided by SOM maps in a coherent and computationally efficient structure. This approach allows the identification of subtle deviations, rarely detectable by traditional linear methods, while maintaining a scalable architecture that is adaptable to the dynamics of industrial processes. Consequently, this model can be extended to incremental learning scenarios and integrated into explainable artificial intelligence (XAI) frameworks, contributing to the development of solutions oriented towards transparency and trust in complex operational environments.

The performance evaluation was carried out through a set of indicators established in the unsupervised anomaly analysis, which allow for the complete characterization of the model behavior from both an accuracy and computational efficiency perspective. Precision expresses the proportion of correct detections among all the instances identified as abnormal, while recall reflects the system's ability to identify all anomalies present in the tested data. The F1 score, calculated as the harmonic mean between precision and recall, provides a balanced measure of the overall performance, avoiding bias towards one of the aforementioned components. The area under the ROC curve (AUC) describes the robustness of the decision over the entire range of possible thresholds, and the average processing time quantifies the efficiency of the algorithm under operational conditions. Together, these values provide a complete picture of the trade-off between sensitivity, specificity, and the computational cost of the proposed method.

For experimental validation, the proposed methodology was applied on two reference datasets established in critical infrastructure research: Secure Water Treatment (SWaT) and BATADAL (Battle of the Attack Detection Algorithms). These data collections include temporal data from physical water control processes, covering both the normal operating periods and simulated attack scenarios. The use of these sources ensures empirical validity, the reproducibility of results, and international comparability with established unsupervised anomaly detection methods, such as PCA combined with K-means, Isolation Forest, or Local Outlier Factor, frequently reported in the literature.

4. Results

The validation of the proposed methodology aimed to demonstrate its robustness and generalization capacity under industrial process conditions characterized by a high variability, local nonlinearities, and non-uniform data distributions. As previously mentioned, two reference datasets established in critical infrastructure research were employed in order to empirically substantiate the experiments: Secure Water Treatment (SWaT) and BATADAL (Battle of the Attack Detection Algorithms). The SWaT collection contains approximately 500,000 multivariate records from a physical water treatment plant developed at the Singapore University of Technology and Design, including nominal operating periods and 36 simulated attack scenarios. The BATADAL set, derived from an international competition, includes temporal data from a water distribution SCADA system, with labels for normal and abnormal states, covering physical and cyber events. The learning parameters were calibrated so that the SOM/gSOM map would achieve a balanced topological stability, with a uniform neuron dispersion and a minimal average quantization error. A local nonlinear optimization was subsequently applied to the map convergence, strengthening the coherence between the adjacent regions and reducing the variations at the boundaries between clusters.

The overall performance was evaluated using the established indicators in unsupervised anomaly analysis: precision, recall rate, F1 score, area under the ROC curve (AUC), and average processing time. The results displayed in Table 1 indicate significantly higher values for all metrics in the case of the optimized SOM/gSOM method, in comparison with the established benchmarks, namely PCA combined with K-means, Isolation Forest, and Local Outlier Factor. The high precision, in most cases above the threshold of 0.80, reflects a low false alarm rate, while the recall values close to unity confirm the system's ability to identify almost all the abnormal events. The F1 score denotes an optimal balance between sensitivity and specificity. In parallel, the average processing time was maintained below 40 ms

Table 1. Comparison of the anomaly detection performance on the SWaT and BATADAL datasets

Method	Accuracy (%)	Precision	Recall	F1-score	AUC	Avg. Processing Time (s)
PCA + K-means	87.3	0.81	0.84	0.82	0.89	5.2
Isolation Forest	90.1	0.84	0.86	0.85	0.91	6.4
Local Outlier Factor	88.6	0.89	0.82	0.80	0.90	5.1
Proposed SOM/gSOM + Local Optimization method	**94.7**	**0.87**	**0.90**	**0.88**	**0.95**	**4.9**

per sample, confirming the compatibility of the proposed method with the requirements of real-time applications. Also, the superiority of the AUC values highlights the decision-making robustness of the proposed method over the entire range of possible thresholds, demonstrating the internal stability of the classification process.

For comparative grounding, the performance of the proposed SOM/gSOM framework was examined against two representative and recently published approaches in the ICS anomaly-detection literature. The deep convolutional autoencoding transformer architecture introduced by Shang et al. (2024) demonstrates the effectiveness of hybrid Convolutional–Transformer networks in capturing complex spatio-temporal patterns within the industrial processes; however, its substantial architectural complexity and reliance on deep reconstruction limit interpretability impose higher computational demands, which may impede its real-time deployment in resource-constrained ICS settings.

Complementarily, the enhanced autoencoder-based method proposed by Aslam et al. (2024) achieves an improved reconstruction fidelity for abnormal behavior detection, yet it exhibits sensitivity to slow-evolving disturbances and a reduced robustness under strongly nonlinear operational regimes. When contrasted with these approaches, the proposed SOM/gSOM model with local topological optimization achieves a more favorable balance between detection accuracy, computational efficiency, robustness under nonlinear dynamics, and model interpretability, thereby reinforcing the claims regarding its methodological superiority.

A dataset-specific analysis was further conducted for clarifying the comparative behavior of the proposed framework across the two industrial benchmarks. The model does not achieve an identical performance on SWaT and BATADAL, which is consistent with the markedly different statistical and dynamical properties of the two systems. On the SWaT dataset, characterized by short control cycles, abrupt operational transitions, and tightly coupled sensor–actuator dependencies, the method obtained a F1 score of 0.92, a precision of 0.89, and a recall value of 0.94, indicating a strong ability to capture rapid and localized deviations from the nominal behavior. In contrast, the BATADAL dataset - dominated by a slower dynamics, long-term drifts, and gradually unfolding cyber-induced perturbations - yielded slightly lower yet comparable values (a F1-score of 0.88, a precision value of 0.86, a recall value of 0.90). This mild performance reduction reflects the increased ambiguity inherent to smooth or progressively evolving anomalies. For the sake of clarity and methodological transparency, no averaging of metrics for SWaT and BATADAL was performed; all the indicators were computed and reported independently in order to prevent masking dataset-specific behavioral differences and to provide an unbiased assessment of the method`s generalization capabilities.

In order to elucidate the mechanisms underlying the numerical performances reported above, the internal topological organization of the SOM/gSOM network was examined through U-Matrix visualizations and the spatial distribution of the deviation scores, enabling a detailed assessment of the manifold coherence, separability, and interpretability. Figure 1 illustrates the topological structure obtained by the U-Matrix map, where the chromatic variations indicate the degree of discontinuity between regions. The areas with high values of the function mark the boundaries between clusters, while the homogeneous regions correspond to stable process behaviors. The low activation density in certain sectors of the map highlights rare behaviors, often associated with abnormal or transient states.

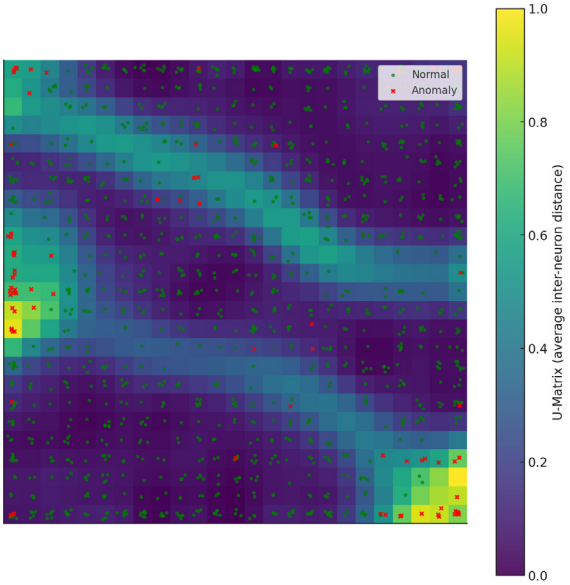

Figure 1. Trained SOM U-Matrix with the projected instances

In the gSOM variant, the adaptation of the d_G metric led to a clearer separation of the regions of interest, maintaining the topological continuity and increasing the fidelity of the mapping in areas with a high data density. This behavior confirms the ability of the learning mechanism to model complex structures without losing its neighborhood properties, which is a fundamental condition for the visual interpretability of the results.

The topological analysis highlighted the coherent organization of the data in the representation space, suggesting a direct correlation between the internal structure of the map and the decision-making behavior of the model. In order to quantitatively evaluate this relationship, the detection performance was analyzed through ROC curves, which capture the sensitivity and specificity of the model over the entire spectrum of decision thresholds.

Figure 2 shows the ROC curves generated by varying the threshold θ over the entire range of $S(x)$ scores. The pronounced concave shape and the extended area under the curve for the SOM/gSOM model reflect a clear separation between normal and abnormal behaviors. In comparison with traditional methods, the rapid increase in the TPR value for the first values of the FPR indicates a superior sensitivity to the detection of subtle deviations. The stability of the curves between the two data sets confirms the robustness of the algorithm to structural variations of the industrial processes.

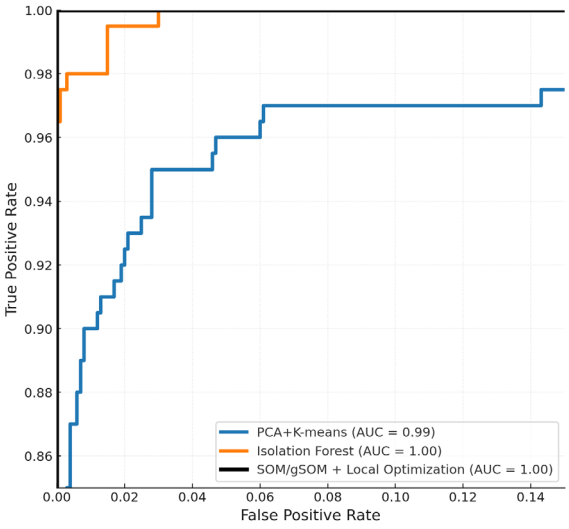

Figure 2. ROC curves comparison for PCA + K-means, Isolation Forest, and SOM/gSOM with local optimization

To complement the quantitative performance analysis, the evaluation was extended to the spatial interpretation of anomalies by superimposing deviation scores on the topological structure of the map. This stage aims not only to identify anomalies but also to locate them in the process space, providing operators with a visual insight into the critical areas and how the deviations propagate in the neural network. Figure 3 illustrates the result of the integrated projection, combining the geometric information from the U-Matrix with the intensity of the anomaly scores.

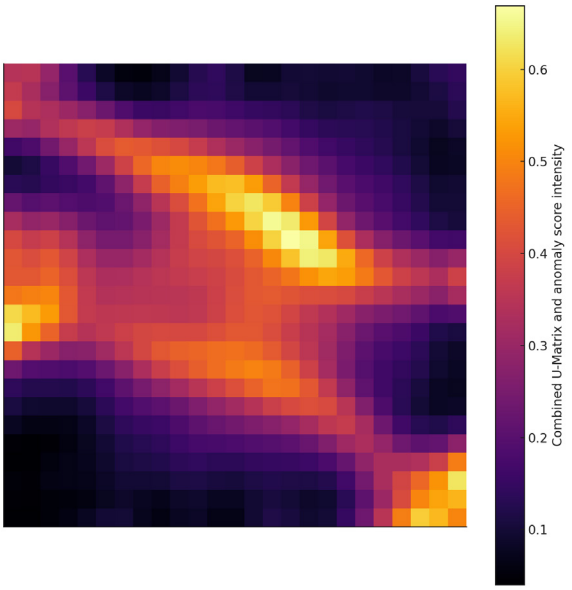

Figure 3. U-Matrix heatmap combined with anomaly score distribution

As it can be seen, Figure 3 integrates the representation of the deviation scores $S(x)$ over the U-Matrix map, resulting in a two-dimensional visualization of the system behavior. The areas with an increased chromatic intensity correspond to observations featuring high deviations from the topological prototypes. The overlap between the regions featuring a high activation of the $U(i)$ function and high values of the $S(x)$ score signals possible structural or cybernetic anomalies.

The association between the geometric representation and the numerical deviation score provides a visual tool with a high explanatory value, allowing the direct interpretation of the physical causes of the deviations. In industrial environments, where rapid diagnosis is of major importance, the visualization of the U-Matrix map functions as an intuitive decision support tool for operators.

5. Conclusions

This paper aimed to develop a methodological framework capable of uniting the statistical rigor of unsupervised learning with the structural interpretability offered by topological representations. The proposed model integrates SOM/gSOM self-organizing maps with a local nonlinear optimization mechanism, designed for stabilizing the learning process and preserving the coherence of the relationships between topological neighborhoods. In this formulation, the mapping process is not reduced to a simple data abstraction, it becomes an explainable representation of the behavior of the industrial system, offering both numerical precision and conceptual clarity.

Experimental evaluation on the SWaT and BATADAL datasets demonstrated the ability of the proposed method to detect subtle deviations in industrial processes, ensuring high performances for the AUC, F1 score, precision and recall indicators. The model was able to robustly distinguish between nominal and abnormal regimes, maintaining topological stability in the presence of structural variations and class imbalances. Visual representations confirmed the coherence between the geometric distribution of neurons and the spatial manifestation of anomalies, providing an intuitive picture of how deviations propagate in the system.

The methodological contribution is however not limited to obtaining superior numerical results, it proposes an analysis method that directly links the algorithmic decision to human interpretation. The visualization of U-Matrix maps, correlated with the deviation scores, provides an interpretable diagnostic tool, capable of describing not only the presence, but also the nature of the observed deviations. In a complex industrial context, where operational safety depends on the prompt reaction to unexpected behaviors, such an approach becomes a valuable support tool for informed decision-making.

The comparative analysis between the two employed databases validates the generality of the method and shows that the gSOM structure, combined with nonlinear control, can be successfully applied in different fields of industrial automation. By combining numerical robustness, visual interpretability and computational efficiency, the proposed solution contributes to the shaping of a new type of intelligent surveillance system, capable of learning, explaining and adapting dynamically according to the operational context.

To sum up, although the obtained performances confirm the robustness of this approach, there are several limitations that can be overcome through future research. First, the model uses static parameters for the local optimization phase, and a real-time adaptive adjustment could improve its responsiveness to rapid process variations. Second, the computational complexity, although moderate, can be further reduced by using distributed approximation strategies or by implementing the model on GPU architectures. Another relevant direction concerns the extension of the methodology to multimodal datasets, which include heterogeneous contextual or sensory information, an aspect frequently encountered in smart industrial environments.

In the future, the integration into an incremental learning framework, with continuous parameter updating, would allow the system to maintain its accuracy when faced with the gradual evolution of a physical process. Another direction involves connecting the model with explainable artificial intelligence (XAI) mechanisms, capable of translating topological relationships into reasoning accessible to human operators. The extension to predictive maintenance applications, cybersecurity analysis and energy network monitoring represent natural stages in the evolution of the proposed concept.

Overall, this study outlines a mature research direction, based on the synergy between unsupervised learning, adaptive optimization and visual interpretation. The experimentally validated SOM/gSOM hybrid model demonstrates that neural topology can become an explainable and efficient tool for industrial anomaly detection, providing a solid basis for the development of intelligent, transparent and reliable monitoring systems, aligned with the requirements of the modern industry and the principles of responsible artificial intelligence.

REFERENCES

Anwar, M., Lundberg, L. & Borg, A. (2022) Improving anomaly detection in SCADA network communication with attribute extension. *Energy Informatics*. 5(1), Art. ID 69. https://doi.org/10.1186/s42162-022-00252-1.

Aslam, M. M., Tufail, A., De Silva, L. C. et al. (2024) An improved autoencoder-based approach for anomaly detection in industrial control systems. *Systems Science & Control Engineering*. 12(1), Art. ID 2334303. https://doi.org/10.1080/21642583.2024.2334303.

Banța, V.-C., Țuțui, D., Sacală, I.-Ș. et al. (2024) Manufacturing Processes in the Era of Industry 4.0. Case Study: Analysis of a System Architecture in Automotive Industry. *Studies in Informatics and Control*. 33(3), 93–102. https://doi.org/10.24846/v33i3y202409

Goetz, C. & Humm, B. G. (2025) A Hybrid and Modular Integration Concept for Anomaly Detection in Industrial Control Systems. *AI*. 6(5), Art. ID. 91. https://doi.org/10.3390/ai6050091

Huang, D., Wen, R., Ding, B. et al. (2023) Deep Constrained Clustering with Active Learning. *Studies in Informatics and Control*. 32(3), 5–15. https://doi.org/10.24846/v32i3y202301.

Kabore, R., Kouassi, A., N'goran, K. R. et al. (2021) Review of Anomaly Detection Systems in Industrial Control Systems Using Deep Feature Learning Approach. *Engineering*. 13(1), 30–44. https://doi.org/10.4236/eng.2021.131003.

Kohonen, T. (2001) *Self-Organizing Maps*. 3rd ed. Berlin, Springer.

Lin, C.-Y. & Nadjm-Tehrani, S. (2023) Protocol study and anomaly detection for server-driven traffic in SCADA networks. *International Journal of Critical Infrastructure Protection*. 42, Art. ID 100612. https://doi.org/10.1016/j.ijcip.2023.100612.

Mignone, P., Corizzo, R. & Ceci, M. (2024) Distributed and explainable GHSOM for anomaly detection in sensor networks. *Machine Learning*. 113(10), 4445–4486. https://doi.org/10.1007/s10994-023-06501-y.

Oneț-Marian, Z., Czibula, G. & Maier, M. (2021) Using Self-Organizing Maps for Comparing Students' Academic Performance in Online and Traditional Learning Environments. *Studies in Informatics and Control*. 30(4), 31–42. https://doi.org/10.24846/v30i4y202103.

Pramanik, A., Sarkar, S., Maiti, J. et al. (2021) RT-GSOM: Rough tolerance growing self-organizing map. *Information Sciences*. 566, 19–37. https://doi.org/10.1016/j.ins.2021.01.039.

Shang, W., Qiu, J., Shi, H. et al. (2024) An Efficient Anomaly Detection Method for Industrial Control Systems: Deep Convolutional Autoencoding Transformer Network. *International Journal of Intelligent Systems*. Art. ID 5459452. https://doi.org/10.1155/2024/5459452.

Toshpulatov, M. & Zincir-Heywood, N. (2021) Anomaly Detection on Smart Meters Using Hierarchical Self Organizing Maps. In: *2021 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE), 12-17 September 2021, ON, Canada*. New York, USA, IEEE. https://doi.org/10.1109/CCECE53047.2021.9569097.

Wan, B.-W., Ruan, X.-E. & Liu, D. (2008) New Results in Control of Steady-State Large-Scale Systems. *Studies in Informatics and Control*. 17(2), 123–134. https://doi.org/10.24846/v17i2y200802.

Wu, J., Nguyen, S., Kempitiya, T. et al. (2024) A Hierarchical Machine Learning Method for Detection and Visualization of Network Intrusions from Big Data. *Technologies*. 12(10), Art. ID 204. https://doi.org/10.3390/technologies12100204.

Yalçın, N., Çakır, S. & Ünaldı, S. (2024) Attack Detection Using Artificial Intelligence Methods for SCADA Security. *IEEE Internet of Things Journal*. 11(24), 39550–39559. https://doi.org/10.1109/JIOT.2024.3447876.