ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.

A Deep Learning Diagnostic Observer for Time Series Anomaly Detection

Assmaa Alsamadi*, Fannia Pacheco, Paul Honeine.

Univ Rouen Normandie, LITIS UR 4108, F-76000 Rouen, France

Abstract. Mathematical models for anomaly detection (AD) in dynamic systems have demonstrated high performance, particularly diagnostic observer models. Deep learning (DL) models are also effective, but they often require complex architectures, large training datasets and costly fine-tuning to achieve generalization across diverse time series (TS) types. This paper introduces a DL AD model based on the algebraic design of a diagnostic observer, marking its first adaptation for TS data. Experiments on large TS benchmark datasets demonstrate its superiority over various recent DL models.

1 Introduction

Time series (TS) are widely monitored across various applications, such as weather forecasting and fault detection tasks. Anomaly detection (AD) in TS data helps identify issues in advance, supporting decision-making processes such as preventing cyberattacks and enabling predictive maintenance in industrial machinery [1]. Machine learning (ML) models are widely used for AD due to their ability to differentiate between anomalous and normal samples [2]. Deep learning (DL) models, on the other hand, embed complex TS data into suitable representations for AD while effectively capturing sequential dependencies. The majority of DL AD models are designed to generate anomaly-free data [3, 4, 5] or predict future steps based on historical data [6]. For the most recent advancements, foundation models have been developed using large language models for TS forecasting and subsequent AD [7, 8]. In all cases, an anomaly is detected when the reconstruction error (anomaly score) between the predicted sample and the real one exceeds predefined levels. The main drawback of these DL models is their need for per-dataset tuning, large training datasets, or data generation.

Before the rise of DL models, various mathematical frameworks were developed for AD in dynamic systems. One such example is the diagnostic observer (DO) [9]. A DO estimates the system's output and computes a residual error between this output and the real output as illustrated in Fig. 1. The residual is then used as an anomaly score. A DO requires both the system input and output, and also a mathematical model of the system. To design a DO, the most commonly adopted approach is the algebraic formulation introduced by [10]. This paper introduces the first DL-based DO (DL DO) that leverages its algebraic design for AD in univariate TS. Specifically, system modeling is bypassed by learning specific DO's parameters. Finally, since we are working in

*Corresponding author. Email: assmaa.alsamadi@univ-rouen.fr

ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.

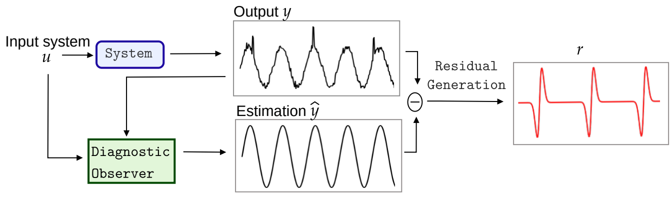

Fig. 1: A dynamical diagnostic observer for fault modeling.

a data-driven scenario where only the TS output of a system is available, we estimate the unknown system input using DL learnable layers.

2 Preliminaries

2.1 Anomaly detection in TS

A TS data y is a finite ordered sequence of l observations: $y(1), y(2), \ldots, y(l)$. AD methods compute an anomaly score for each $y(k)$. If the score exceeds a certain threshold, the observation is labeled as an anomaly; otherwise, it is considered normal. Therefore, AD models aim to find a reliable anomaly score through various approaches.

2.2 Discrete State Space Model

State Space Models (SSMs) in DL have been recently investigated to map an input TS $u \in \mathbb{R}^l$ to an output TS $y \in \mathbb{R}^l$ through a latent variable $x \in \mathbb{R}^{n \times l}$, enabling the modeling of long temporal dependencies. The discrete mapping is denoted as $y(k) = \mathrm{SSM}_{\overline{\mathbf{A}},\overline{\mathbf{B}},\overline{\mathbf{C}},\overline{\mathbf{D}}}(u(k))$ and described as:

$$\mathrm{SSM}_{\overline{\mathbf{A}},\overline{\mathbf{B}},\overline{\mathbf{C}},\overline{\mathbf{D}}} = \begin{cases} x(k) = \overline{\mathbf{A}}x(k-1) + \overline{\mathbf{B}}u(k) \\ y(k) = \overline{\mathbf{C}}x(k) + \overline{\mathbf{D}}u(k) \end{cases} \tag{1}$$

where $\overline{\mathbf{A}} \in \mathbb{R}^{n \times n}$, $\overline{\mathbf{B}} \in \mathbb{R}^{n \times 1}$, $\overline{\mathbf{C}} \in \mathbb{R}^{1 \times n}$, and $\overline{\mathbf{D}} \in \mathbb{R}$. This SSM can be efficiently implemented in a DL framework as in [11].

2.3 Diagnostic Observer

A Diagnostic Observer (DO) maps a pair input-output $(u(k), y(k))$ to a latent state space $z(k) \in \mathbb{R}^s$ of order s, to estimate a residual $r(k) \in \mathbb{R}$ through [10]:

$$\begin{aligned} z(k+1) &= \mathbf{G}z(k) + \mathbf{H}u(k) + \mathbf{L}y(k) \\ r(k) &= vy(k) - wz(k) - qu(k), \end{aligned} \tag{2}$$

where $v \in \mathbb{R}$, $w \in \mathbb{R}^{1 \times s}$, $q \in \mathbb{R}$, $\mathbf{G} \in \mathbb{R}^{s \times s}$, $\mathbf{H} \in \mathbb{R}^{s \times 1}$ and $\mathbf{L} \in \mathbb{R}^{s \times 1}$. When $y(k)$ is contaminated by a fault from the system, a spike at $r(k)$ appears indicating the presence of an anomaly [12].

ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.

2.4 DO's Algebraic Design

The algebraic design for the DO introduced in [10] relies on a parity vector $\mathbf{p} \in \mathbb{R}^s$ to define the matrices ($\mathbf{G}$, $\mathbf{H}$, $\mathbf{L}$) and vectors (v, w, q). For instance, $\mathbf{H}$ is defined as $\mathbf{H}_{i,j} = f_{\mathbf{H}}(\mathbf{p}, i, j)$ where $f_{\mathbf{H}}$ is a function that describes how the values of $\mathbf{p}$ are distributed across rows and columns. A similar procedure is applied to obtain $\mathbf{L}, v$ and q. This design assumes that the dynamic system is given by an SSM, i.e., $y(k) = \text{SSM}_{\overline{\mathbf{A}},\overline{\mathbf{B}},\overline{\mathbf{C}},\overline{\mathbf{D}}}(u(k))$ with $\overline{\mathbf{A}}, \overline{\mathbf{B}}, \overline{\mathbf{C}}, \overline{\mathbf{D}}$ known. The parity vector $\mathbf{p}$ is computed as a null space as follows:

$$\mathbf{p} \begin{bmatrix} \overline{\mathbf{C}} \\ \overline{\mathbf{CA}} \\ \vdots \\ \overline{\mathbf{CA}}^{\mathbf{s}} \end{bmatrix} = \mathbf{0}, \text{ with } \mathbf{p} = \begin{bmatrix} p_0 & p_1 & \cdots & p_s \end{bmatrix} \text{ and } p_i \in \mathbb{R}. \tag{3}$$

The DO offers a mathematical foundation for fault modeling, but requires system modeling and accessible input. Our proposal addresses these challenges.

3 Proposal

We implement the DO in (2) in a DL framework to serve as an anomaly detector, as illustrated in Fig. 2. This framework comprises three main components: (i) an input definition plus a patching block that enable the definition of the unknown input u, (ii) the initialization of the unknown matrices $\overline{\mathbf{A}}, \overline{\mathbf{B}}, \overline{\mathbf{C}}, \overline{\mathbf{D}}$ to produce $\mathbf{p}$ and (iii) the parametrization of the DO model via the algebraic design. In training, the model learns the optimal parameters of the DO and the most accurate reconstruction of the input u. This is accomplished by minimizing the residual to zero using anomaly-free training data. At inference, the absolute value of the residual serves as an anomaly score that depicts high values when anomalies are present.

Input u. In a dynamical context, y is a linear transformation of u as in (1). Motivated by the inversion of this SSM, we aim to recover the input u directly from y through another linear transformation. To achieve this, we first enrich the feature representation of y by applying a Fast Fourier Transform (FFT), allowing the model to adaptively emphasize the most informative frequency components. Next, the FFT is passed through a block of two linear layers followed by a GELU activation and produces the inversion in the frequency domain. The resulting inversion is then modulated by a learnable complex-valued weight. An inverse FFT (IFFT) is applied to bring the signal back to the time domain. Finally, to enable the model to memorize the TS sequence, the signal with window size l is divided into M patches $\{\text{S}_1, \text{S}_2, \ldots, \text{S}_M\}$. Each patch S_i corresponds to a segment of the TS with a predefined patch size m. This yields a patch-based representation of the TS in $\mathbb{R}^{M\times m}$. Each patch is then embedded into another dimension with a linear layer to model temporal dependencies, flattened, and followed by another linear layer to restore the sequence to its original length l, resulting in the desired u ready to be sent to the DO.

ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.

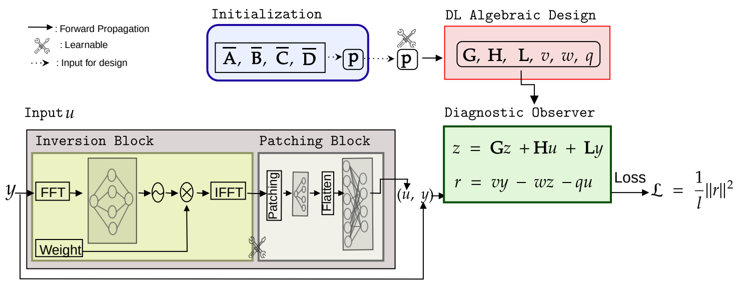

Fig. 2: Our proposed DL DO: two DL blocks reconstruct the input u. The pair (u, y) is then processed by a DO block initialized using the algebraic design. Both block parameters are learned in a DL framework.

Initialization. We randomly initialize the system matrices $\overline{\mathbf{A}}, \overline{\mathbf{B}}, \overline{\mathbf{C}}, \overline{\mathbf{D}}$ to define the prior system. Then we compute the initialization of $\mathbf{p}$ as a null space as in (3) using a Singular Value Decomposition (SVD). Following this initialization, $\mathbf{p}$ is then set as a learnable parameter.

DL Algebraic Design. We compute $\mathbf{G}, \mathbf{H}, \mathbf{L}, v, w, q$ using $\mathbf{p}$ as it follows: $\mathbf{G} = \begin{bmatrix} \mathbf{G}_0 & g \end{bmatrix}$,

$$\mathbf{G}_0 = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 1 & 0 \\ 0 & \cdots & 0 & 1 \end{bmatrix} \quad \text{and} \quad \mathbf{L} = -\begin{bmatrix} p_0 \\ p_1 \\ \vdots \\ p_{s-1} \end{bmatrix} - g p_s, \tag{4}$$

where $\mathbf{G}_0 \in \mathbb{R}^{s\times(s-1)}$ and $g \in \mathbb{R}^{s\times 1}$ is a user-initialized learnable column vector that yields the system stable. Then, $\mathbf{H}$, q, v, and w are obtained as follows: $w = \begin{bmatrix} 0 & \cdots & 0 & 1 \end{bmatrix}$, $v = p_s$ and

$$\begin{bmatrix} \mathbf{H} \\ q \end{bmatrix} = \begin{bmatrix} p_0 + g_1 p_s & p_1 & \cdots & p_{s-1} & p_s \\ p_1 + g_2 p_s & p_2 & \cdots & p_s & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ p_{s-1} + g_s p_s & p_s & \cdots & 0 & 0 \\ p_s & 0 & \cdots & 0 & 0 \end{bmatrix} \begin{bmatrix} \overline{\mathbf{D}} \\ \overline{\mathbf{CB}} \\ \overline{\mathbf{CAB}} \\ \vdots \\ \overline{\mathbf{CA}}^{s-2}\overline{\mathbf{B}} \\ \overline{\mathbf{CA}}^{s-1}\overline{\mathbf{B}} \end{bmatrix}. \tag{5}$$

Finally, after computing the DO's matrices and vectors, the residual r is used as the loss to optimize $\mathbf{p}$ via backpropagation, enabling the model to converge from the initial conditions specified by the algebraic design.

ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.

	DL DO	$MOMENT_0$	$MOMENT_{LP}$	GPT4TS	TimesNet	CrossAD	Anomaly Transformer	DGHL	k-NN
Mean	**0.776**	0.670	0.684	0.611	0.679	*0.750*	0.661	0.646	0.706
Median	**0.767**	0.677	0.692	0.615	0.692	*0.765*	0.658	0.635	0.727
Std.	0.142	0.133	0.146	0.114	0.141	0.158	0.147	0.137	0.155

Table 1: VUS-ROC averaged over 248 TS from the UCR Anomaly Archive, **best performance** in bold and *second best* in italics.

	DL DO	CrossAD	MOMENT
Total Param.	**262,869**	673,315	53,009,160
Total Size (MB)	**5.79**	30.55	793.27

Table 2: Efficiency analysis for top three best AD models.

	DL DO	w/o FFT	w/o Patching
Mean	**0.776**	0.757	0.741
Median	**0.767**	0.751	0.742
Std.	0.142	0.145	0.141

Table 3: Ablation study results for the layers in the block Input u.

4 Experiments

Datasets. We use the UCR Anomaly Archive [13], which contains 248 curated TS datasets from a wide range of domains, including medicine, meteorology, and industry.

Baselines. We primarily compare our model with DL architectures like Anomaly Transformer [3], TimesNet [6], DGHL [5], and CrossAD [4]. Additionally, we include pretrained foundation models such as GPT4TS [8], $Moment_0$ [7], and $Moment_{LP}$ [7]. Lastly, we also consider k-NN, a classic ML model for AD [2].

Experimental Settings. For fair comparison, we follow [7] data train-test split strategy, which downsamples all TS longer than 2560 timesteps by a factor of 10 to speed up the training. Each TS is windowed with a fixed length ($l = 512$). We define our DO model with $s = 5, n = 40, g = -0.1 \times \mathbb{1}_s^T$, $m = 1$. We train it for 120 epochs using the AdamW optimizer with a learning rate of 10^{-3}.

Evaluation metrics. We evaluate our model with the recently proposed Volume Under the Surface (VUS-ROC). The VUS-ROC is parameter-free, threshold-independent, and robust to lags, and noise [14]. It is regarded as one of the most reliable accuracy measures for evaluating both point-based and range-based anomalies.

Results. Experiments demonstrate that our model surpasses the baseline models. It consistently ranks first in both mean and median values across the 248 datasets, as given in Table 1. Moreover, efficiency analysis of the top three models reveals that our model outperforms the others in terms of parameters, with a significantly small model size, as given in Table 2.

Ablation Study. We analyze the contributions of the components in our model's architecture by: i) performing inversion in the frequency domain, which is the proposed method (denoted DL DO), ii) removing the FFT (denoted w/o FFT) and iii) removing the patching block (denoted w/o Patching). The ablation study given in Table 3 demonstrates the significance of both frequency domain inversion and the patching block for effective AD.

ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.

5 Conclusion

This paper introduces DL DO, the first DL approach of a DO for TS AD. We propose an input reconstruction and design parameters that lead to an accurate AD model. Our DL DO outperforms state-of-the-art models in univariate TS AD, demonstrating superior efficiency in model size and parameters. Future work will focus on scaling this approach to multivariate TS, designing deeper DO architectures, and exploring alternative parameter designs.

Acknowledgments and reproducibility: This work has been funded by the ANR-23-CE23-0004 project.
The code is available at https://github.com/AssmaaSamadi/TSDLAD.

References

[1] Philippe Esling and Carlos Agon. Time-series data mining. *ACM Comput. Surv.*, 45(1), 2012.

[2] Sridhar Ramaswamy, Rajeev Rastogi, and Kyuseok Shim. Efficient algorithms for mining outliers from large data sets. In *Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data*, pages 427–438, 2000.

[3] Jiehui Xu, Haixu Wu, Jianmin Wang, and Mingsheng Long. Anomaly transformer: Time series anomaly detection with association discrepancy. In *ICLR*, 2022.

[4] Beibu Li, Qichao Shentu, Yang Shu, Hui Zhang, Ming Li, Ning Jin, Bin Yang, and Chenjuan Guo. CrossAD: Time series anomaly detection with cross-scale associations and cross-window modeling. In *NeurIPS*, 2025.

[5] Cristian Ignacio Challu, Peihong Jiang, Ying Nian Wu, and Laurent Callot. Deep generative model with hierarchical latent factors for timeseries anomaly detection. In *NeurIPS Workshop on Deep Generative Models and Downstream Applications*, 2021.

[6] Haixu Wu, Tengge Hu, Yong Liu, Hang Zhou, Jianmin Wang, and Mingsheng Long. Timesnet: Temporal 2d-variation modeling for general time series analysis. In *ICLR*, 2023.

[7] Mononito Goswami, Konrad Szafer, Arjun Choudhry, Yifu Cai, Shuo Li, and Artur Dubrawski. MOMENT: A family of open time-series foundation models. In *ICML*, 2024.

[8] Tian Zhou, Peisong Niu, Xue Wang, Liang Sun, and Rong Jin. One fits all: Power general time series analysis by pretrained LM. In *NeurIPS*, 2023.

[9] Yu Ding, Darek Ceglarek, and Jianjun Shi. Fault diagnosis of multistage manufacturing processes by using state space approach. *Journal of Manufacturing Science and Engineering*, Vol.124(No.2):313–322, 2002.

[10] S. X. Ding, S. Yin, Y. Wang, Y. Wang, Y. Yang, and B. Ni. Data-driven design of observers and its applications. *IFAC Proceedings Volumes*, 44(1):11441–11446, 2011. 18th IFAC World Congress.

[11] Albert Gu, Karan Goel, and Christopher Re. Efficiently modeling long sequences with structured state spaces. In *ICLR*, 2022.

[12] Steven X. Ding. *Model-Based Fault Diagnosis Techniques: Design Schemes, Algorithms, and Tools*, volume 381 of *Advances in Industrial Control*. Springer, 2008.

[13] Renjie Wu and Eamonn J. Keogh. Current time series anomaly detection benchmarks are flawed and are creating the illusion of progress. In *2022 IEEE 38th International Conference on Data Engineering*, pages 1479–1480, 2022.

[14] J. Paparrizos, P. Boniol, T. Palpanas, R. S. Tsay, A. Elmore, and M. J. Franklin. Volume under the surface: A new accuracy evaluation measure for time-series anomaly detection. *Proceedings of the VLDB Endowment*, 15(11):2774–2787, 2022.