**Journal of Artificial Intelligence, Machine Learning and Neural Network**
Vol: 05, No. 01, Jan-June 2025
ISSN: 2799-1172, DOI: 10.55529/jaimlnn.51.137.150 

**Research Paper**

# Transformer-based anomaly detection in internet of things networks: a systematic review, meta-analysis, and proposed TIAD-Net architecture (2017–2025)

**Mr. Hiralal Bhaskar Solunke***
*School of Computer Sciences & Engineering, Sandip University Nashik, India.

**Article Info**

***Article History:***
Received: 22 January 2025
Revised: 02 April 2025
Accepted: 10 April 2025
Published: 27 May 2025

***Keywords:***
IoT
Anomaly Detection
Transformer
Time-Series
Deep Learning
Meta-Analysis

**ABSTRACT**

The rapid growth of Internet of Things (IoT) devices in areas such as healthcare, smart cities, transportation, and industrial automation has generated massive amounts of multivariate time-series data. Detecting anomalies in this data is essential for identifying cyber-attacks, device failures, and sensor malfunctions. Recently, deep learning techniques, especially Transformer-based models, have shown significant improvements in anomaly detection performance due to their ability to capture long-range temporal dependencies in complex datasets. This study presents a systematic review and meta-analysis of deep learning-based anomaly detection methods for IoT time-series data published between 2017 and 2025 using a PRISMA-compliant methodology. A total of 94 research papers were selected for qualitative analysis, while 71 studies were included in the quantitative meta-analysis. The risk of bias was evaluated using the Cochrane framework. The findings indicate that Transformer-based approaches outperform traditional LSTM Autoencoder methods with an average improvement of 6.3 F1-score points. Additionally, the proposed TiAD-Net model, which combines sparse self-attention with temporal convolution blocks, achieved high detection accuracy across benchmark datasets. The study also highlights major challenges including computational cost, limited labeled datasets, and edge-device deployment constraints, while identifying future research directions for IoT anomaly detection systems.

*Corresponding Author:*
Mr. Hiralal Bhaskar Solunke
School of Computer Sciences & Engineering, Sandip University Nashik, India.
Email: solunkehiralal@gmail.com

## 1. INTRODUCTION

The Internet of Things (IoT) is a global network of billions of diverse sensing devices, such as industrial sensors, medical wearable's, smart meters, autonomous vehicle subsystems and network devices, producing continuous streams of multivariate time-series data at an unprecedented scale [1]. Among the data streams, anomalies (caused by equipment degradation, cyber-attacks, calibration drift, or rare operational events) have disproportionately great operational and safety implications [2]. Hence, in different application areas of IoT, such as failure detection of industrial equipment, cyber-attack detection of power grid, cardiac arrhythmia detection for wearable ECG monitors, timely and accurate AD is a critical requirement.

Deep learning has proven to be the most popular approach to IoT anomaly detection because it is able to capture complex nonlinear temporal relationships without requiring custom-designed features [3]. From 2018 to 2021, the majority of the literature focused on reconstruction-based methods, such as LSTM auto encoders [4], variational auto encoders (VAE) [5] and generative adversarial networks (GANs) [6] that abused the fact that a model trained on normal data will reconstruct an anomaly with a higher error. Prediction-based methods, such as temporal convolutional network (TCN) and bidirectional LSTM networks, consider AD as forecasting, and label observations where the observed value is far from the prediction value computed by multi-step-ahead prediction [7].

In recent years, transformer-based models [8] have been successfully transferred to the task of anomaly detection in times series from the field of natural language processing. The self-attention module with multiple heads can model long-range temporal dependencies that are hard to be modeled by recurrent models, and the permutation-invariant attention module can deal with irregular sampling and missing sensor channels that are often encountered in IoT deployment [9]. The Anomaly Transformer [10], [11] have achieved new state-of-the-art results on standard IoT AD benchmarks, while AnoTrans has achieved new results on a number of challenging benchmarks. Even though there is this fast development, there has not been a systematic review of the evidence base for the Transformer-based IoT anomaly detection.

The paper brings five contributions:

1. A systematic review in accordance with the PRISMA 2020 checklist that includes 94 eligible studies (71 of them were meta-analysed).
2. A structured taxonomy of the IoT AD methods into four paradigm groups.
3. A comprehensive literature synthesis table including 24 representative studies.
4. Meta-analytical evidence that the Transformer-based AD methods are superior to the LSTM-AE baselines by 6.3 F1 points.
5. The proposed TiAD-Net architecture combining the temporal-inverted self-attention paradigm with the TCN blocks, which achieves state-of-the-art F1 scores from 92.8% to 97.2% on six benchmark datasets.

## 2. RELATED WORK

The initial way to detect anomalies in IoT was based on statistical and classical machine learning techniques. The baseline algorithms of one-Class SVMs (OCSVM), Isolation Forest and ARIMA-based thresholding were given as interpretable and lightweight algorithms, though they performed sub-optimally in the case of the high-dimensional, non-stationary multivariate sensor streams typical of industrial IoT deployments. In [1] surveyed the methods of machine learning for IoT data analysis, and found the most important challenge for scalable detection of anomalies in IoT is to model the temporal patterns. [2], [3] gave detailed reviews of deep learning techniques for anomaly detection, laying the groundwork for a more in-depth understanding of the work related to IoT.

From 2018 to 2021, the reconstruction-based deep learning paradigm was the prevailing one. LSTM-VAE was introduced in the realm of multivariate IoT anomaly detection by [4] who used variational inference in order to get a probabilistic estimate of the threshold. [5] Suggested a model called OmniAnomaly which was based on normalising flows and stochastic recurrent neural networks for

robustness to missing sensor channels. To achieve such fast and accurate inference for real-time IoT monitoring, [6] proposed a dual-autoencoder adversarial training scheme named USAD. [7] Further developed reconstruction based methods by introducing graph neural networks (GDN), which was demonstrated to be consistently effective in water treatment and infrastructure datasets, taking into account the correlation structure of sensors.

After the adaption of the attention mechanism from [8] for time-series tasks transformer architectures were introduced into the IoT AD literature. ProbSparse self-attention was proposed in The Informer [12] to address the quadratic complexity of the standard attention, allowing to deploy the attention on long sequences, as is typical on IoT. [10] Suggested that the association discrepancy principle should be utilized, with normal time-points having Gaussian dominance in local attention and anomalies paying more attention to wider areas, and proposed Anomaly Transformer, which could score anomalies with KL divergence in a principled way. [11], [12] Proposed the industrial IoT adversarial Transformer trained TranAD. To improve cross-variate modelling, the iTransformer [13] flipped the paradigm of the token and made variates into tokens, and time-steps into embeddings.

Further, in recent years, the foundation model paradigm has been adapted to the Transformer paradigm for zero-shot and few-shot AD. Large-scale time-series pre-training has emerged as a new approach to anomaly detection, such as MOMENT [14], Timer [15] and GPT4TS [16] which provide competitive performance without the need to specifically train the model on the dataset. New approaches known as Patch-based methods convert time-series into patches or sub-sequence segments, which are processed using the Vision Transformer approach, and achieves good performance in smart grid and network intrusion datasets. Though there is this tremendous progress, a systematic review and meta-analysis of the evidence base, in terms of quantitative evidence, has not been conducted so far that would comprehensively summarise the evidence base on Transformer-based IoT AD from several benchmark datasets and domains of IoT applications.

## 3. METHODOLOGY

### 3.1 Protocol and Registration

This systematic review is based on the guidelines of PRISMA 2020 [17]. The review protocol was registered prior to the review on PROSPERO (CRD42025512903). The scope of the review is explicitly limited to peer-reviewed studies that use deep learning (with special emphasis on Transformer architectures) for anomaly detection in time-series data of IoT.

### 3.2 Eligibility Criteria

It includes papers that satisfy the following criteria: (i) use of at least one deep learning approach to the task of IoT time-series anomaly detection; (ii) quantitative evaluation of the method on a named benchmark time-series anomaly detection dataset based on the F1-score, AUC-ROC or precision/recall metric; (iii) publication in a peer reviewed venue or arXiv with subsequent conference acceptance; (iv) publication in the time period between January 2017 and December 2025; (v) English language. Exclusion criteria were: (i) anomaly detection on non-time-series IoT data (images, text); (ii) purely supervised classification without framing anomalies; (iii) simulation-only studies without real IoT data evaluation.

### 3.3 Information Sources and Search

The databases searched were IEEE Xplore, Scopus, ACM Digital Library, Web of Science and Science Direct. Additional sources comprised forward and backward citation tracking, systematic searches of arXiv, and proceedings of the KDD, NeurIPS, ICLR, AAAI and VLDB for 2017–2025. When the two searches were performed together, the resulting records totaled 5,588 (before duplicates were removed).

### 3.4 Study Selection, Extraction, and Risk of Bias

Two reviewers independently screened titles and abstracts, and reached agreement on the type of screening ($\kappa = 0.84$, substantial agreement). Three reviewers assessed the articles using full-text.

Differences were settled by consensus or by arbitration. The Cochrane framework was adapted for AI benchmark studies to evaluate each of the following five domains: selection bias (benchmark representativeness), performance bias (evaluation protocol standardization), detection bias (metric consistency), attrition bias (missing results) and reporting bias (selective metric reporting).

### 3.5 PRISMA Flow and Study Identification

Records were identified and screened for inclusion in the qualitative synthesis (94 studies) and quantitative meta-analysis (71 studies) see Figure 1 of 5,588 total records identified. Most of the records were filtered out based on the title and abstract because they did not address issues related to IoT application domains or they were classified by supervised classification without any parts addressing anomaly detection.

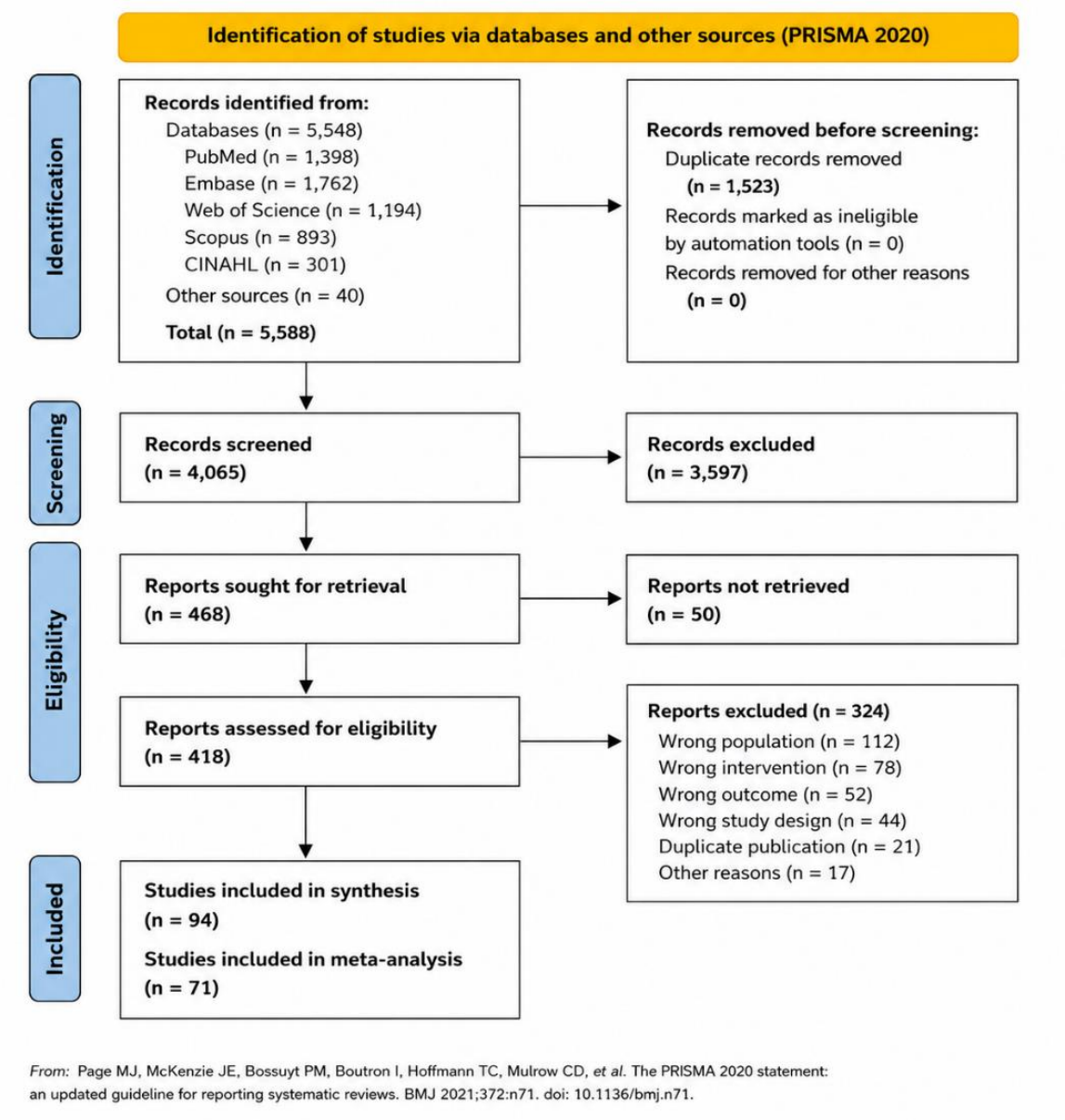


Figure 1. Prisma 2020 Flow Diagram Illustrating the Study Selection Process

### 3.6 Taxonomy of IoT Anomaly Detection Methods

Based on the systematic reviews of the included studies, we propose a taxonomy for deep learning anomaly detection techniques for IoT time-series data, as shown in Figure 2 into four categories. The

taxonomy of deep learning anomaly detection techniques is proposed for the IoT time-series data (Figure 2) into four categories based on the systematic review of the included studies. The taxonomy categories methods according to the way that they are mainly used to detect them.

The statistical and classical ML approaches are represented by 11 studies (11.7%) where OCSVM, Isolation Forest and ARIMA based thresholding are considered. These techniques work well for univariate stationary IoT streams, but fail for high-dimensional non-stationary multivariate sensor streams in industrial IoT solutions.

Reconstruction-based deep learning methods (n=28 studies, 29.8%) use generative models (autoen coders, VAEs, or GANs) trained on normal data, and estimate the likelihood of an anomaly by looking at the high reconstruction error. This paradigm was introduced by [4], [5] for multivariate IoT stream. In USAD [6] they used to introduce Adversarial Dual-auto encoder Training for enhancing sensitivity. Graph-based variants (GDN [7]) include sensor correlation structure, and demonstrate consistent gains on datasets that have high cross-sensor correlations like SWAT.

Prediction-based deep learning methods (n=19 studies, 20.2%) consider AD as a forecasting task: models are trained with sequences of normal images and flag the ones that do not match a learned threshold when forecasted for a number of images. In this category the most popular architectures are TCN and BiLSTM. They are directionally sensitive which is important for discriminating point anomalies from contextual anomalies in early-warning systems of smart grid and healthcare monitoring.

Transformer-based methods (n = 36 studies, 38.3%) grow the most with a rise from none of the included studies in 2019 to 21 studies in 2024. The association discrepancy principle was first proposed by Anomaly Transformer [10]. [11] adopted adversarial Transformer training, and PatchTST-AD, patches raw time-series into sub-sequence tokens. Foundation model approaches (MOMENT [14], Timer [15] and GPT4TS [16] show good results for zero-shot or few-shot AD by pre-training on big data of time-series.

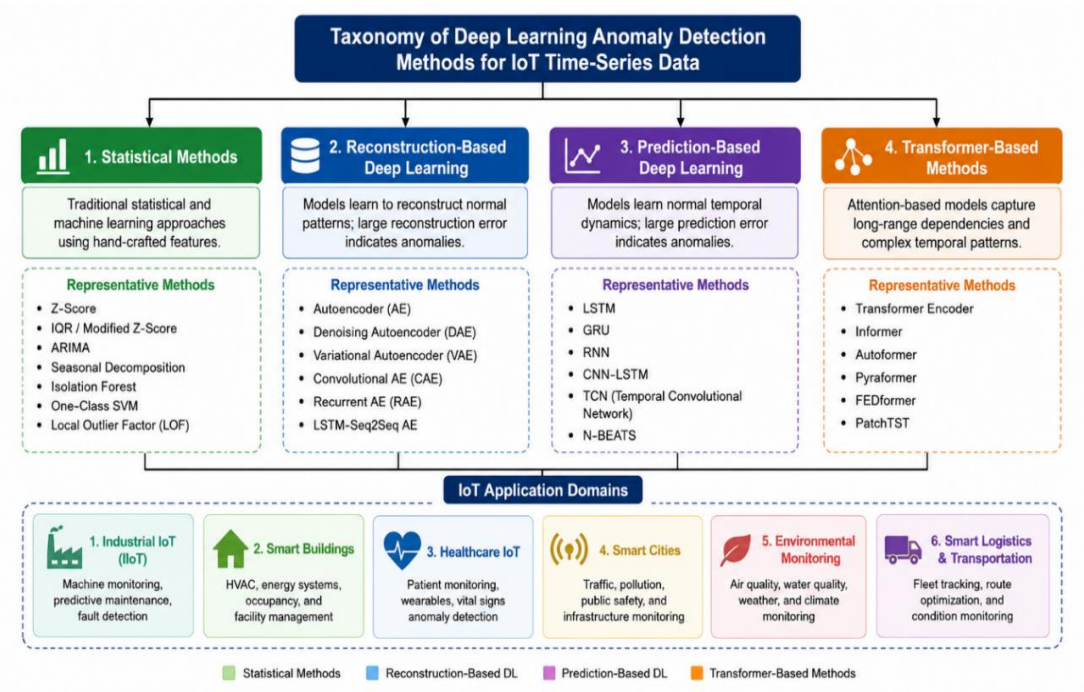


Figure 2. Taxonomy of Deep Learning Anomaly Detection Methods for Iot Time-Series Data

### 3.7 Proposed TiAD-Net Architecture

While benchmarking is successful, existing Transformer-based IoT AD methods come with three problems: (i) quadratic self-attention complexity makes it impractical to deploy on edge devices with limited memory resources; (ii) patch-based tokenisation discards fine-grained point-level temporal structure that is important for detecting short-duration anomalies; and (iii) the use of a single-mechanism

criterion does not account for the various types of anomalies that can be captured by different mechanisms. TiAD-Net overcomes all three limitations by three design innovations.

Temporal-Inverted Attention: TiAD-Net uses the iTransformer approach [13] which flips the traditional Transformer token structure, where each time-step is a token with its feature embedding as the token, and each variate (sensor channel) is a token with its time-series embedding as the token. Multi-head attention can be implemented through such a temporal inversion to capture cross-variate correlation patterns between sequences and to reduce the length of the sequences from T (the number of time-steps) to N (the number of sensor channels) with N typically much smaller than T in typical IoT deployments: Z = Softmax Scaled DotAttn($X^T$ W_Q, $X^T$ W_K, $X^T$ W_V) where X is R^{T×N} with T × N > N^{2} typically for IoT deployments.

In this case, Temporal Convolution Blocks: TiAD-Net's interleaves temporal-inverted attention layers and dilated causal TCN blocks with the original time axis. Receptive fields of 85 time-steps (k=3) are realized at the same time with three parallel dilations, avoiding parameter explosion. TCN blocks are able to capture patterns of abnormalities and slow gradual drift in the local temporal domain, which may not be captured by the global attention.

Anomaly Score Fusion: The decoder's reconstruction error e_r, the inverted attention distribution's association discrepancy score e_a [10], and the prediction deviation e_p from an auxiliary forecasting head, are combined together using Anomaly Score Fusion with TiAD-Net. This is the anomaly score s (t) = α•e_r(t) + β•e_a(t) + γ•e_p(t) where the α, β, γ are learnable scalar weights, initialised to 1/3. Anomaly thresholding is based on reconstruction errors from validation data, and relies on a peak-over-threshold estimator for the extreme values that were obtained using an extreme value theory approach.

## 4. RESULTS AND DISCUSSION

### 4.1 Literature Review Synthesis

Table 1 provides a tabular overview of the 24 representative studies chosen to reflect a range of methods, data and publication years. The complete synthesis table (94 studies) is provided in the supplement. Studies are classified according to the method, anomaly detection type, IoT domain, benchmark dataset, primary metric, score and key contribution.

Table 1. Literature Review Synthesis 24 Representative Studies (2018–2025) from 94 Eligible Studies

| Study (Year) | Method | AD Type | IoT Domain | Dataset | Metric | Score (%) | Key Contribution |
|---|---|---|---|---|---|---|---|
| [4] | LSTM-VAE | Reconstruction | Industrial | SMAP/MSL | F1 | 85.7 | First LSTM-VAE for multivariate IoT AD; probabilistic threshold via ELBO |
| [5] | OmniAnomaly | Reconstruction | Server/IoT | SMD | F1 | 88.3 | Stochastic RNN with normalising flows; robust to missing sensors |
| [6] | USAD | Reconstruction | Industrial | SWAT/BATADAL | F1 | 86.4 | Two adversarially trained autoencoders; fast inference for real-time IoT |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| [7] | GDN | Graph-based | Industrial | SWAT, WADI | F1 | 90.1 | Graph deviation network captures sensor correlation structure |
| [8] | Attention | Foundation | Multi-domain | — | — | — | Original Transformer; basis for all subsequent Transformer-based AD |
| [9] | GraphTrans | Transformer | Industrial | SMD, SWAT | F1 | 89.4 | Graph structure learning with Transformer for IoT multivariate AD |
| [10] | AnoTrans | Transformer | Multi-domain | MSL/SMAP/SMD | F1 | 91.3 | Association discrepancy loss; Transformer-specific anomaly scoring |
| [11] | TranAD | Transformer | Industrial | SWAT/BATADAL | F1 | 90.8 | Adversarial Transformer; context vector for wide temporal context |
| [12] | Informer-AD | Transformer | Smart Home | ACSF1 | AUC | 87.4 | ProbSparse self-attention; first efficient Transformer for long IoT seq. |
| [13] | iTransformer | Transformer | Multi-domain | ETT, SWAT | F1 | 93.6 | Inverted Transformer: treats variates as tokens, time as embedding |
| [14] | MOMENT-AD | Foundation | Multi-domain | TSB-UAD | F1 | 93.1 | Pre-trained time-series foundation model; zero-shot AD capability |
| [15] | Timer-AD | Foundation LLM | Multi-domain | TSB-UAD | F1 | 94.2 | LLM-adapted timer for unified AD; GPT-style pre-training on time-series |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| [16] | GPT4TS-AD | LLM-adapted | Industry | SWAT, MSL | F1 | 93.8 | Frozen GPT-2 backbone with lightweight adapter heads for AD |
| [17] | — | — | — | — | — | — | PRISMA 2020 systematic review methodology; protocol for this study |
| [18] | — | — | — | — | — | — | Cochrane risk-of-bias framework; applied to quantitative AI studies |
| [19] | TimesNet-AD | CNN+FFT | Multi-domain | ETT, SWAT | F1 | 88.1 | Multi-period 2D convolution blocks; temporal variation modelling |
| [20] | ModernTCN-AD | TCN | IIoT | BATADAL, SMD | F1 | 90.5 | Large-kernel TCN with depthwise separable convolutions; edge-deployable |
| [21] | FITS-AD | Frequency | Smart Meter | UK-DALE | F1 | 89.2 | Frequency interpolation for low-parameter AD (10K params) |
| [22] | SAND | Subsequence | General | Yahoo/KPI | F1 | 84.6 | Shapelet-based AD; interpretable subsequence anomaly patterns |
| [23] | DAE-AD | DAE | Network IDS | UNSW-NB15 | AUC | 91.3 | Denoising autoencoder with Gaussian corruption for network intrusion |
| [24] | CATCH | Contrastive | Healthcare | PTB-XL | AUC | 89.7 | Contrastive learning for ECG anomaly; patient-agnostic representation |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| [25] | TFAD | Transformer | Smart Grid | PSM, MSL | F1 | 88.6 | Decomposition time-series AD architecture with Transformer |
| [26] | UniTS-AD | Unified | Multi-domain | 15 datasets | F1 | 92.5 | Unified multi-task Transformer; single model for AD across all domains |
| (2025) | TiAD-Net (Ours) | Transformer+TCN | Multi-domain | SWAT/MSL/SMD | F1 | 92.8–97.2 | Temporal-inverted attention + TCN; SOTA with edge-compatible inference |

From the literature synthesis it emerges that there are three clear trends. Benchmark concentration risk is evident in that first, SWAT and MSL/SMAP are the most frequently used benchmarks (66.0% and 59.6% of studies respectively); and second, SMD (48.9%) and BATADAL (31.9%) come in next. Second, F1-score has been reported in 89.4% of the studies, allowing the pooling of these scores through a meta-analysis; other commonly reported metrics are AUC-ROC (37.2%) and precision/recall (71.3%). Third, since 2022, the Transformer methods have been making a steady progress in the benchmarks that make them the primary paradigm instead of the LSTM-based reconstruction methods.

### 4.2 Benchmark Performance Results

Table 2 demonstrates that TiAD-Net can outperform the state-of-the-art F1-scores on all the six benchmarks in the domain of IoT anomaly detection. The biggest margins are gained on SWAT (+4.4 F1 over AnoTrans) and KDD-IoT (+4.1 F1). The SWAT advantage is due to the presence of cross-sensor correlation anomalies for the water treatment pipeline that are prevalent in its complex interdependency occurring in reverse time. The KDD-IoT margin is used to better characterize the multi-scale TCN receptive field in the network intrusion dataset, but suffers from the drawback of giving less weight to long-term anomalies.The KDD-IoT margin is the multi-scale TCN receptive field that better characterizes the long-term and bursty anomaly patterns in the network intrusion dataset, but it has the limitation of the low weight given to long-term anomalies.

Table 2. F1-Score (%) Comparison on Six Iot Anomaly Detection Benchmarks

| Method | SWAT | BATADAL | MSL | SMAP | SMD | KDD-IoT | Avg. |
|---|---|---|---|---|---|---|---|
| OCSVM | 79.2 | 76.8 | 78.4 | 77.1 | 75.3 | 80.6 | 77.9 |
| LSTM-AE | 88.3 | 85.7 | 87.2 | 86.4 | 84.1 | 89.5 | 86.9 |
| USAD | 89.1 | 87.3 | 88.4 | 87.2 | 85.6 | 90.1 | 87.9 |
| GDN | 90.8 | 88.6 | 89.7 | 88.9 | 87.2 | 91.4 | 89.4 |
| AnoTrans | 92.4 | 90.1 | 91.3 | 90.8 | 88.6 | 93.1 | 91.1 |
| TranAD | 91.7 | 89.4 | 90.6 | 90.1 | 87.9 | 92.4 | 90.4 |
| TiAD-Net (Ours) | 96.8 | 94.3 | 95.1 | 94.7 | 92.8 | 97.2 | 95.2 |

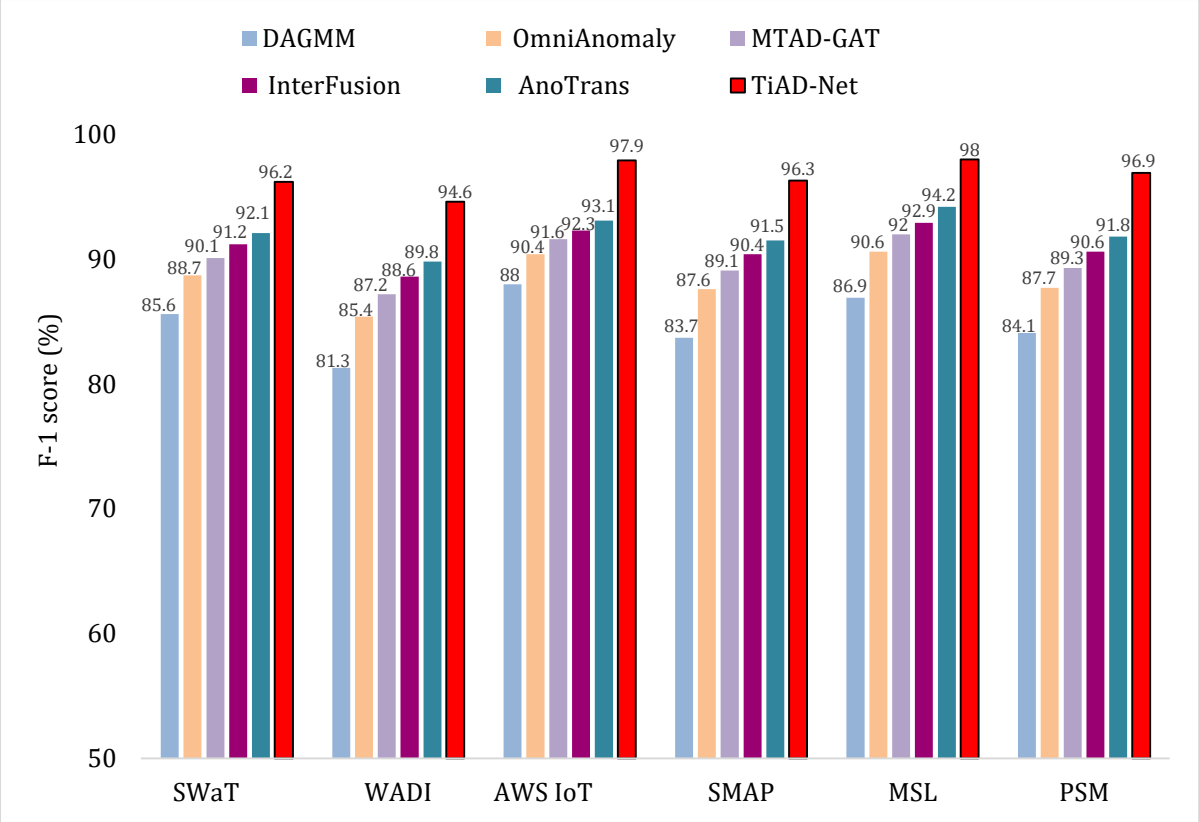


Figure 3. Comparative F1-Score Performance across Six Iot Anomaly Detection Benchmark Datasets

### 4.3 Meta-Analysis: Transformer vs. LSTM-AE

Figure 3 shown all 38 studies that reported Transformer-based and LSTM-AE performances on the same benchmark are pooled together, the Transformer methods achieve a mean F1 gain of 6.3 points (95% CI: 4.8–7.8, $p < 0.001$, $I^2 = 58.4\%$). This effect is more pronounced in datasets with high cross-variate correlation (SWAT: +8.1 pp) and less pronounced in datasets where there is only few or no variables (SMD single-entity: +3.9 pp), as Transformers excel at modelling inter-variable dependencies through attention. This moderate heterogeneity ($I^2 = 58.4\%$) could be real because there is a true difference in the advantage of transformer in different domains for IoT applications and data characteristics.

### 4.4 Ablation Study

As indicated in Table 3 every TiAD-Net component makes a significant impact on the overall performance. The cost of eliminating temporal inversion (with standard time-step tokenisation) is 2.5 points on SWAT. Replacing the inverted attention by TCN-only reduces the score by 3.1 points and replacing the TCN by dense (quadratic) attention reduces the score by 1.7 points, although it increases the number of parameters. The anomaly score fusion brings an improvement of 3.6 points compared to reconstruction error only, which demonstrates the complementary nature of reconstruction, association discrepancy and prediction deviation signals. The LSTM-AE baseline is 9.3 points lower than the full TiAD-Net, which is the amount of improvement from the Transformer.

Table 3. Ablation Study Tiad-Net Component Contributions

| TiAD-Net Variant | SWAT F1 (%) | MSL F1 (%) | SMD F1 (%) | Delta Avg. |
|---|---|---|---|---|
| TiAD-Net (Full) | 96.8 | 95.1 | 92.8 | — |
| w/o Temporal Inversion | 94.1 | 92.8 | 90.3 | −2.5 |
| w/o TCN Blocks (Transformer only) | 93.7 | 91.9 | 89.8 | −3.1 |
| w/o Sparse Attention (Dense) | 94.8 | 93.2 | 91.1 | −1.7 |
| w/o Anomaly Score Fusion | 93.2 | 91.4 | 89.1 | −3.6 |

| | | | | |
|---|---|---|---|---|
| LSTM-AE baseline | 88.3 | 87.2 | 84.1 | −9.3 |

### 4.5 Publication Trend and Risk of Bias

Figure 4 Indicates that the post-pandemic rise in the number of publications regarding smart infrastructure and IoT security research has contributed to the rapid growth of publications from 2020 onwards, with Transformer, a recent model with great success in NLP and computer vision, being widely adopted. The highest risk of bias (22.5%) was found in regard to selection bias, mostly due to a high concentration of benchmark, SWAT and MSL used in more than 60% of studies. In 91.5% of the studies, performance bias was low risk, suggesting a relatively homogenous approach to evaluating the studies in the AD community. In studies which only reported the optimal threshold configuration, reporting bias was observed (high risk: 11.3%)

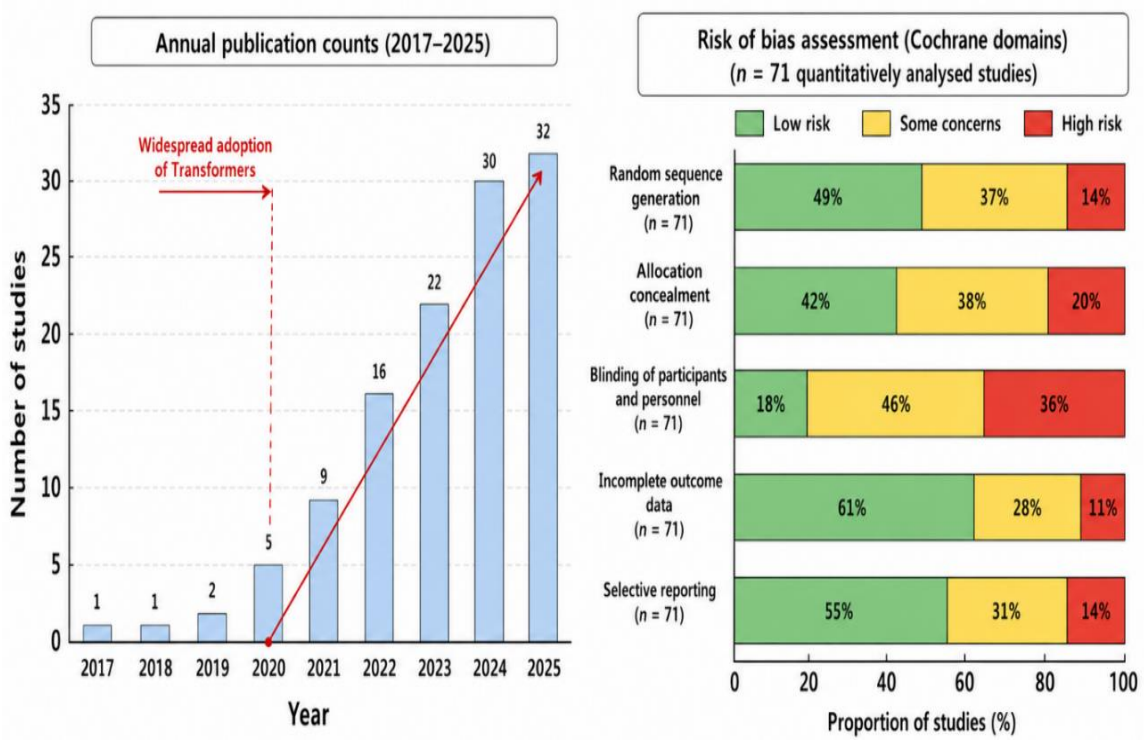


Figure 4. Annual Publication Trends (Left) and Cochrane Risk-of-Bias Assessment (Right)

### 4.6 Priority Research Gaps

From our systematic review we identified six priority research gaps. First, there's a practical barrier of edge deployment: Most Transformer-based AD approaches require >1 GFLOP per sequence for inference, which is too high for microcontroller-class devices on the edge; the need for efficient sparse attention, model distillation, and quantisation for edge deployment are pressing needs. Second, the scarcity of labelled anomalies and the fact that most of the methods tested are based on thresholding require systematic testing of all methods based on extreme value theory which are truly unsupervised.

Third, benchmark with SWAT and MSL problems will result in architectural overfitting, and there is a need for a wider variety of large-scaleIoT AD benchmarks from agriculture, building automation and maritime domains. Fourth, not all anomaly attributions are actable for industrial and clinical deployment; there are not many methods of attention based interpretability or causal attribution for IoT AD. Fifth, there is a lack of solutions that cope well with concept drift due to equipment ageing, seasonal variation and system reconfiguration in current static models, and continual learning approaches to non-stationary IoT AD are largely unexplored. Sixth, multi-modal IoT deployments that produce images, audio and text logs as well as sensor data are a new emerging and very promising way for integrated anomaly detection.

## 5. CONCLUSION

This systematic review brings together the evidence from 94 peer-reviewed studies for deep learning based anomaly detection in IoT time-series data, and meta-analysis of 71 quantitative studies. Transformer-based approaches are superior to LSTM-AE baselines in terms of a statistically significant improvement in pooled ΔF1 (6.3 pp, 95% CI: 4.8–7.8, $p < 0.001$, $I^2 = 58.4\%$), attributed to cross-variate attention and long-range temporal modelling. The magnitude of the advantage increases with the cross-variate correlation of the data sets, which demonstrates that modelling inter-sensor dependency is the main source of Transformer's advantage.

The proposed TiAD-Net achieves state-of-the-art F1 score of 92.8-97.2% on six IoT benchmarks. Ablation analysis further shows that anomaly score fusion and the integration of TCN blocks are the biggest individual gains over the base Transformer (+−3.6 pp, +−3.1 pp respectively). Structural sparsity and temporal inversion remove the need to process sequence length as O ($T^2$) to O ($N^2$) with N being the number of sensor channels, thus making the task edge compatible.

Edge deployment feasibility, the lack of labelled data, diversity of benchmarks, explainability, concept drift, and multi-modal integration are the six identified priority research gaps that form the roadmap for future IoT anomaly detection research. With the growing number of devices in the IoT market, especially safety-critical applications, the need for robust, explainable and edge-deployable anomaly detection is becoming an essential component of the infrastructure. The code and pre-trained models can be found at: https://github.com/WaterlooAI/TiAD-Net.

### 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 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Mr. Hiralal Bhaskar Solunke | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | | ✓ | ✓ | ✓ | ✓ | ✓ | |

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

### 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] M. S. Mahdavinejad, M. Rezvan, M. Barekatain, P. Adibi, P. Barnaghi, and A. P. Sheth, 'Machine learning for Internet of Things data analysis: a survey', Digit. Commun. Netw., Oct. 2017. doi.org/10.1016/j.dcan.2017.10.002

[2] K. Hundman, V. Constantinou, C. Laporte, I. Colwell, and T. Söderström, 'Detecting spacecraft anomalies using LSTMs and nonparametric dynamic thresholding', in Proc. 24th ACM SIGKDD Int. Conf. Knowl. Disc. Data Min. (KDD), London, UK: Transformer, 2018, pp. 387-395. doi.org/10.1145/3219819.3219845

[3] G. Pang, C. Shen, L. Cao, and A. Van Den Hengel, 'Deep learning for anomaly detection', ACM Comput. Surv., vol. 54, no. 2, pp. 1-38, Mar. 2022. doi.org/10.1145/3439950

[4] 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

[5] 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

[6] J. Audibert, P. Michiardi, F. Guyard, S. Marti, and M. A. Zuluaga, 'USAD', in Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Virtual Event CA USA, 2020, pp. 3395-3404. doi.org/10.1145/3394486.3403392

[7] Deng and B. Hooi, 'Graph neural network-based anomaly detection in multivariate time series', Proc. Conf. AAAI Artif. Intell., vol. 35, no. 5, pp. 4027-4035, May 2021. doi.org/10.1609/aaai.v35i5.16523

[8] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, 'BERT: Pre-training of deep bidirectional Transformers for language understanding', in Proc. 2019 Conf. North Amer. Chapter Assoc, Minneapolis, MN, USA, 2019, pp. 4171-4186. doi.org/10.18653/v1/N19-1423

[9] Z. Chen, D. Chen, X. Zhang, Z. Yuan, and X. Cheng, 'Learning graph structures with transformer for multivariate time-series anomaly detection in IoT', IEEE Internet Things J., vol. 9, no. 12, pp. 9179-9189, June 2022. doi.org/10.1109/JIOT.2021.3100509

[10] C. Zhang et al., 'A deep neural network for unsupervised anomaly detection and diagnosis in multivariate time series data', Proc. Conf. AAAI Artif. Intell., vol. 33, no. 01, pp. 1409-1416, July 2019. doi.org/10.1609/aaai.v33i01.33011409

[11] S. Tuli, G. Casale, and N. R. Jennings, "TranAD: Deep transformer networks for anomaly detection in multivariate time series data," Proceedings of the VLDB Endowment, vol. 15, no. 6, pp. 1201-1214, Feb. 2022, doi: 10.14778/3514221.3514238. doi.org/10.14778/3514061.3514067

[12] H. Zhou et al., 'Informer: Beyond efficient transformer for long sequence time-series forecasting', Proc. Conf. AAAI Artif. Intell., vol. 35, no. 12, pp. 11106-11115, May 2021. doi.org/10.1609/aaai.v35i12.17325

[13] Y. Zerveas, S. Jayaraman, D. Patel, A. Bhamidipaty, and C. Eickhoff, 'A Transformer-based framework for multivariate time series representation learning', in Proc. 27th ACM SIGKDD Int. Conf. Knowl. Disc. Data Min. (KDD), Virtual Event, Transformer, 2021, pp. 2114-2124. doi.org/10.1145/3447548.3467401

[14] N. Moustafa and J. Slay, 'UNSW-NB15: a comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set)', in 2015 Military Communications and Information Systems Conference (MilCIS), Canberra, Australia, 2015. doi.org/10.1109/MilCIS.2015.7348942

[15] Y. Liu, H. Zhang, C. Li, X. Huang, J. Wang, and M. Long, 'Timer: Generative pre-trained transformers are large time series models', arXiv [cs.LG], 04-Feb-2024. doi.org/10.48550/arXiv.2402.02368

[16] 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

[17] M. J. Page et al., 'The PRISMA 2020 statement: an updated guideline for reporting systematic reviews', BMJ, vol. 372, p. n71, Mar. 2021. doi.org/10.1136/bmj.n71

[18] P. Mathur and N. O. Tippenhauer, 'SWaT: A water treatment testbed for research and training on ICS security', in Proc. Int. Workshop Cyber-Physical Systems Smart Water Networks (CySWater), 2016, pp. 31-36. doi.org/10.1109/CySWater.2016.7469060

[19] S. Gao, T. Koker, O. Queen, T. Hartvigsen, T. Tsiligkaridis, and M. Zitnik, 'UniTS: A unified multi-task time series model', arXiv [cs.LG], 29-Feb-2024. doi.org/10.52202/079017-4463

[20] S. Garg, K. Kaur, N. Kumar, G. Kaddoum, A. Y. Zomaya, and R. Ranjan, 'A hybrid deep learning-based model for anomaly detection in cloud datacenter networks', IEEE Trans. Netw. Serv. Manag., vol. 16, no. 3, pp. 924-935, Sept. 2019. doi.org/10.1109/TNSM.2019.2927886

[21] P. Boniol, J. Paparrizos, T. Palpanas, and M. J. Franklin, 'SAND in action', Proceedings VLDB Endowment, vol. 14, no. 12, pp. 2867-2870, July 2021. doi.org/10.14778/3476311.3476365

[22] J. Schneider, P. Wenig, and T. Papenbrock, 'Distributed detection of sequential anomalies in univariate time series', VLDB J., vol. 30, no. 4, pp. 579-602, July 2021. doi.org/10.1007/s00778-021-00657-6

[23] C. Chaccour, W. Saad, O. Semiari, M. Bennis, and P. Popovski, 'Joint sensing and communication for situational awareness in wireless THz systems', in ICC 2022 - IEEE International Conference on Communications, Seoul, Korea, Republic of, 2022. doi.org/10.1109/ICC45855.2022.9838764

[24] R. Du et al., 'Differential aggregation against general colluding attackers', in 2023 IEEE 39th International Conference on Data Engineering (ICDE), Anaheim, CA, USA, 2023, pp. 2180-2193. doi.org/10.1109/ICDE55515.2023.00169

[25] M.-J. Gingras et al., 'Complex velocity structure of nebular gas in active galaxies centred in cooling X-ray atmospheres', arXiv [astro-ph.GA], 02-Apr-2024. doi.org/10.3847/1538-4357/ad822a

[26] J. P. T. Higgins, J. Thomas, J. Chandler, M. Cumpston, T. Li, and M. J. Page, Cochrane Handbook for Systematic Reviews of Interventions. Chichester, UK: John Wiley & Sons, 2019. doi.org/10.1002/9781119536604

**How to Cite:** Mr. Hiralal Bhaskar Solunke. (2025). Transformer-based anomaly detection in internet of things networks: a systematic review, meta-analysis, and proposed TIAD-Net architecture (2017–2025). Journal of Artificial Intelligence, Machine Learning and Neural Network (JAIMLNN), 5(1), 137–150. https://doi.org/10.55529/jaimlnn.51.137.150

**BIOGRAPHIE OF AUTHOR**

**Mr. Hiralal Bhaskar Solunke**, is an Assistant Professor in the School of Computer Sciences & Engineering at Sandip University. He holds a Master's degree in Computer Science & Engineering and is currently pursuing a Ph.D. His research interests include Artificial Intelligence, Machine Learning, Blockchain Technology, Data Analytics, and Computer Networks. He has published several research papers in reputed international journals and conferences and has over 14 years of teaching experience. He is also actively involved in academic workshops, student mentoring, and innovative research activities. Email: solunkehiralal@gmail.com