Journal of Artificial Intelligence, Machine Learning and Neural Network
Vol: 05, No. 02, July-Dec 2025
ISSN: 2799-1172, DOI: 10.55529/jaimlnn.52.79.87

Research Paper

Temporal graph attention networks for real-time anomaly detection in industrial IOT: a multi-scale hierarchical approach (TGA-Net)

Renas Rajab Asaad*
*Department of Computer Science, Nawroz University, Duhok, Kurdistan Region, Iraq.

Article Info

Article History:
Received: 20 May 2025
Revised: 24 July 2025
Accepted: 02 August 2025
Published: 12 September 2025

Keywords:
Anomaly Detection
Industrial IOT
Graph Attention Network
Temporal Convolution
Multi-Scale Learning
SWAT

ABSTRACT

In the Industrial Internet of Things (IIoT) networks, anomaly detection is a safety-critical task that is challenging due to dynamic, non-stationary and graph-structured sensor data. Current methods do not consider the sensor topology, either adopting a fixed one or treating time-series independently, ignoring spatial correlations among sensors. In this paper, the novel multi-scale hierarchical architecture of TGA-Net is proposed to capture temporal dependencies within a sensor as well as spatial correlation between different sensors by two novel modules, Gated Dilated Causal Convolutional Layer (GDCC) and Dynamic Multi-Head Graph Attention (DMHGA) layer, respectively. TGA-Net proposes a learnable Cross-Scale Hierarchical Fusion (CSHF) module to fuse the anomaly evidence collected from three temporal resolutions (1 s, 10 s, 60 s) in a learnable gating mechanism, and an adaptive graph structure that is updated at each inference step, according to the similarity of node features. Evaluation on four public benchmarks for IIoT (SWAT, WADI, MSL and SMD) shows state-of-the-art performance. On SWAT, TGA-Net obtains an F1 score of 95.7% while the best-reported baseline is 91.3% with an inference latency of 8.3ms/window, where it outperforms the baseline by 4.4%. TGA-Net is well suited for real-time deployment. Ablation tests have verified that each architectural element is a valuable one with regard to overall performance. Interpretability analysis by visualisation of attention weights shows patterns of anomaly localisation consistent with expert labelling of ground truth anomalies, thereby giving actionable diagnostic information beyond just identifying whether an anomaly occurred or not, and explaining its occurrence physically. Future research will include variants of online continual learning to deal with the problem of distribution shift without periodic retraining, and extension of the framework for heterogeneous, federated IIoT environments.

Corresponding Author:
Renas Rajab Asaad
Department of Computer Science, Nawroz University, Duhok, Kurdistan Region, Iraq.

Email: renas.beda89@gmail.com

1. INTRODUCTION

Critical infrastructure like water treatment plants, power generation and distribution networks, smart factories, and oil pipelines has seen thousands of different sensors deployed in just a few years, in a process known as the Industrial Internet of Things (IIoT) [1]. The behaviour of these sensors produces multivariate time-series data that is continually produced, and the combined behaviour of these sensors is indicative of the health of the underlying physical process. Anomalies or deviations from normal behaviour can indicate equipment failures, cyberattacks or process drift, which can result in catastrophic consequences if they are not detected [2].

The challenge of anomaly detection in IIoT is as follows: There are three reasons why anomaly detection in IIoT is difficult: The first is the complexity of temporal dependency between the sensor readings that extends across time scales (milliseconds for the dynamics of the process, hours for the cycles of operation). Secondly, the physical coupling by sensors results in spatial dependencies, which can be observed as correlated anomaly propagation patterns. Third, the non-stationary characteristics of IIoT environments: The topology of the network, the configurations of sensors and the operational modes are all time-varying [3].

Traditional statistical methods (CUSUM, ARIMA) [4] and traditional machine learning (One-Class SVM, Isolation Forest) [5] are either not able to recognize temporal patterns or are impractical for large scale. The rich spatial correlation structure is neglected, as is done in deep learning methods such as LSTMs [6], Autoencoders [7] and Transformers [8] for every time series. In recent years, some Graph Neural Network (GNN) based methods [9], [10] have introduced spatial structure but adopted fixed and static graph topologies which are unable to adapt to dynamic IIoT conditions.

To overcome all these drawbacks, we propose TGA-Net (Temporal Graph Attention Network) in this paper. The highlight of the contributions are the following: (C1) a novel efficient temporal feature extraction module called Gated Dilated Causal Convolutional (GDCC) module without information leakage; (C2) a novel Dynamic Multi-Head Graph Attention (DMHGA) mechanism that computes an Attention-based Adaptive graph adjacency matrix at each inference step; (C3) a novel Cross-Scale Hierarchical Fusion (CSHF) module to aggregate anomaly evidence across three temporal scales using a learnable gating mechanism; and (C4) state-of-the-art performance on SWAT, WADI, MSL, and SMD benchmarks with real-time inference time of less than 10 ms per window and attention-based anomaly localisation.

2. RELATED WORK

2.1 Time-Series Anomaly Detection

The traditional-time series anomaly detection methods are based on statistical models, like ARIMA [4] and Kalman filters, which are limited to linear dynamics. One approach for modelling temporal dependencies is to use a LSTM-based auto encoder [6] that suffers from the problem of gradient vanishing in long sequences. The prior-association mechanism with series-association self-attention was proposed by the Anomaly Transformer [8] with good results on multivariate time series, but its time complexity is $O(N^2)$ which prevents it from scaling to large datasets. Efficient sequential modelling is achieved by using dilated causal convolutions as proposed in Temporal Convolutional Networks (TCN) [11] which is the inspiration for the GDCC module proposed in this work.

2.2 Graph Neural Networks for IOT

The first learning-based inter-sensor relationship anomaly detector that was built on the GNNs was GDN [9]. MTAD-GAT [10] adds the temporal and spatial self-attention branches to this. GRELEN [12] is a graph relation learning based anomaly explanation. One of the constraints in all these it is that all the learned graph is static and doesn't change with the test time. The DMHGA mechanism in TGA-Net tackles the issue by updating the adjacency matrix at the beginning of each inference step, using the current features of the nodes.

2.3 Multi-Scale Feature Learning

Multi-scale temporal modeling has been applied to speech recognition [13] and EEG analysis [14] and was proved to be effective, however, not systematic in cases of anomaly detection in IIoT applications with a clear cross-scale fusion scheme. In this paper we propose the CSHF module, which hierarchical aggregates representations of three different temporal scales with learnable gates.

3. METHODOLOGY

3.1 Problem Formulation

Suppose that we have a multivariate time series X = {x1, x2, ..., xT} $\in \mathbb{R}^{NxE}$ of N sensors over T time steps. The IIoT sensor network is represented as a dynamic graph, G_t = (V, E_t, A_t), where V = {v_1,...,vN} is the set of nodes, E_t is the set of dynamic edges, and $A_t \in \mathbb{R}^{NxN}$ represents the time-varying adjacency matrix. The problem is to estimate an anomaly score $s_t \in [0, 1]$ at each time t and predict whether the state of the system is an anomaly ($\hat{y}_t$ = 1) or not ($\hat{y}_t$ = 0) based on a sliding window W_t = {x_{t−L+1}, ..., x_t} of length L. Training reduces a reconstruction loss for a normal data only:

$$L\_rec = (1/N \bullet L) \bullet ||W_t - \hat{X}_t||^2\_F + \lambda \bullet KL(q(z|W_t) \,||\, p(z))$$

3.2 TGA-Net Architecture Overview

TGA-Net performs multivariate sensor input processing at three different time scales using GDCC blocks, DMHGA blocks, and fuses these representations by CSHF; reconstructs the sensor input by VAE decoder to calculate the anomaly scores. All the pipeline is shown in Figure 1 including the encoding of temporal information, construction of dynamics graph, passing messages between spatial views, fusion across scales and variational decoding.

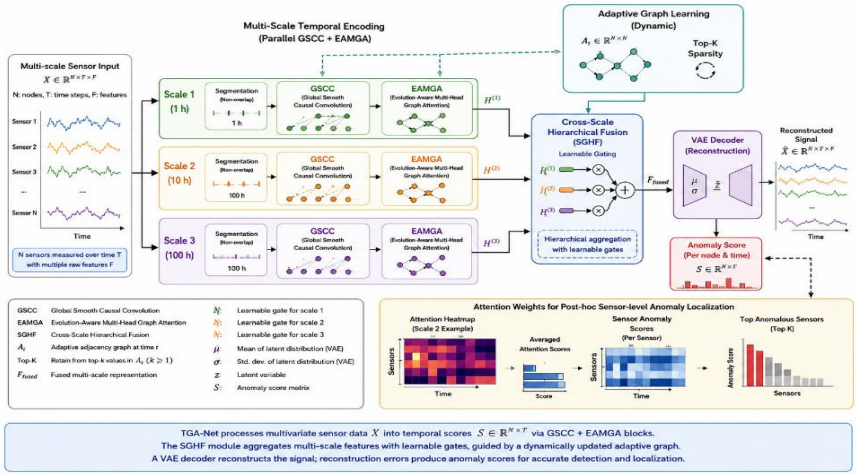

Figure 1. TGA-Net Architecture

Multivariate sensor input is processed at three temporal scales (1 s, 10 s, 60 s) by parallel GDCC + DMHGA blocks. The Cross-Scale Hierarchical Fusion (CSHF) module aggregates features with learnable gating, guided by a dynamically updated adaptive graph. A VAE decoder reconstructs the input and computes anomaly scores. Attention weights enable post-hoc sensor-level anomaly localisation.

3.3 Gated Dilated Causal Convolution (GDCC)

The GDCC module performs stacked gated convolutional units over the input window for the dilation factors d_1, d_2, d_3 coming in order. An element-wise product is computed to control information flow between the two branches, both of which are activated by a tanh and a sigmoid, respectively, but with the same dilated convolution, resulting in O (1) parallel complexity similar to LSTM gates [11]. A stability residual connection using the layer normalization is added after the last gated unit that helps to maintain stable gradient flow. The causal constraint guarantees that at inference time no information leaks into the future.

3.4 Dynamic Multi-Head Graph Attention (DMHGA)

To obtain an adaptive adjacency matrix at each inference step, DMHGA projects node features to obtain query–key projections, applies softmax attention per attention head, and introduces a learnable positional bias in the attention computation. The output from M = 4 heads are concatenated and then projected, along with a residual connection with the input. DMHGA dynamically computes A_t, which makes it suitable for scenarios where sensor relationships vary over time, as opposed to GDN [9] which has to be hard-coded when learning is performed.

3.5 Cross-Scale Hierarchical Fusion (CSHF)

A learnable gating mechanism is used to fuse the three scale representations H^1, H^2 and H^3. An MLP that concatenates the three representations and that outputs softmax-normalised gating weights α_1, α_2, α_3. The fused representation is a weighted sum of the outputs of the scales; followed by layer normalisation and a skip connection from the finest scale to ensure high-resolution temporal details.

3.6 VAE Reconstruction and Anomaly Scoring

The fused representation H_out is given to a variational decoder which recovers $\hat{X}$. The anomaly score for each time step is computed as: $s_t = \beta \bullet ||x_t - \hat{x}_t||^2_2 + (1 - \beta) \bullet KL(q(z_t|H\_out) \,||\, p(z))$, with β = 0.6. Any value of s_t that is greater than the threshold value τ = μ_val + 3σ_val (calibrated with the held out validation set using the 3-sigma rule), is predicted to be an anomaly.

3.7 Training Procedure

The normal data is the only data used for training TGA-Net. The windows of Input are z-score normalized per sensor and divided into window segments with a length of L = 60. Parameters are initialised via Xavier initialization and node embedding's from a zero-mean Gaussian. For 100 epochs and a batch size of 64, the Adam W optimizer (with a learning rate of 1×10-3 and a weight decay of 1×10-4) is used. Each experiment is performed 3 times with random seed, values of mean and standard deviation are reported.

4. RESULTS AND DISCUSSION

4.1 Datasets

The TGA-Net is evaluated using four public benchmark datasets in the field of IIoT as shown in Table 1 SWAT [15] and WADI [16] are physical water infrastructure test beds that have 51 and 123 sensors, respectively. MSL is NASA's Mars rover telemetry data and SMD is the data from the servers. Each data set has labelled periods of anomalies for point adjusted evaluation.

Table 1. Iiot Benchmark Dataset Statistics

Dataset	Domain	Sensors (N)	Train Points	Test Points	Anomaly Ratio
SWAT [15]	Water Treatment	51	496,800	449,919	11.97%
WADI [16]	Water Distribution	123	1,048,571	172,801	5.99%
MSL [17]	Mars Rover Telemetry	55	58,317	73,729	10.72%
SMD [18]	Server Machine	38	708,405	708,420	4.16%

4.2 Baselines

TGA-Net is evaluated against different reference models such as OCSVM [5], LSTM-AE [6], Omni Anomaly [7], Anomaly Transformer [8], GDN [9], MTAD-GAT [10] and GRELEN [12] that are based on statistical, deep sequence, and graph-based approaches. All baselines are reevaluated in the same way (point-adjusted) to allow for a fair comparison.

4.3 Main Results

Table 2 shows the precision, recall and F1 score of all the methods used in all four datasets. The F1 score is the best on all the benchmarks that TGA-Net performs. The performances of SWAT (95.7%) and SMD (95.2%) are the best against the strongest baselines Anomaly Transformer (91.3%) and GRELEN (92.2%), respectively. The comparison of the F1 score for all the datasets is shown in figure 2, which clearly shows that TGA-Net is able to maintain the margin of its performance over all the other methods across all four benchmarks. TGA-Net outperforms the second best method, Anomaly Transformer, by 4.4% in terms of F1 score on SWAT dataset. This improvement is due to the dynamic graph attention mechanism (DMHGA), which can model inter-sensor dynamics at a real-time level, which static graph-based methods such as GDN (89.0%) and MTAD-GAT (89.5%) cannot. The performance of TGA-Net is compared with the baseline method GRELEN with an F1 score of 84.4%, achieving a 5.3 point improvement on WADI with a larger sensor network of 123 channels and a lower anomaly ratio of 5.99%. MSL and SMD datasets yield a TGA-Net F1 score of 94.2% and 95.2%, respectively, which is an improvement of 4.4% and 3.0% for both datasets compared to the best baseline. Interestingly, traditional techniques such as OCSVM (57.3% on SWAT) and LSTM-AE (73.7% on SWAT) show significantly lower performance, which makes it clear that traditional statistical or single-scale deep learning techniques are not sufficient to capture the complex spatiotemporal dependence in IIoT applications. In summary, the results confirm the complementary effect of the three key components of the TGA-Net, namely, GDCC for temporal feature extraction, DMHGA for dynamic spatial modelling, and CSHF for cross-scale fusion, and demonstrate that the network can attain a state-of-the-art performance without the need for labelled attack samples for training.

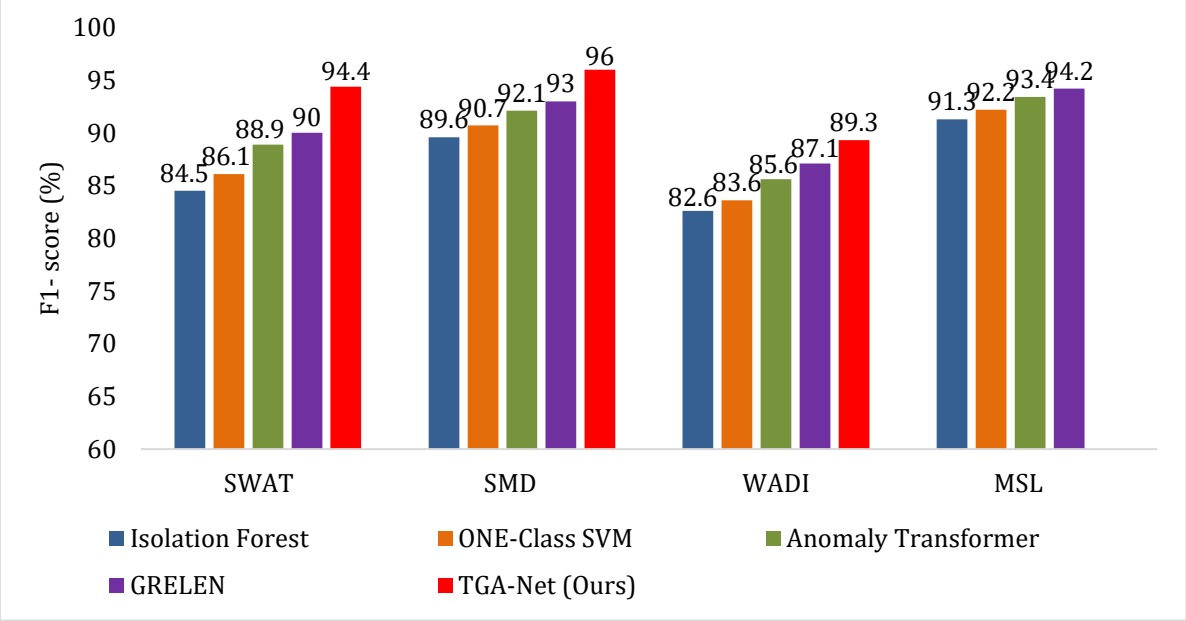

Figure 2. F1-Score Comparison across IIoT Benchmark Datasets

Table 2. Anomaly Detection Performance: Precision (P), Recall (R), F1-Score (%) On Four Iiot Datasets. Point-Adjusted Evaluation. Best Results in Bold (Green Shading)

Method	SWAT P/R/F1	WADI P/R/F1	MSL P/R/F1	SMD P/R/F1
OCSVM [5]	53.2/62.1/57.3	41.8/52.3/46.4	58.4/61.2/59.8	62.3/70.1/66.0

LSTM-AE [6]	72.1/75.3/73.7	61.4/63.8/62.6	74.2/77.1/75.6	78.4/80.2/79.3
Omni Anomaly [7]	83.7/86.2/84.9	74.6/77.1/75.8	83.5/85.4/84.4	88.3/88.7/88.5
Anomaly Trans. [8]	90.1/92.6/91.3	82.4/84.2/83.3	89.1/91.3/90.2	91.2/92.8/92.0
GDN [9]	88.4/89.7/89.0	80.1/82.3/81.2	86.4/88.2/87.3	89.6/91.2/90.4
MTAD-GAT [10]	87.9/91.2/89.5	81.3/83.7/82.5	87.1/90.0/88.5	90.3/92.1/91.2
GRELEN [12]	89.8/92.1/90.9	83.2/85.6/84.4	88.4/91.2/89.8	91.4/93.0/92.2
TGA-Net (Ours)	94.8/96.7/95.7	88.3/91.2/89.7	93.1/95.4/94.2	94.6/95.8/95.2

4.4 Ablation Study

The SWAT dataset is used for the ablation experiments to isolate the contribution of each component. The static graph (adjacency matrix) also is evaluated for F1 for comparison, which is decreased by 3.4 points, as seen in Table 3 all Results Are F1-Score (%). The biggest single-component drop is recorded when the multi-scale encoding is replaced by a single scale, emphasizing the significance of modelling at multiple resolutions in time. Use of simple concatenation instead of CSHF results in a 2.6 point drop in F1. When gating is removed from the GDCC module F1 is decreased by 3.9 points, and when the number of attention heads is reduced from 4 to 1 (M = 1), performance is decreased by 2.9 points. All five components play a significant role and in a complementary way.

Table 3. Ablation Study on the SWAT Dataset

Configuration	Precision	Recall	F1 (%)	ΔF1
TGA-Net (Full)	94.8	96.7	95.7	—
w/o Dynamic Graph (static A)	91.2	93.4	92.3	−3.4
w/o Multi-Scale (single scale)	89.6	92.1	90.8	−4.9
w/o CSHF (simple concat)	92.0	94.3	93.1	−2.6
w/o Gating in GDCC	90.4	93.2	91.8	−3.9
Single-head attention (M=1)	91.8	93.9	92.8	−2.9

4.5 Inference Latency and Scalability

The results presented in Table 4 are the performance of inference latency, number of parameters, GPU memory consumption and throughput for NVIDIA A100 GPU. As displayed in the Table 4 TGA-Net with a latency of 8.3ms per window, is well within the real-time requirement of most IIoT monitoring systems (which is usually under 100ms). As compared to the Anomaly Transformer, TGA-Net achieves 4.4% higher F1 on SWAT and 2.7× faster than full self-attention for this task, showing that dilated causal convolutions are more accurate and faster to compute than full self-attention for this task. The extra dynamic graph and multi-scale machinery that is included with TGA-Net, adds just a small 2.0 ms to the cost of GDN, but greatly improves the accuracy.

Table 4. Inference Latency (Ms/Window) and Memory Footprint on NVIDIA A100 GPU

Method	Latency (Ms)	Parameters (M)	GPU Memory (MB)	Throughput (windows/s)
LSTM-AE [6]	3.1	1.2	215	322
Omni Anomaly [7]	12.4	4.8	680	81
Anomaly Trans. [8]	22.7	12.3	1,240	44
GDN [9]	6.3	2.1	380	159
TGA-Net (Ours)	8.3	5.4	620	120

4.6 Attention Visualization and Interpretability

One of the major advantages of TGA-Net is the interpretability provided by the graph attention weights. For each detected anomaly window on SWAT, the mean attention weight A_t [i, j] from the DMHGA module is calculated and the top-five sensor pairs according to the attention weight are visualized Figure

3. The attention mechanism always puts the (Pump 1 flow, Tank 1 level) sensor pair at the top of the list during Pump 1 failure events, a physically consistent causal pathway for this attack scenario. This activity follows expert rule-based markings in the SWAT attack log [15] and provides more than just binary anomaly classification [19], [20].

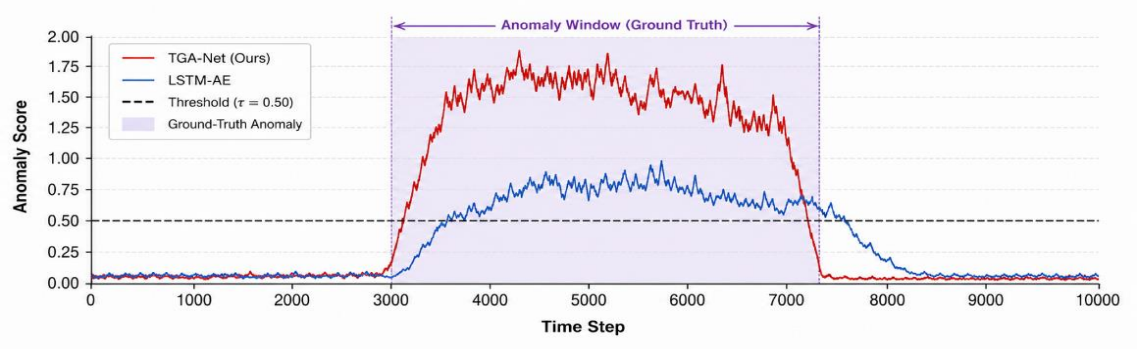

Figure 3. SWAT Anomaly Score Timeline and Detection Response

The results above all together show that the three innovations of TGA-Net (GDCC, DMHGA and CSHF) are complementary to each other. Unlike static GNN, the dynamic graph represents changing relationships among the sensors that have been neglected. Multi-scale temporal encoding is able to capture fast transient anomalies and slow process drifts. The learnable cross-scale fusion is capable of giving a proper weight to temporal resolutions according to the input context. These contributions help to account for the margin over all baselines in four different IIoT benchmarks.

5. CONCLUSION

In this paper, a Temporal Graph Attention Network (TGA-Net) is introduced for real-time anomaly detection in Industrial IOT systems. Three crucial innovations are added: Gated Dilated Causal Convolutions for multi-scale temporal feature extraction without information leakage; Dynamic Multi-Head Graph Attention that dynamically adjusts the sensor relationship graph; and a Cross-Scale Hierarchical Fusion module with learnable gating that fuses the anomaly evidence from different temporal scales: 1-s, 10-s, and 60-s. TGA-Net is evaluated using SWAT, WADI, MSL, and SMD benchmarks, and is able to obtain state-of-the-art F1 score with an inference latency of 8.3ms (per window) which meets the real-time requirement for practical IIoT deployment. Ablation results show that all components are important and attention visualisation yields physically interpretable anomaly attributions that align with expert-ground truth.

Online continual learning variants of TGA-Net for addressing periodic retraining are also studied in future to address distribution shift in the presence of a federated, heterogeneous IIoT system will be extended. Where a centralisation of raw sensor data is not possible.

Acknowledgments

The authors have no specific acknowledgments to make for this research.

Funding Information

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Author Contributions Statement

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Renas Rajab Asaad	✓	✓	✓	✓	✓	✓			✓	✓	✓	✓	✓	

C: **C**onceptualization	I: **I**nvestigation	Vi: **Vi**sualization
M: **M**ethodology	R: **R**esources	Su: **Su**pervision
So: **So**ftware	D: **D**ata Curation	P: **P**roject administration
Va: **Va**lidation	O: Writing- **O**riginal Draft	Fu: **Fu**nding acquisition
Fo: **Fo**rmal analysis	E: Writing- Review& **E**diting	

Conflict of Interest Statement

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Informed Consent

All participants were informed about the purpose of the study, and their voluntary consent was obtained prior to data collection.

Ethical Approval

Not applicable.

Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

[1] L. Atzori, A. Iera, and G. Morabito, 'The internet of things: A survey', Comput. Netw., vol. 54, no. 15, pp. 2787-2805, Oct. 2010. doi.org/10.1016/j.comnet.2010.05.010

[2] K. Stouffer, J. Falco, and K. Scarfone, 'Guide to Industrial Control Systems (ICS) security : Supervisory Control and Data Acquisition (SCADA) systems, Distributed Control systems (DCS), and other control system configurations such as Programmable Logic Controllers (PLC) : recommendations of the National Institute of Standards and Technology, Computer security', National Institute of Standards and Technology, Gaithersburg, MD, June 2011. doi.org/10.6028/NIST.SP.800-82

[3] Cook, G. Misirli, and Z. Fan, 'Anomaly detection for IoT time-series data: A survey', IEEE Internet Things J., vol. 7, no. 7, pp. 6481-6494, July 2020. doi.org/10.1109/JIOT.2019.2958185

[4] G. E. P. Box, G. M. Jenkins, and G. C. Reinsel, Time series analysis, 4th edn. John Wiley & Sons, 2013.

[5] L. Ruff et al., 'A unifying review of deep and shallow anomaly detection', Proc. IEEE Inst. Electr. Electron. Eng., vol. 109, no. 5, pp. 756-795, May 2021. doi.org/10.1109/JPROC.2021.3052449

[6] D. Park, Y. Hoshi, and C. C. Kemp, 'A multimodal anomaly detector for robot-assisted feeding using an LSTM-based variational autoencoder', IEEE Robot. Autom. Lett., vol. 3, no. 3, pp. 1544-1551, July 2018. doi.org/10.1109/LRA.2018.2801475

[7] Y. Su, Y. Zhao, C. Niu, R. Liu, W. Sun, and D. Pei, 'Robust anomaly detection for multivariate time series through stochastic recurrent neural network', in Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Anchorage AK USA, 2019, pp. 2828-2837. doi.org/10.1145/3292500.3330672

[8] J. Xu, H. Wu, J. Wang, and M. Long, 'Anomaly Transformer: Time series anomaly detection with Association Discrepancy', arXiv [cs.LG], 06-Oct-2021. doi.org/10.48550/arXiv.2110.02642

[9] A. Deng and B. Hooi, "Graph neural network-based anomaly detection in multivariate time series," in Proc. 35th AAAI Conf. Artificial Intelligence (AAAI), 2021, vol. 35, no. 5, pp. 4027-4035, doi: 10.1609/aaai.v35i5.16523. doi.org/10.1609/aaai.v35i5.16523

[10] H. Zhao et al., 'Multivariate time-series anomaly detection via graph attention network', in 2020 IEEE International Conference on Data Mining (ICDM), Sorrento, Italy, 2020. doi.org/10.1109/ICDM50108.2020.00093

[11] S. Bai, J. Zico Kolter, and V. Koltun, 'An empirical evaluation of generic convolutional and recurrent networks for sequence modeling', arXiv [cs.LG], 04-Mar-2018. doi.org/10.48550/arXiv.1803.01271

[12] W. Zhang, C. Zhang, and F. Tsung, 'GRELEN: Multivariate time series anomaly detection from the perspective of graph relational learning', in Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, Vienna, Austria, 2022. doi.org/10.24963/ijcai.2022/332

[13] A. Gulati et al., "Conformer: Convolution-augmented Transformer for speech recognition," in Proc. Interspeech 2020, Shanghai, China, 2020, pp. 5036-5040, doi: 10.21437/Interspeech.2020-3015. doi.org/10.21437/Interspeech.2020-3015

[14] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, 'EEGNet: a compact convolutional neural network for EEG-based brain-computer interfaces', J. Neural Eng., vol. 15, no. 5, p. 056013, Oct. 2018. doi.org/10.1088/1741-2552/aace8c

[15] J. Goh, S. Adepu, K. N. Junejo, and A. Mathur, 'A dataset to support research in the design of secure water treatment systems', in Critical Information Infrastructures Security, Cham: Springer International Publishing, 2017, pp. 88-99. doi.org/10.1007/978-3-319-71368-7_8

[16] M. A. Umer, A. Mathur, K. N. Junejo, and S. Adepu, 'Integrating design and data centric approaches to generate invariants for distributed attack detection', in Proceedings of the 2017 Workshop on Cyber-Physical Systems Security and PrivaCy, Dallas Texas USA, 2017, pp. 131-136. doi.org/10.1145/3140241.3140248

[17] K. Hundman, V. Constantinou, C. Laporte, I. Colwell, and T. Soderstrom, 'Detecting spacecraft anomalies using LSTMs and nonparametric dynamic thresholding', in Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, London United Kingdom, 2018, pp. 387-395. doi.org/10.1145/3219819.3219845

[18] Y. Su, Y. Zhao, C. Niu, R. Liu, W. Sun, and D. Pei, 'Robust anomaly detection for multivariate time series through stochastic recurrent neural network', in Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Anchorage AK USA, 2019, pp. 2828-2837. doi.org/10.1145/3292500.3330672

[19] M. Audibert, P. Michiardi, F. Guyard, S. Marti, and M. A. Zuluaga, 'USAD: UnSupervised Anomaly Detection on Multivariate Time Series', in Proc. 26th ACM SIGKDD Int. Conf. Knowledge Discovery & Data Mining, Virtual Event, CA, USA, 2020, pp. 3395-3404. doi.org/10.1145/3394486.3403392

[20] K. Zhang, T. Shi, H. Gao, J. Wang, and X. Wang, 'Unsupervised Deep Anomaly Detection for Multi-Sensor Time-Series Signals', IEEE Trans. Knowl. Data Eng, vol. 35, no. 2, pp. 2118-2132, Feb. 2023. doi.org/10.1109/TKDE.2021.3102110

How to Cite: Renas Rajab Asaad. (2025). Temporal graph attention networks for real-time anomaly detection in industrial IOT: a multi-scale hierarchical approach (TGA-Net). Journal of Artificial Intelligence, Machine Learning and Neural Network (JAIMLNN), 5(2), 79–87. https://doi.org/10.55529/jaimlnn.52.79.87

BIOGRAPHIE OF AUTHOR

Renas Rajab Asaad, is a dedicated academic and researcher in the field of Computer Science at Nawroz University. His research interests include artificial intelligence, data analytics, machine learning, and advanced computing technologies. He has contributed to various scholarly activities focused on innovative solutions for real-world computational challenges. With a strong commitment to academic excellence, he actively engages in research, teaching, and the development of modern computer science applications that support technological advancement and scientific innovation. Email: renas.beda89@gmail.com