Entropy–Attention Integration for Anomaly Detection in Nonlinear and Uncertain Dynamics*

Oleksandr Palagin[1,†], Denys Symonov[1,*,†], Yehor Symonov[1,†], Bohdan Zaika[1,†] and Ivan Demenko[1,†]

[1] *V.M. Glushkov Institute of Cybernetics of the National Academy of Sciences (NAS) of Ukraine, Akademician Glushkov Avenue 40, 03187, Kyiv, Ukraine*

Abstract

The article addresses the problem of detecting and interpreting anomalies in the time series of complex dynamical systems operating under conditions of uncertainty, high dimensionality and nonlinear dynamics. A multilevel approach is proposed that combines the Kozachenko-Leonenko entropy (KLE) method, chaos theory methods (analysis of Lyapunov exponents and correlation dimension), and attention mechanisms (in particular, Multi-Head Attention). This integration allows not only to accurately localize anomalous areas, but also to determine their dynamic nature (random or structured), as well as to identify hidden causal and temporal dependencies between them. Particular attention is paid to the phase space reconstruction to identify fractal structures in the dynamics of anomalies. Experimental validation is carried out on sensor array data containing time series of different stages of production. The obtained results indicate the efficiency of the proposed approach for the analysis of nonstationary multivariate signals in computer science and data processing applications. This approach can be applied in decision support systems for technical diagnostics, environmental monitoring, and management of complex information systems.

Keywords

anomaly analysis, entropy analysis, Kozachenko-Leonenko entropy method, chaotic properties of anomalies, attention mechanisms, Multi-Head Attention, dynamic systems

1. Introduction

In today's environment of increasing data in complex dynamic systems characterized by high uncertainty, adaptability, and nonlinear behavior, the task of timely detection and interpretation of anomalies is becoming extremely important [1-3]. Anomalous observations in time series can signal critical changes in the system, including unstable modes, defects, or external influences, making their analysis a key element in the structure of decision support systems [4, 5]. This problem is relevant to many real-world domains, including but not limited to cybersecurity [6], healthcare [7], and finance [8].

Traditional statistical methods are often ineffective in cases of multidimensional and chaotic processes [9, 10]. In response to these challenges, this paper proposes a multilevel approach to anomaly analysis that combines mathematical and machine learning methods. The central tool is the Kozachenko-Leonenko entropy estimator (KLE), which provides high sensitivity to local changes in the distribution density. Additional consideration of the phase space allows to detect the hidden structure in the signal dynamics. Particular emphasis is placed on the study of the chaotic properties of the identified anomalies. The use of the nonlinear dynamics tools, in particular, Lyapunov's exponents and correlation dimension, helps to identify the degree of their determinism

*ITS-2025: Information Technologies and Security, December 11, 2025, Kyiv, Ukraine

* Corresponding author.

† These authors contributed equally.

palagin_a@ukr.net (O. Palagin); denys.symonov@gmail.com (D. Symonov); e.symonov@gmail.com (Y. Symonov); zaikabohdan5@gmail.com (B. Zaika); ivandemenkofifa@gmail.com (I. Demenko)

0000-0003-3223-1391 (O. Palagin); 0000-0002-6648-4736 (D. Symonov); 0009-0008-2581-2001 (Y. Symonov); 0009-0001-9567-8361 (B. Zaika); 0009-0008-2683-6523 (I. Demenko)

© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR-WS.org/Vol-4337/paper17.pdf

CEUR Workshop Proceedings ceur-ws.org ISSN 1613-0073

published 2026-11-26

and the presence of fractal objects. Another important component is the use of attention mechanisms that enable the identification of causal and temporal relationships between anomalies and the formation of more accurate predictions of future system behavior. Multi-Head Attention models provide an in-depth interpretation of spatial and temporal dependencies in time series.

To verify the approach, a dataset with complex sensory dynamics is used. The experimental results confirm the effectiveness of the developed method and its suitability for building adaptive diagnostic and prognostic systems under uncertainty.

2. Problem statement

In complex dynamic systems operating under conditions of uncertainty and multifactorial influence, the key task is to timely detect and analyze anomalous values in time series. The presence of such anomalies may indicate the approach of critical modes, internal disturbances, or external perturbations that directly affect the quality of decisions [4-7]. Classical methods of statistical analysis are not able to provide proper accuracy and sensitivity in conditions of nonlinearity, high dimensionality, and chaotic signals. There is a need to integrate methods of entropy analysis, nonlinear dynamics, and attention mechanisms to detect structural patterns, predict abnormal states, and improve the reliability of decision support systems in complex information environments.

3. Goal and objectives of the work

The goal of this work is to develop a multilevel method for detecting, analyzing, and predicting anomalous values in the time series of complex dynamical systems using the tools of entropy analysis, chaos theory, and attention mechanisms.

In order to achieve this goal, it is necessary to solve the following tasks:

1. Justification of the expediency of using the Kozachenko-Leonenko entropy (KLE) method to identify local anomalies in time series.
2. Construction of the reconstructed phase space and study of the chaotic properties of anomalous regions using Lyapunov's indices and correlation dimension.
3. Integration of attention mechanisms to detect latent causal and temporal dependencies between anomalous points.
4. Experimental validation of the proposed approach on real sensory data and evaluation of its effectiveness.

4. Detection of anomalies

Various applications of entropy for anomaly detection remain an active area of research. For example, the study [11] by Kleyton da Costa investigated the detection of anomalies in global financial markets using Graph Neural Networks (GNNs) in an uncertainty scenario measured by nonextensive entropy. Another example is the study [12] by Shuo Liang et al., which proposed a multivariate time series anomaly detection model based on transfer entropy and a graph attention network.

The Kozachenko-Leonenko method is one of the modern and effective entropy estimations methods [13, 14, 15]. According to this method, entropy of the random vector X can be estimated based on its samples $x_1, x_2, \ldots, x_n$ (where $x_i \in R^d, i=\overline{1,n}$) with the following formula:

$$H_{KLE}=\psi(n)-\psi(k)+\log V_d+\frac{d}{n}\sum_{i=1}^{n}\ln \rho_k(x_i)+\gamma, \quad (1)$$

where $\psi(\cdot)$ is the digamma function; k is the number of nearest neighbors; V_d is the volume of a unit ball in a d-dimensional space, $V_d = \frac{\pi^{d/2}}{\Gamma(d/2+1)}$; $\rho_k(x_i)$ is the distance from x_i to its k-th nearest neighbor; γ is the Euler–Mascheroni constant.

A sharp decrease or increase in local entropy values may indicate the presence of anomalies in the time series. This is because anomalous values are usually characterized by significantly different statistical properties compared to the main sample, which leads to a significant change in the distribution density and, accordingly, KLE values.

The algorithm for detecting anomalous values of the time series is as follows. First, for each element x_i a local KLE estimate is calculated according to the formula:

$$H_{KLE}^{local}(x_i) = \psi(n) - \psi(k) + \log V_d + d \cdot \ln \rho_k(x_i) + \gamma. \quad (2)$$

Next, a threshold value is determined for identifying anomalies based on the empirical distribution of local entropy estimates $H_{KLE}^{local}(x_i)$ is determined (for example, by using the quantile method):

$$H_{threshold} = Q_\alpha\left(H_{KLE}^{local}(x_i)\right), \quad (3)$$

where Q_α is the quantile of the α level, defined as:

$$Q_\alpha(Y) = inf\left\{y \in R : F_Y(y) \geq \alpha\right\} \quad (4)$$

where F_Y is the distribution function of a random variable Y, α is the significance level, selected in accordance with the required sensitivity of the method, $\alpha \in [0,1]$.

The last step is to determine the anomalies, meaning the values for which the condition is met:

$$\left|H_{KLE}^{local}(x_i)\right| > H_{threshold}. \quad (5)$$

The effectiveness of this approach is due to the ability of the Kozachenko-Leonenko method to take into account the structural features of the sample. By taking into account local features and the proximity of points in space, this method is able to effectively distinguish noise and minor fluctuations from true anomalous values, which makes it especially relevant when analyzing complex multivariate data. Thus, the use of KLE estimation improves the accuracy of detecting anomalous values and ensures adequate subsequent mathematical modeling and analysis of the studied time series.

Given that the time series is being studied, it is advisable to supplement the above algorithm with the method of analyzing the deviation in the phase space. The distance from x_i to its k-th nearest neighbor x_i^k in the reconstructed phase space can be determined by the following formula:

$$\rho_k(x_i) = \left(\sum_{l=1}^{L}\left[x_i(t_p + (l-1)\tau) - x_i^k(t_p^k + (l-1)\tau)\right]^2\right)^{\frac{1}{2}}. \quad (6)$$

where t_p, t_p^k are the initial time points for the vector with index i and its k-th nearest neighbor correspondingly; p is the number of the starting point of the phase space reconstruction, $p \in \overline{1,n}$; l is the dimensionality of the reconstructed phase space embedding (chosen by the method of false nearest neighbors); τ is the time delay for phase space reconstruction (chosen by the method of mutual information or autocorrelation).

Formula (6) takes into account the structure of the data in the reconstructed phase space, allowing for more accurate detection of local features and anomalies.

5. Analysis of chaotic properties of anomalies

After identifying anomalous values in time series, an important next step is to examine their chaotic properties to determine the presence or absence of hidden patterns characteristic of complex dynamic systems. The analysis of chaotic properties relies on nonlinear methods from dynamical systems theory, particularly chaos theory, which provides deeper insights into the nature of anomalies and enables their future prediction. Lyapunov exponent analysis is widely studied as a method for assessing the chaotic properties of anomalies in various fields [16, 17]. Although the Grassberger-Procaccia algorithm is less commonly applied in the context of anomalies (making the present study particularly noteworthy), it is actively used to investigate the chaotic properties of variables [18, 19].

For the analysis of the anomaly behavior, it is necessary to apply time-delayed phase space reconstruction methods with the optimal embedding dimension determined by the false nearest neighbors method:

$$X^{(a)}(t)=\left[x_i^{(a)}(t), x_i^{(a)}(t+\tau), x_i^{(a)}(t+2\tau), \ldots, x_i^{(a)}\left(t+\left(l^{(a)}-1\right)\tau\right)\right], \tag{7}$$

where $x_i^{(a)}(t)$ are the abnormal values of the parametric series at time t; $l^{(a)}$ is the optimal dimension of the investment; τ is the optimal time delay calculated by the mutual information method or by the autocorrelation function.

The Lyapunov exponents, which quantitatively characterize the sensitivity of the trajectories of a dynamic system to initial conditions, are calculated by the formula:

$$\lambda_{max}^{(a)}=\lim_{t\to\infty}\frac{1}{t}\ln\frac{\left\|\delta X^{(a)}(t)\right\|}{\left\|\delta X^{(a)}(0)\right\|}, \tag{8}$$

where $\left\|\delta X^{(a)}(0)\right\|$ is the initial distance between two trajectories in phase space at $t=0$; $\left\|\delta X^{(a)}(t)\right\|$ is the distance between two initially close trajectories in phase space at time t, which is determined by the formula:

$$\left\|\delta X^{(a)}(t)\right\|=\left(\sum_{l=1}^{L}\left[x_i^{(a)}\left(t+(l-1)\tau\right)-x_j^{(a)}\left(t+(l-1)\tau\right)\right]^2\right), \tag{9}$$

where $x_j^{(a)}$ is the point representing the nearest neighbor of the point $x_i^{(a)}$ in the reconstructed phase space.

The next step is to determine the fractal dimension of the set of points in phase space using the Grassberger-Procaccia correlation dimension method to assess the presence of hidden geometric structures in the anomalies.

The correlation sum for anomalous values can be determined with the following formula:

$$C^{(a)}(r)=\lim_{N^{(a)}\to\infty}\frac{2}{N^{(a)}\left(N^{(a)}-1\right)}\sum_{i=1}^{N^{(a)}}\sum_{j=i+1}^{N^{(a)}}\Theta\left(r-\left\|X_i^{(a)}-X_j^{(a)}\right\|\right), \tag{10}$$

where Θ is the Heaviside step function; r is the correlation radius; $N^{(a)}$ is the number of anomalous points in the phase space.

The assessment of the presence of structural (geometric) properties of the set of anomalous points $X^{(a)}$ is performed by calculating the correlation dimension value:

$$D^{(a)}=\lim_{r\to\infty}\frac{\ln C^{(a)}(r)}{\ln r}. \tag{11}$$

A comprehensive analysis of the results is carried out as follows:

- if $\lambda_{max}^{(a)} > 0, D^{(a)} \ll l^{(a)}$, then the anomalies are the result of chaotic dynamics and have a hidden structure;
- if $\lambda_{max}^{(a)} \le 0, D^{(a)} \approx l^{(a)}$, it indicates the random nature of the anomalies or the absence of a defined structure.

The use of these methods enables effective analysis of the chaotic properties of anomalous values, determination of their nature, and prediction of future system dynamics, taking into account hidden nonlinear patterns.

6. Attention mechanisms for analyzing anomalies in time series

Attention mechanisms can significantly improve the analysis of anomalies in time series by helping to find complex and hidden relationships between them. They allow considering anomalies not in isolation, but in the context of an overall dynamic process, which can be useful for predicting and interpreting chaotic changes in systems [20-23].

Suppose we have a complete time series of a dynamic system containing both normal values and clearly marked anomalies: $X^{¿} = \left\{ x(t_i) \right\}_{i=1}^{N-l\tau}$, $x(t_i) = \left[x_1, x_2, \ldots, x_i, \ldots, x_{i+\tau}^{(a)}, \ldots, x_{i+(l-1)\tau}^{(a)} \right]$.

Multi-Head Attention is defined as follows:

$$MultiHead(Q, K, V) = Concat(h_1, h_2, \ldots, h_j) W^O, j = \overline{1, h}, \tag{12}$$

where Q, K, V are, respectively, the query, key and value matrices formed from the reconstructed phase space $X^{¿}$; $h_j = softmax\left(\frac{Q W_j^Q \left(K W_j^K \right)^T}{\sqrt{d_k}} \right) V W_j^V$ is the attention head; W^O is a trainable weight matrix that projects the concatenated attention heads back into the model's dimension; W_j^Q, W_j^K, W_j^V are the trainable weight matrices that project input queries, keys, and values into lower-dimensional spaces for each attention head; h is the number of attention heads; d_k is the dimension of the keys, $d_k = d/h$, d is the dimension of the output feature space of the model.

Since the model now takes into account all values, it is important to pay special attention to anomalous points. Losses are determined by the weighted formula:

$$L(W) = \frac{1}{N - l \cdot \tau} \sum_{i=1}^{N - l \cdot \tau} \alpha_i \left\| x(t_i) - \hat{x}(t_i, W) \right\|^2, \tag{13}$$

where $\hat{x}(t_i, W)$ is the prediction of the Multi-Head Attention model; α_i is a weighting coefficient that increases for anomalous points, $\alpha_i = \begin{cases} \beta, if\ x(t_i) \supseteq x_i^{(a)} \wedge \beta > 1, \\ 1, if\ x(t_i) \not\supseteq x_i^{(a)} \vee \beta \le 1 \end{cases}$; β is the level of attention paid to anomalous points.

The resulting softmax attention coefficients are analyzed to determine the significance of each point in the series. Patterns in the anomalies are identified by the coefficients that connect anomalous points to other points (anomalous or normal). High attention values between anomalies and certain previous (or subsequent) points signal hidden causal relationships or patterns, which helps to identify temporal or spatial patterns of anomalies.

The proposed approach, which uses a full time series with clear indexing of anomalies, provides an effective way to detect hidden patterns and predict future anomalous states of complex dynamic systems. The use of the full series enriches the analysis with context and allows you to create detailed and accurate models of anomaly behavior.

7. Model validation for anomaly detection and analysis

7.1. Validation dataset

For experimental validation, the Gas Sensor Array Low-Concentration dataset [24] was used. The dataset consists of multivariate time series recorded from a gas sensor array comprising ten sensors exposed to various volatile organic compounds under controlled laboratory conditions. Each experimental trial contains approximately 9000 temporal measurements sampled at 1 Hz and captures the full sensor response cycle, including baseline stabilization, gas exposure, and recovery phases.

In this study, a subset of the dataset was employed, corresponding to sensor responses to ethanol at a concentration of 100 ppb. The selected subset includes measurements from all ten sensors across all experimental phases, forming a high-dimensional multivariate time series suitable for evaluating anomaly detection and dynamical analysis methods. The resulting data matrix has dimensionality $T \times d$, where $T \approx 9000$ denotes the temporal length and $d=90$ represents the aggregated sensor features. This subset was chosen to ensure controlled experimental conditions while preserving sufficient temporal and structural complexity for validating the proposed entropy-based, dynamical, and attention-driven analysis pipeline.

7.2. Entropy-based detection of anomalous modes

Unlike classical statistical methods, the proposed approach is not based on assumptions about the distribution shape and allows analyzing local state density changes in a high-dimensional feature space. This is fundamentally important in the case of sensory time series characterized by non-stationarity and mode transitions.

Entropy estimation was performed locally for each time step of the test segment based on distances to a fixed number of nearest neighbors in feature space. To avoid the influence of global distribution shifts and ensure the adaptability of the method, sliding thresholds determined by quantile analysis in a moving time window were used. In particular, for each point in time, the thresholds were calculated based on the empirical distribution of local entropies in the previous window of fixed length, which allows automatic consideration of changes in the statistical properties of the signal.

Anomaly detection was performed on a test data segment containing 2,701 time points, which was not used at the parameter tuning stage. This approach corresponds to an unsupervised analysis scenario and prevents information leakage between the training and test samples. Those time points for which the local entropy estimate exceeded the adaptive quantile thresholds were considered anomalous.

The test results shown in 1 indicate that the entropy detection procedure identifies a limited number of time points that form compact groups in time. In total, 47 anomalous points out of 2701 were identified in the test segment, which corresponds to a share of approximately 1.7%. This detection level indicates moderate sensitivity of the method and the absence of excessive reaction to random fluctuations.

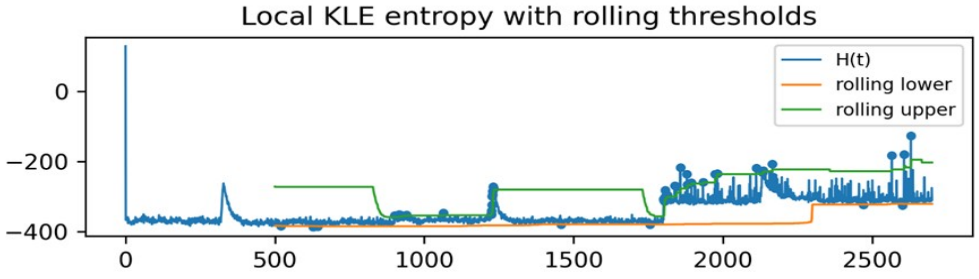


Figure 1: Local entropy detection of anomalous modes.

It is important to note that the detected anomalous points are not distributed evenly throughout the entire time interval. Instead, they are concentrated in relatively narrow time intervals, which is interpreted as a manifestation of regime changes or transitional processes in the dynamics of the sensor array. The absence of isolated single events confirms that the detection procedure responds to structural deviations rather than local noise.

7.3. Formation of anomalous episodes and analysis of their chaotic properties

After completing local entropy-based anomaly detection, the identified anomalous points were aggregated into anomalous episodes, defined as time intervals. Anomalous points separated by less than a fixed temporal threshold were merged into a single temporal episode. For each such episode, an additional symmetric temporal context was incorporated on both sides in order to preserve information about the transition dynamics between normal and anomalous regimes.

The application of the described procedure resulted in 10 continuous anomalous episodes with durations ranging from 57 to 282 time steps, as shown in 1. The detected episodes are characterized by compact temporal localization and the absence of isolated single events.

Table 1
Anomalous episodes formed from the results of anomaly detection

Episode	Time interval	Number of time steps	Episode	Time interval	Number of time steps
ep0	[490-550]	57	ep5	[1428–1488]	57
ep1	[597-673]	73	ep6	[1725–2010]	282
ep2	[871-971]	97	ep7	[2081–2195]	111
ep3	[1033-1093]	57	ep8	[2439–2499]	57
ep4	[1193-1262]	66	ep9	[2535–2659]	121

The formed anomalous episodes were used to analyze their dynamic properties by using invariants of nonlinear dynamics. For each episode, a phase trajectory reconstruction was performed with fixed input parameters, identical for all experiments. The maximum Lyapunov exponent was estimated independently for each episode based on the evolution of distances between close trajectories, which allows quantifying the sensitivity of dynamics to initial conditions.

The results presented in 2 demonstrate significant heterogeneity in the dynamic properties of anomalous episodes. The values of the maximum Lyapunov exponent λ_{max}^{real} vary in the range from approximately 0.004 to 0.18. Two episodes are characterized by consistently high values (close to 0.18). The rest of the episodes remain close to zero, indicating the absence of exponential divergence of trajectories and suggesting the different nature of the detected anomalous modes.

To verify that the positive values of the Lyapunov exponent λ_{max}^{real} were not the result of random fluctuations, a surrogate analysis was performed. For each anomalous episode, time surrogates were constructed by rearranging observations while preserving the distribution of values. For the obtained surrogate realizations, the Lyapunov exponent $\lambda_{max}^{surrogate}$ was evaluated using identical parameters, which ensures a correct comparison with real episodes.

A comparison of the maximum Lyapunov exponents for real and surrogate episodes is shown in 2 and 2. For most episodes, the values of the Lyapunov exponent for real and surrogate episodes are comparable. At the same time, for two anomalous episodes, the real values significantly exceed the corresponding surrogate estimates, indicating the presence of exponential sensitivity to initial conditions. Thus, only a subset of anomalous episodes shows signs of chaotic dynamics, while other deviations are of a different, non-chaotic nature.

Table 2
Estimates of the maximum Lyapunov exponent for anomalous episodes

Episode	Time interval	λ_{max}^{real}	$\lambda_{max}^{surrogate}$
ep0	[490-550]	0.0307	0.01838
ep1	[597-673]	0.0244	0.01401
ep2	[871-971]	0.1834	0.04918
ep3	[1033-1093]	0.0195	0.02434
ep4	[1193-1262]	0.0661	0.03603
ep5	[1428–1488]	0.0039	0.01883
ep6	[1725–2010]	0.1828	0.04482
ep7	[2081–2195]	0.0392	0.06413
ep8	[2439–2499]	0.0253	0.02681
ep9	[2535–2659]	0.0125	0.05868

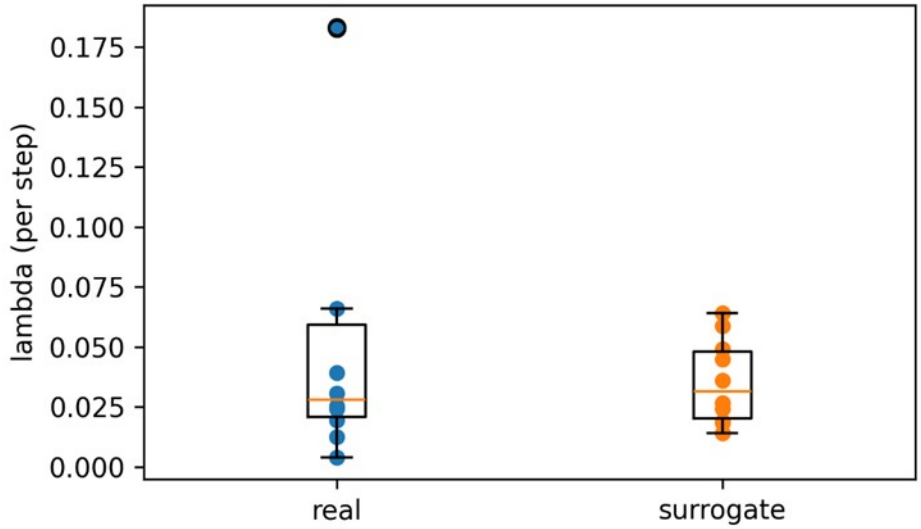


Figure 2: Comparison of Lyapunov exponents for real and surrogate episodes.

To further verify the structural properties of anomalous episodes, the correlation dimension of reconstructed phase trajectories was evaluated. This analysis was used not as the main criterion, but as an auxiliary tool aimed at verifying the consistency of the results obtained using the Lyapunov exponent. The assessment was performed only for selected anomalous episodes, the duration of which was sufficient for correct statistical analysis.

As can be seen from 3, correlation dimension estimates for most anomalous episodes are unstable or degenerate, especially for short trajectories. For individual episodes, the dimension values approach zero for both real and surrogate data, indicating the limited informativeness of this indicator in conditions of high dimension and finite signal length.

Table 3
Correlation dimension estimates for selected real anomalous episodes and their surrogates

Episode	Time interval	D^{real}	$D^{surrogate}$	Comment
ep2	[871-971]	0.0043	0.0	short chaotic episode
ep5	[1428–1488]	NaN	0.0	high-dimensional / finite-size effects
ep6	[1725–2010]	0.0	0.0	insufficient length for GP estimation

These results confirm the advisability of using the Lyapunov exponent as the main indicator of chaotic dynamics.

7.4. Attention-based predictive validation

For additional validation of the structural nature of the detected anomalous modes, an experiment was conducted on short-term prediction of a multidimensional sensory signal using an attention model. The model was trained in one-step prediction mode on a test segment without using anomaly labels. Fixed time windows of reconstructed trajectories were used as input, which allows evaluating the model's ability to reproduce the local dynamics of the system in both normal and anomalous modes.

Figure 3 shows the projection of the reconstructed phase space onto the first two principal components (PC1 as the left axis and PC2 as the bottom axis), which together explain a significant portion of the total data variance and are used exclusively for visual-analytical assessment of the geometric structure of trajectories. The noticeable overlap of normal and abnormal points in this projection indicates the absence of clearly separated clusters associated with abnormal modes. This suggests that anomalies cannot be reliably identified using static projections in feature space and require analysis of temporal dynamics.

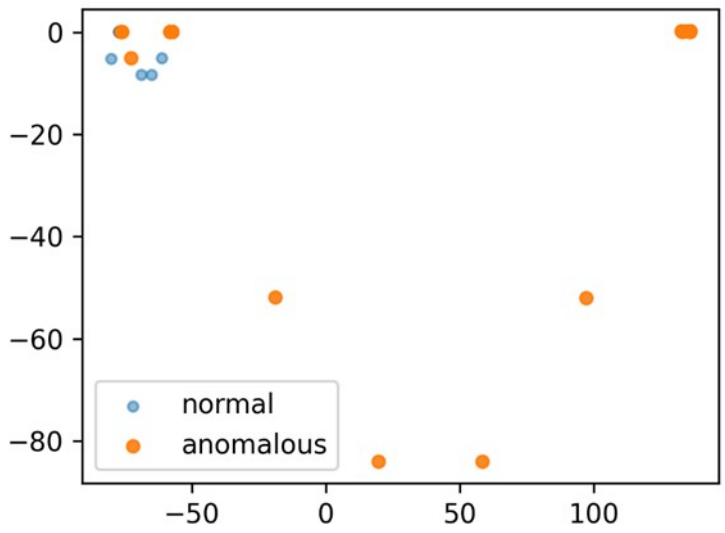


Figure 3: PCA projection of reconstructed phase space for normal and abnormal states.

The quantitative assessment of the predictive complexity of different modes was performed using the root mean square error (RMSE) of the forecast, calculated separately for the entire test segment and for the subset of time steps belonging to anomalous episodes. This division allows us to directly compare the quality of forecasting in normal and anomalous modes and to assess the extent to which the detected anomalies complicate short-term predictions of system dynamics.

The root mean square error (RMSE) over the entire test segment is 0.1866, whereas for anomalous episodes this value increases to 1.3319. Consequently, the prediction error in anomalous regimes exceeds the average level by more than a factor of seven. This observation indicates substantially higher dynamic complexity of anomalous regimes and is consistent with the results of the chaotic properties analysis. The obtained results further confirm that the detected anomalies are structural in nature rather than attributable to noise.

Thus, attention-based forecasting provides an independent empirical validation that the identified anomalous regimes are characterized by increased dynamic complexity.

8. Conclusion

This work proposes and experimentally validates a multilevel computational approach for anomaly detection and analysis in multivariate parametric time series of complex dynamical systems. The approach integrates local entropy-based detection, episodic representation of anomalies, and analysis of their dynamical and predictive properties in an unsupervised setting, which is typical for computer-based analysis of real-world sensor data.

Entropy-based detection using the Kozachenko–Leonenko method with adaptive thresholds enabled stable identification of regime deviations without relying on a priori labels. Within the test segment, 47 anomalous points out of 2,701 were identified and aggregated into 10 anomalous episodes. Subsequent dynamical analysis revealed that only a subset of these episodes exhibits signatures of chaotic dynamics, while others demonstrate non-chaotic yet structurally distinct behavior.

Predictive validation using an attention-based model demonstrated a substantial increase in forecasting error on anomalous episodes (RMSE=1.33 compared to 0.19 on all test data), quantitatively confirming their elevated dynamic complexity. The obtained results demonstrate the effectiveness of the proposed approach as a computational tool for the analysis of nonstationary multivariate signals in computer science and data processing applications.

9. Directions for further research

Further research should be focused on expanding the functionality of the proposed multilevel model by integrating hybrid approaches that combine attention mechanisms with deep recurrent or graph neural networks to model spatial and temporal dependencies in large sensor networks. Particular attention should be paid to the automatic selection of phase space reconstruction parameters (delay, dimensionality), adaptive adjustment of sensitivity thresholds in entropy analysis, and optimization of the interpretation of attention coefficients. In addition, it is promising to apply the developed methods to multicomponent systems with nonlinear feedback or to situations with weak change signaling. It is also advisable to consider online analysis tasks with gradual updating of models in real time. Expanding the testing to other types of data will allow to assess the versatility and stability of the approach.

Acknowledgements

The work was supported by the state budget research project “To develop evolutionary modeling methods for complex systems aimed at analyzing the dynamics of selforganized structures under conditions of uncertainty” (state registration number 0126U000358) of the V.M. Glushkov Institute of Cybernetics of the National Academy of Sciences (NAS) of Ukraine.

Declaration on Generative AI

During the preparation of this work, the authors used DeepL in order to translate research notes and results from Ukrainian to English. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

References

[1] S. Lin, G. Mengaldo, R. Maulik, Online data-driven changepoint detection for high-dimensional dynamical systems, Chaos: An Interdisciplinary Journal of Nonlinear Science 33 (10) (2023) 103112. doi:10.1063/5.0160312.

[2] J. M. Spoor, J. Weber, J. Ovtcharova, A definition of anomalies, measurements, and predictions in dynamical engineering systems for streamlined novelty detection, in: 2022 8th International Conference on Control, Decision and Information Technologies (CoDIT), vol. 1, 2022, pp. 675–680. doi:10.1109/CoDIT55151.2022.9804125.
[3] J. S. Lekscha, Complex systems methods for detecting dynamical anomalies in past climate variability, Doctoral dissertation, Humboldt-Universität zu Berlin, 2020. doi:10.18452/21047.
[4] R. Frehner, K. Wu, A. Sim, J. Kim, K. Stockinger, Detecting anomalies in time series using kernel density approaches, IEEE Access 12 (2024) 33420–33439. doi:10.1109/ACCESS.2024.3371891.
[5] S. Schmidl, P. Wenig, T. Papenbrock, Anomaly detection in time series: A comprehensive evaluation, Proceedings of the VLDB Endowment 15 (9) (2022) 1779–1797. doi:10.14778/3538598.3538602.
[6] A.-R. Al-Ghuwairi, Y. Sharrab, D. Al-Fraihat, A. Al-Elaimat, A. Alsarhan, A. Algarni, Intrusion detection in cloud computing based on time series anomalies utilizing machine learning, Journal of Cloud Computing 12 (2023) 1–17. doi:10.1186/s13677-023-00491-x.
[7] X. Yang, X. Qi, X. Zhou, Deep learning technologies for time series anomaly detection in healthcare: A review, IEEE Access 11 (2023) 117788–117799. doi:10.1109/ACCESS.2023.3325896.
[8] G. Chen, An improved BiGAN model for anomaly detection in finance, Applied and Computational Engineering, 2024. doi:10.54254/2755-2721/53/20241281.
[9] K. Fu, H. Li, MCGnet with multi-level gated unit transformations: A time series prediction-based model for unknown pattern abnormality detection, Expert Systems with Applications 258 (2024) 125088. doi:10.1016/j.eswa.2024.125088.
[10] L. Zhang, T. Yu, J. Shen, Anomaly detection method based on GNN, in: 2024 IEEE 4th International Conference on Data Science and Computer Application (ICDSCA), 2024, pp. 76–80. doi:10.1109/ICDSCA63855.2024.10859844.
[11] K. Da Costa, Anomaly detection in global financial markets with graph neural networks and nonextensive entropy, arXiv preprint, 2023. doi:10.48550/arXiv.2308.02914.
[12] S. Liang, D. Pi, X. Zhang, Anomaly detection model for large-scale industrial systems using transfer entropy and graph attention network, Measurement Science and Technology 35 (9) (2024) 095104. doi:10.1088/1361-6501/ad545e.
[13] W. Xu, S. Zhu, Q. Li, X. Chen, X. Zhou, Star uniform selection algorithm based on maximizing Kozachenko-Leonnko entropy, in: Fourth International Conference on Artificial Intelligence and Electromechanical Automation (AIEA 2023), SPIE Proceedings, vol. 12709, 2023, p. 1270936. doi:10.1117/12.2684884.
[14] D. Lombardi, S. Pant, Nonparametric k-nearest-neighbor entropy estimator, Physical Review E 93 (1) (2016) 013310. doi:10.1103/PhysRevE.93.013310.
[15] O. Palagin, D. Symonov, Formalized model of attitude formation as a tool for analyzing behavioral patterns, Cybernetics and Systems Analysis 61 (2025) 705–713. doi:10.1007/s10559-025-00804-9.
[16] E. Savran, E. Karpat, F. Karpat, Energy-efficient anomaly detection and chaoticity in electric vehicle driving behavior, Sensors 24 (17) (2024) 5628. doi:10.3390/s24175628.
[17] P. Niu, Y. Sun, Z. Gong, Research on the chaotic characteristics and noise reduction prediction of information system anomalies in equipment manufacturing enterprises, Sustainability 13 (9) (2021) 4911. doi:10.3390/su13094911.
[18] L. Ma, X. Liu, X. Liu, Y. Zhang, Y. Qiu, K. Li, On the correlation dimension of discrete fractional chaotic systems, International Journal of Bifurcation and Chaos 30 (12) (2020) 2050174. doi:10.1142/S0218127420501746.
[19] F. Sun, Q. Xue, R. Wang, W. Yue, Detection of low-dimensional chaos in drill bit torsional vibration time series, Journal of Vibroengineering 17 (2015) 516–526.

[20] S. Lunawat, J. Rao, P. Patil, ASBO-MAT-BiLSTM: Social media network anomaly detection using optimized multihead attention transformer-based model, in: 2025 3rd International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT), 2025, pp. 2209–2216. doi:10.1109/IDCIOT64235.2025.10915016.
[21] Y. Qiao, J. Lü, T. Wang, K. Liu, B. Zhang, H. Snoussi, A multihead attention self-supervised representation model for industrial sensors anomaly detection, IEEE Transactions on Industrial Informatics 20 (2) (2024) 2190–2199. doi:10.1109/TII.2023.3280337.
[22] O. Palagin, D. Symonov, T. Semykopna, Mathematical modeling of the evolution of the rehabilitation process for patients with oncological diseases, in: V. Kazymyr et al. (Eds.), Mathematical Modeling and Simulation of Systems. MODS 2023, Lecture Notes in Networks and Systems, vol. 1091, Springer, Cham, 2024, pp. 99–112. doi:10.1007/978-3-031-67348-1_8.
[23] L. Zhou, Q. Zeng, B. Li, Hybrid anomaly detection via multihead dynamic graph attention networks for multivariate time series, IEEE Access 10 (2022) 40967–40978. doi:10.1109/ACCESS.2022.3167640.
[24] F. Tian, L. Zhao, S. Deng, Gas sensor array low-concentration, UCI Machine Learning Repository, 2023. doi:10.24432/C5CK6F.