Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Transformer-Based Models for Streaming Anomaly Detection in Solar Microgrids: A Systematic Review

Zablon Ombiro

zablon.ombiro@strathmore.edu

Strathmore University https://orcid.org/0009-0009-5822-780X

Bernard Shibwabo

Strathmore University https://orcid.org/0000-0002-0827-9857

Systematic Review

Keywords: Streaming anomaly detection, Concept drift, Edge deployment, Interpretability, Real-time monitoring, Non-stationary environments

Posted Date: November 2nd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11564622/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: The authors declare no competing interests.

Abstract

*Background:*Solar Photovoltaic (PV) microgrids operate in highly non-stationary conditions, where irradiance, weather, and load variability cause frequent concept drift in streaming telemetry. Traditional anomaly detection models degrade under these dynamics.
Objective: To systematically review Transformer-based models for streaming anomaly/fault detection in energy systems, emphasizing solar PV and microgrids and appraise evidence on (i) drift adaptation, (ii) edge deployment efficiency, and (iii) interpretability for operator use.

Methods: We followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 guidelines. Records published between January 2023, and May 2026 were screened reflecting the recent emergence of Transformer-based anomaly detection research addressing streaming operation, concept drift, and edge deployment. Inclusion required a Transformer/attention-based architecture applied to anomaly/fault detection on time-series or operational telemetry, with rule-based removal of clearly irrelevant studies followed by manual validation.

Results: A total of 2968 records were identified. After duplicate removal and screening, 113 studies were included in the final synthesis. Transformer-based architecture is widely reported across included studies but shows uneven application across domains. Photovoltaic systems account for most studies, while microgrid-specific applications remain limited. Reporting of drift adaptation, edge efficiency, and interpretability is inconsistent.

Conclusions: Transformers show strong anomaly detection potential, yet gaps persist in drift-aware training, standardized edge metrics, and validated interpretability. We recommend lightweight attention mechanisms, integrated drift detectors, and operator-tested explanations to enable reliable PV microgrid deployment.

1 Introduction

Solar photovoltaic (PV) microgrids are increasingly deployed to improve resilience, decentralization, and sustainability in energy systems. These systems operate under highly dynamic and non-stationary conditions, where variability in irradiance, temperature, and load demand introduces continuous changes in system behavior. Such variability creates significant challenges for anomaly detection and fault diagnosis, which are essential for maintaining system reliability and enabling predictive maintenance in distributed energy environments [1], [2].

Traditional machine learning and deep learning approaches, including autoencoders and recurrent neural networks such as Long Short-Term Memory (LSTM), have been widely applied for anomaly detection in energy systems [113]. However, these models typically assume stable data distributions and are often designed for offline or batch processing. As a result, their performance degrades in streaming environments where concept drift and temporal variability are inherent [3], [4], [5].

Transformer-based architectures have emerged as a promising alternative due to their ability to model long-range dependencies and capture complex relationships in multivariate time-series data. Their self-attention mechanisms enable flexible representation of temporal patterns and interactions across multiple signals, making them suitable for anomaly detection in energy systems, with particular emphasis on solar photovoltaic and microgrid applications [6], [7], [8].

Recent studies have applied Transformer models across a range of tasks, including microgrid fault detection and photovoltaic system monitoring [1], [9], [10]. In addition, hybrid approaches combining Transformers with other architectures such as autoencoders, graph neural networks, and signal decomposition methods have been explored to

enhance performance in specific scenarios. However, evidence from the current review shows that such hybrid approaches remain limited relative to the dominance of standard Transformer implementations.

Despite the growing interest in Transformer-based anomaly detection, important gaps remain. Existing studies are unevenly distributed across application domains, with a strong concentration in photovoltaic systems and relatively limited attention to microgrid and operational telemetry contexts. Furthermore, many models are evaluated under static or semi-static conditions, with limited consideration of streaming data, concept drift, or real-time deployment constraints, even in studies proposing streaming or near real-time frameworks such as [11]. Reporting efficiency metrics such as latency and memory usage is inconsistent, and interpretability is often limited to attention visualization without validation in operational settings [12], [13].

This systematic review synthesizes evidence from 113 empirical studies published between 2023 and 2026, focusing on Transformer-based models for anomaly detection in energy systems, with particular emphasis on solar PV and microgrid applications. The review evaluates three key aspects: (i) applicability of Transformer models in anomaly detection tasks, (ii) handling of non-stationary and streaming data, and (iii) evidence of deployment readiness, including efficiency and interpretability.

The objective is to provide a structured analysis of current research trends, identify methodological gaps, and highlight areas requiring further investigation to support reliable deployment of anomaly detection systems in real-world energy environments.

2 Methods

2.1 Protocol and Reporting

This systematic review was conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 guidelines [14]. The review protocol was registered on the Open Science Framework (OSF) with an embargo to maintain blinding during the initial screening phase. All methodological decisions, including eligibility criteria, search strategy, and synthesis approach, were pre-specified to ensure transparency and reproducibility.

2.2 Eligibility Criteria

Articles were eligible for inclusion if they were published between January 2023 and May 2026, written in English, and comprised empirical studies applying Transformer or attention-based architectures for anomaly detection, fault diagnosis, or predictive maintenance in energy systems, with an emphasis on solar PV and microgrid applications.

Studies had to report at least one of the following: streaming context, adaptation to concept drift, computational efficiency for edge deployment (latency, memory footprint), or interpretability features. Peer-reviewed journal or conference publications were considered.

Studies were excluded if they were non-empirical, lacked streaming or operational relevance, or did not provide sufficient metadata for quality appraisal. Included studies were subsequently grouped by domain, model architecture, and data modality. The full inclusion and exclusion criteria are summarized in Table 1.

Table 1 Inclusion and exclusion criteria.

Criterion	Inclusion	Exclusion
Study type	Empirical studies applying machine learning models for anomaly detection or fault diagnosis	Reviews, surveys, tutorials, opinion papers
Model requirement	Studies implementing Transformer or attention-based architectures (including hybrid variants)	Studies without Transformer or attention-based components
Task definition	Anomaly detection, fault diagnosis, or predictive maintenance	Forecasting-only studies (e.g., load or power prediction)
Domain scope	Energy systems, including solar PV, microgrids, and power systems	Studies outside energy or unrelated domains
Data modality	Time-series or operational telemetry (e.g., SCADA, inverter logs, sensor data)	Vision-only/image-based studies without operational telemetry
Operational context	Studies reflecting temporal/system behavior (streaming, sequential, or time-dependent inputs)	Static, non-temporal, or batch-only studies without system dynamics relevance
Search alignment	Studies matching search query constructs: (Transformer/attention) AND (anomaly/fault) AND (energy systems, PV, microgrid)	Studies retrieved by broadened queries but failing task, domain, or model criteria
Selection scoring (Appendix B, Table B1)	Studies meeting threshold score (TS ≥ 9.5), prioritizing anomaly detection, time-series, Transformer models, and real-world datasets	Studies below threshold score (TS < 9.5), including weak relevance to anomaly detection or deployment context
Publication criteria	Peer-reviewed journal or conference papers (2023–2026, English)	Non-peer-reviewed sources, non-English, or outside time window
Reporting quality	Studies with sufficient methodological detail (model, data, evaluation)	Studies with missing, unclear, or non-reproducible descriptions

Abbreviations: SCADA Supervisory Control and Data Acquisition,

2.3 Research Questions

1. What types of Transformer-based and attention-based architectures are used for anomaly detection in energy systems?

2. How are these models applied across different domains, data modalities, and operational contexts?

3. To what extent do existing approaches address real-world operational requirements, including streaming data, concept drift, and deployment constraints?

4. What evaluation practices and performance metrics are used, and how consistent and comparable are they across studies?

5. What are the key limitations and research gaps in current Transformer-based anomaly detection approaches for real-world energy systems?

2.4 Information Sources

A structured and comprehensive search was conducted across five major electronic databases: IEEE Xplore, ACM Digital Library, SpringerLink, ScienceDirect (Elsevier), and Web of Science. In addition, Google Scholar was used to

capture additional records through extended retrieval.

2.5 Search Strategy

The search strategy was developed and executed using structured query formulations across all selected databases. Queries were designed to capture three core dimensions: model type, analytical task, and application domain, with additional constraints addressing streaming conditions, concept drift, and operational deployment.

Model-related terms included *"transformer"*, *"attention"*, *"anomaly transformer"*, *"Informer"*, and *"Reformer"*. Task-related terms included *"anomaly detection"*, *"fault detection"*, and *"fault diagnosis"*. Domain-specific terms included *"microgrid"*, *"solar microgrid"*, *"photovoltaic"*, *"PV"*, *"inverter"*, and *"SCADA"*.

Additional operational constraints were included through terms such as *"streaming"*, *"online"*, *"real-time"*, *"concept drift"*, and *"non-stationary"*. Where supported by database syntax, exclusion filters were applied to remove forecasting- and vision-based studies (e.g., *forecast*, *YOLO*, *object detection*, *thermal image*, *DETR*, *Swin*), ensuring alignment with the defined scope.

Searches were conducted across IEEE Xplore, SpringerLink, ScienceDirect (Elsevier), ACM Digital Library, Web of Science, and Google Scholar. Database-specific queries (Q1–Q17) and configurations are detailed in Appendix A (Table A1).

All searches were restricted to publications between January 2023 and May 2026. Filters were applied based on document type (journal and conference papers), subject relevance, and language (English). Search dates ranged from 07 May 2026 to 19 May 2026, depending on database and query configuration.

2.6 Selection Process

The study selection process was conducted using a structured, multi-stage screening protocol in accordance with PRISMA 2020 guidelines. Title and abstract screening were performed independently by two reviewers using a predefined classification framework implemented in Excel. Studies were assessed against the inclusion and exclusion criteria, and disagreements were resolved through discussion and consensus.

Full-text screening was subsequently conducted using the same criteria to confirm eligibility. No automated or machine-learning-assisted screening tools were used. The outcomes of the selection process are reported in the Results section.

To minimize bias, author names, journal titles, and publication years were masked during initial screening.

2.7 Data Collection Process

Data extraction was conducted independently by two reviewers using a structured and standardized template (Additional file 1) aligned with the PRISMA 2020 [14] reporting framework and informed by the Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling Studies (CHARMS) [15] checklist for prediction modelling studies. The extraction protocol was piloted on a subset of studies and refined iteratively to ensure consistency and completeness across reviewers.

Each included study was reviewed in full, and relevant information was extracted from the main text, tables, figures, and supplementary materials where available. Extracted data were cross-validated between reviewers, and any discrepancies were resolved through discussion and consensus.

No direct contact with study authors was undertaken. Where required information was missing, incomplete, or ambiguous, it was recorded as “Not Reported” to preserve transparency and avoid introducing assumptions into the dataset.

The extraction process was designed to capture methodological, technical, and operational characteristics necessary for structured synthesis and comparison across studies.

2.8 Data Items

Data items were predefined based on the objectives of the review and guided by the CHARMS framework to ensure comprehensive capture of study characteristics, modelling approaches, and evaluation practices.

The extracted variables included:

· *Study characteristics:* publication year, venue, and authorship

· *Application context:* domain of application (e.g., power systems, microgrids, photovoltaic systems and related operational systems)

· *Data characteristics:* data source and dataset type (e.g., real-world, benchmark, simulated), as well as telemetry modality (e.g., SCADA, inverter logs, network traffic, image-based, multimodal)

· *Model characteristics:* architecture type (e.g., Transformer, hybrid Transformer, Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), graph-based models) and learning paradigm (supervised, unsupervised, or hybrid)

· *Operational context:* presence of streaming or real-time processing, consideration of non-stationarity or concept drift, and reported deployment context (edge, cloud, or offline)

· *Performance measures:* including accuracy, precision, recall, Harmonic Mean of Precision and Recall (F1-score), Area Under the Receiver Operating Characteristic Curver (AUROC), and other reported metrics

· *Efficiency measures:* including inference latency, memory requirements, and model complexity indicators where available

· *Adaptation strategies:* including drift detection methods, online learning mechanisms, and model update protocols

· *Interpretability features:* including attention visualization, explainability techniques, and any reported human or operator validation

The selection of these variables ensured alignment with both predictive modelling evaluation principles and system-level considerations relevant to anomaly detection in operational and streaming environments.

2.9 Risk of Bias Assessment

Risk of bias was evaluated using an adapted version of the Prediction Model Risk of Bias Assessment Tool (PROBAST) [16], tailored to machine learning-based anomaly detection studies. The assessment considered key domains including data quality and representativeness, predictor definition and measurement, outcome definition and labelling, analytical methods and evaluation practices, reproducibility, operational feasibility (e.g., latency and resource constraints), and interpretability.

Each study was independently assessed by two reviewers and classified as Low, Moderate, or High risk of bias, with disagreements resolved through consensus. This structured approach enabled consistent evaluation of methodological quality across the included studies and informed subsequent interpretation of the evidence.

2.10 Effect Measures

Effect measures were defined to capture both predictive performance and operational characteristics. Primary performance measures included classification metrics such as F1-score, precision, recall, and accuracy, as well as probabilistic measures such as AUROC where available. Domain-specific metrics, including mean average precision (mAP), were recorded for image-based studies.

Efficiency-related measures, including inference latency, memory usage, and model complexity indicators, were extracted where reported. Qualitative indicators of interpretability were also documented. Due to variability in reporting practices, all measures were synthesized descriptively rather than standardized.

2.11 Synthesis Methods

Given substantial heterogeneity in datasets, modelling approaches, evaluation metrics, and experimental settings, quantitative meta-analysis was not appropriate. Instead, a structured descriptive and narrative synthesis was conducted in accordance with PRISMA guidance for heterogeneous evidence.

Studies were grouped and analyzed across key dimensions: application domain, model architecture, data modality, evaluation practices, and adaptation strategies, including concept drift handling. Descriptive statistics summarized the distribution of these characteristics, while narrative synthesis identified patterns and inconsistencies across studies.

2.12 Reporting Bias Assessment

Reporting bias was assessed qualitatively by examining the completeness and transparency of reported outcomes. Attention was given to the omission of key performance metrics (e.g., AUROC, false alarm rate), absence of operational metrics (e.g., latency, memory, deployment context), and inconsistencies between reported claims and supporting evidence. This assessment was integrated into the synthesis to identify systematic reporting gaps.

2.13 Certainty Assessment

Certainty of evidence was evaluated qualitatively using an adapted evidence confidence framework, considering risk of bias, consistency of findings, applicability to real-world deployment contexts, and completeness of reporting. Given the methodological heterogeneity of the included studies, certainty was not quantified but assessed descriptively to reflect overall confidence in the evidence base.

3. Results

3.1 Study Selection

The database search identified a total of **2968 records** across IEEE Xplore, ACM Digital Library, ScienceDirect (Elsevier), SpringerLink, Web of Science, and Google Scholar. After identification, **777 duplicate instances** were detected and consolidated, resulting in **253 records being removed**, yielding **2715 unique records** for title and abstract screening.

Title and abstract screening were conducted using a structured and reproducible classification framework. Two independent reviewers assessed all records against predefined inclusion and exclusion criteria, with disagreements resolved through consensus. The inclusion criteria required: (i) the application of Transformer or attention-based models, (ii) anomaly detection or fault diagnosis as the primary task, and (iii) relevance to time-series or operational telemetry within energy systems.

Application of these criteria led to the exclusion of **2551 records**, leaving **164 studies** for full-text assessment. Exclusions were primarily due to: (i) absence of Transformer or attention-based architectures, (ii) focus on forecasting rather than anomaly detection, (iii) reliance on image-based inspection without integration of operational telemetry, and (iv) lack of relevance to energy system contexts.

Full-text assessment resulted in the exclusion of **51 studies**, primarily due to lack of methodological relevance, insufficient reporting, or misalignment with anomaly detection tasks. A final set of **113 studies** was retained for synthesis. The study selection process is summarized in Tables 2 and 3 and further illustrated in the PRISMA flow diagram (Figure 1).

Table 2 PRISMA Screening Summary.

Stage	Count
Records identified	2968
Duplicate instances detected	777
Records removed after consolidation	253
Records screened (title/abstract)	2715
Full-text articles assessed	164
Studies included in synthesis	113

Table 3 PRISMA Search and Screening Log Distribution Summary.

Source	Total hits	After de-duplication	Screened (title/abstract)	Full-text assessed	Included
IEEE Xplore	427	490	490	51	33
SpringerLink	851	925	925	35	28
ScienceDirect (Elsevier)	988	808	808	41	25
ACM DL	356	396	396	22	21
Web of Science (Cross Ref)	96	96	96	15	6
Google Scholar	250	-	-	-	-
Totals	**2968**	**2715**	**2715**	**164**	**113**

Figure 1 PRISMA flowchart for study selection.

3.2 Study Characteristics

The final synthesis comprises 113 studies, selected through a structured multi-stage screening pipeline. The dataset exhibits clear patterns across domain focus, data modalities, model architecture, evaluation practices, and deployment considerations. These patterns reveal both dominant research directions and critical methodological gaps.

3.2.1 Domain Distribution

Table 4 shows that included studies are distributed across multiple application domains, with a strong concentration in power systems and cybersecurity contexts.

Table 4 Domain Distribution of Included Studies (n = 113).

Domain	Count (n)	Representative References
Smart Grid / Power Systems / Microgrids	30	[17], [18], [19], [1], [20], [21]
Cybersecurity / IDS / IIoT	23	[7], [22], [23], [24], [25]
Time-Series (General / Benchmark / Methods)	17	[26], [27], [28], [9.]
Industrial / CPS / Manufacturing	14	[29], [30], [31], [32], [33]
Renewable Energy (PV / Wind)	13	[34], [35], [2], [36]
IoT / Sensor Networks	9	[37], [38], [39]
Multimodal / Cross-domain	7	[40], [41]

Abbreviations: IDS intrusion detection system; IIoT Industrial Internet of Things; CPS Cyber-Physical Systems; IoT Internet of Things

The distribution of studies indicates a strong concentration in power systems, with additional representation from cybersecurity and related domains, which together account for a large proportion of the dataset. This reflects the

operational importance of anomaly detection in critical infrastructure systems, particularly in smart grids and industrial control environments.

However, this distribution is uneven. Microgrid-focused studies remain relatively limited compared to broader smart grid research, while multimodal and cross-domain studies are still emerging despite their relevance for real-world monitoring.

3.2.2 Architecture Distribution

A range of modelling architectures is observed across the dataset, Table 5, with a strong emphasis on Transformer-based approaches, particularly hybrid configurations.

Table 5 Model Architecture Distribution.

Architecture Type	Count (n)	Representative References
Hybrid Transformer-based	39	[7], [23], [18], [24]
Pure Transformer	28	[42], [12], [2]
CNN-based models	18	[44], [29], [32]
RNN / LSTM-based	14	[20], [36], [111]
Autoencoder / VAE-based	11	[26], [10]
Graph-based models (GNN/GCN)	8	[21], [45]
Optimization + ML hybrids	7	[46], [25]

Abbreviations: GNN Graph Neural Network; GCN Graphical Convolutional Network, ML Machine Learning

The dataset shows that Transformer-based architectures dominate, primarily through hybrid implementations rather than standalone models. This reflects the need to combine attention mechanisms with established feature extraction and temporal modelling approaches.

Hybrid models frequently integrate CNNs for feature extraction, LSTMs for sequential dependencies, and Transformers for global context modelling. Pure Transformer approaches are less prevalent and are often evaluated in controlled or benchmark environments. Graph-based architecture remains relatively limited despite their relevance to structured energy systems.

Overall, this reflects a fragmented architectural landscape in which no single modelling approach has emerged as dominant, indicating unresolved challenges in achieving unified and efficient design strategies.

3.2.3 Data Modality and Telemetry

Studies utilize a variety of data types, with a clear dominance of time-series telemetry as summarized in Table 6.

Table 6 Data Modality / Telemetry Distribution.

Data Type	Count (n)	Representative References
Multivariate Time-Series (SCADA, sensors)	62	[26], [47], [2], [36], [33]
Network traffic (IDS / CPS)	21	[7], [23], [111]
Image-based (PV / vision)	14	[34], [35]
Multimodal datasets	9	[40], [41]
Graph/log-based data	7	[48], [49]

The dataset is dominated by multivariate time-series data, particularly from SCADA systems and sensor networks, highlighting the relevance of temporal telemetry in anomaly detection tasks. Network traffic data also appears frequently in cybersecurity applications, while image-based approaches are primarily limited to photovoltaic inspection.

Despite this diversity, multimodal data integration remains limited. Many studies rely on single-source telemetry and do not incorporate complementary information such as environmental conditions or control signals.

This indicates insufficient exploitation of heterogeneous and multimodal data sources, which are critical for improving robustness and generalizability in real-world anomaly detection systems.

3.2.4 Evaluation Metrics

A wide range of evaluation metrics is reported across studies, Table 7, though with considerable inconsistency.

Table 7 Evaluation Metrics Distribution.

Metric	Count (n)	Representative References
Accuracy	71	[7], [23]
Precision / Recall	66	[24], [25]
F1-score	64	[26], [8]
AUROC / AUC	28	[26], [50]
MAE / RMSE / MAPE	18	[51], [52]
FAR / FPR	11	[37], [53]
Latency / efficiency	16	[54], [12]

Abbreviations: MAE Mean Absolute Error; MARE Mean Absolute Percentage Error; RMSE Root Mean Square Error; AUC Area Under the Curve; FAR False Alarm Rate; FPR False Positive Rate

Most studies rely on classification metrics, while probabilistic and operational measures are less reported. This applies across Transformer [55], [24], classical [56], [57], and hybrid models [58], often lacking full operational evaluation [5]. This inconsistency limits reproducibility and real-world assessment.

3.2.5 Learning Regimes

Table 8 shows how different learning paradigms are employed across studies, with a strong emphasis on supervised approaches.

Table 8 Learning Strategy Distribution.

Learning Type	Count (n)	Representative References
Supervised	52	[55], [23], [33]
Unsupervised	31	[8], [59], [36]
Hybrid / Semi-supervised	22	[50], [53]
Reinforcement / Adaptive	8	[60], [62]

Supervised learning dominates [61], [43], including CPS models [57], despite anomaly detection often being unsupervised. Hybrid and semi-supervised approaches ([56], [63], [5]) are less common, and adaptive strategies remain limited or poorly reported.

3.2.6 Concept Drift and Adaptation

Concept drift handling is inconsistently addressed across the dataset.

Table 9 Concept Drift Handling.

Drift Strategy	Count (n)	Representative References
Explicit drift handling	12	[47], [64], [46]
Mentioned only	36	[22], [53]
No drift handling	65	Majority (e.g., [36], [33])

Only a few studies implement drift-aware mechanisms, Table 9, while most ignore or only mention concept drift. This limitation is evident in CPS and energy models [57] and microgrid studies [65], which rely on static or controlled datasets without adaptation.

3.2.7 Dataset Heterogeneity

A variety of dataset types are used across the included studies and summarized in Table 10.

Table 10 Dataset Types.

Dataset Type	Count (n)	Representative References
Benchmark datasets	43	[26], [8], [44]
Real-world datasets	39	[2], [111]
Simulated datasets	18	[1], [33]
Multimodal datasets	13	[40], [41]

The dataset reflects a relatively balanced use of benchmark and real-world datasets, although simulated data remains common across forecasting and anomaly detection studies [1], [33], [5]. Multimodal datasets are limited, and dataset usage is fragmented. This limits reproducibility and hinders meaningful comparison across studies.

3.3 Risk of Bias in Studies

The risk-of-bias assessment indicates widespread methodological limitations across the included studies.

Table 11 Risk of Bias Across Included Studies (N = 113).

Domain	Low Risk	Moderate Risk	High Risk	Evidence / Examples (Studies)
Data representativeness	21	34	58	High risk is common in studies using simulated or constrained datasets [1], [33], with validation in controlled settings rather than real-world telemetry, as also seen in domain-specific systems [56] and microgrid datasets [65].
Measurement validity	19	38	56	Measurement limitations are evident in studies using sensor or network data without detailed calibration or noise characterization, including [36] and [32].
Outcome definition	17	35	61	Many studies rely on implicit or reconstruction-based thresholds without clear ground truth validation, as seen in [66] and [36].
Analytical reporting	15	37	61	Incomplete reporting of evaluation setups and assumptions is observed in high-performing models such as [43] and [67], where detailed experimental conditions are limited.
Concept drift handling	12	36	65	Most studies assume static data distributions with no adaptive mechanisms, as observed in [68], [36], and [33].
Edge feasibility	18	26	69	Limited reporting of computational constraints and deployment feasibility is evident even in high-performance models such as [69], [70], and [43], [80], [71]
Reproducibility	24	28	61	Reproducibility is constrained in studies using complex pipelines or non-public datasets, including [69] and [44].
Interpretability validation	0	9	104	Interpretability is largely limited to attention or feature importance without validation, as seen in [67] and [72], with no operator-level evaluation.

High risk is concentrated (Table 11) in data representativeness, outcome definition, and reporting, particularly in studies using simulated datasets [1], [20]. Measurement and outcome definitions are inconsistently reported, including SCADA-based studies [7], [2], [73] and reconstruction-based models [8], [59], [4]. Drift handling remains limited to few studies [47], [46], [64], while reproducibility and interpretability are constrained despite some benchmark usage [74], [8]. Overall, the evidence base remains methodologically fragile.

3.4 Results of Individual Studies

The included studies demonstrate diverse modelling approaches across multiple application contexts, with Transformer-based models commonly applied to time-series anomaly detection in energy and energy and related operational systems. Representative implementations include reconstruction-based Transformer models [8],[75], lightweight streaming-oriented variants [12], and attention-based SCADA applications in photovoltaic systems [2].

Hybrid architecture is also widely reported. These include Transformer-based integration with convolutional layers for intrusion detection [7], as well as combinations with sequential or deep learning models in studies such as [68], [66], [23], [18], and [24].

Other modelling approaches include CNN-based image anomaly detection [35], [34], [76], [43], and LSTM-based sequential models [1], [20]. Classical and hybrid methods are also present, including supervised CPS detection [57], XGBoost-based behavioral modelling [58], and attention-based extensions such as [77] and [63].

Applications primarily span energy systems, including renewable and operational environments [2], [51], with additional applications in supporting domains such as cybersecurity [22], [25]. Studies report performance metrics such as F1-score, accuracy, and precision across datasets [8], [23].

3.5 Synthesis of the Results

The results demonstrate a clear dominance of Transformer-based architectures in multivariate time-series anomaly detection, with hybrid configurations emerging as the most prevalent design pattern. Table 12 synthesizes model characteristics across architectures, data, and evaluation. Transformer-based models dominate, particularly hybrid variants applied to multivariate time-series in SCADA, smart grids, and PV systems [55], [23], [8], [12], [2], [75]. Performance is consistently high (often F1 > 0.90) [8], [23], but evaluation is largely offline or pseudo-streaming, with limited reporting of latency or efficiency [54], [12], and minimal drift handling [47], [46].

Hybrid deep learning models also achieve strong results in cybersecurity and industrial contexts [69], [70], [36], [62] but exhibit high complexity and limited deployment reporting, including weak edge feasibility [69], [43]. Classical models remain effective in structured domains [61], [33], [32], including gas systems [56], CPS [57], and behavioral detection [58], but assume static distributions and lack adaptability. Attention-based hybrids (e.g., [77]) improve temporal modelling, while reconstruction-based methods rely on thresholding with limited ground truth validation [30], [59], [4], [78].

Across models, evaluation is dominated by classification metrics [7], [24], with limited use of probabilistic and operational measures [50], [51]. Dataset usage is heterogeneous, spanning benchmark [74], simulated [1], and real-world sources [2], with limited multimodal integration [40], [41]. Despite this diversity, the evidence base remains constrained by limited streaming validation, minimal drift handling, and insufficient reporting of reproducibility and interpretability, consistent with the high-risk bias profile in Table 11.

Table 12 Model-Level Synthesis of Included Studies (N = 113).

Model Category	Representative Studies	Data & Domain Context	Operational Characteristics	Performance & Evaluation	Key Limitations
Transformer-based (Pure & Hybrid)	[7], [23], [112], [12], [2], [72], [79], [80], [71], [114], [116], [82], [83], [84]	Predominantly multivariate time-series data from SCADA systems, smart grids, IoT networks, and photovoltaic systems; some multimodal integrations (e.g., sensor + image or graph-based data).	Limited true streaming implementations; most models are evaluated under offline or pseudo-streaming conditions. Drift handling is rarely implemented explicitly. High computational complexity is common, with only a few studies addressing edge deployment.	High reported predictive performance (commonly F1 > 0.90 or AUC > 0.90), typically evaluated on benchmark or controlled datasets. Operational metrics such as latency and memory are inconsistently reported.	Lack of drift-aware learning, high computational cost, limited real-world deployment validation, and weak reporting of edge feasibility and interpretability validation.
Hybrid Deep Learning (Non-Transformer Dominant)	[69], [70], [44], [36], [85], [86], [84], [87]	Multivariate time-series and network-based datasets applied in cybersecurity, industrial systems, and energy monitoring contexts.	Complex architecture combining CNNs, LSTMs, attention mechanisms, and optimization methods. Rarely designed for real-time or edge deployment. Limited support for continual learning or adaptation.	Very high reported accuracy (often >95%), mainly on benchmark datasets. Evaluation focuses on classification metrics with limited operational validation.	High model complexity, limited scalability, lack of real-world validation, and strong dependence on supervised learning and feature engineering.
Classical Machine Learning Models/Frameworks	[61], [87], [88], [33], [32], [89], [90], [114]	Structured time-series, tabular datasets, and simulated power system data used in anomaly detection and fault diagnosis.	Lower computational cost and easier deployment compared to deep learning models. Typically operate in static environments without adaptation mechanisms.	Moderate to high performance depending on feature engineering (accuracy ~0.90–1.00). Often evaluated in controlled or simulated conditions.	Limited ability to model complex temporal dependencies, lack of scalability, and absence of drift-aware or adaptive mechanisms.
Reconstruction-Based / Unsupervised Models	[30], [59], [85], [83], [109], [86], [78]	Time-series telemetry from industrial systems and energy infrastructures, often without labeled anomaly data.	Depend on reconstruction error thresholds for anomaly detection. Minimal support for streaming adaptation or dynamic thresholding.	Effective in controlled datasets where anomaly patterns are distinguishable. Interpretability is typically limited to reconstruction error without explanation.	Sensitivity to threshold selection, lack of ground truth validation, and limited robustness in real-world operational settings.

Model Category	Representative Studies	Data & Domain Context	Operational Characteristics	Performance & Evaluation	Key Limitations
Domain-Specific / Non-Core Applications	[68], [91], [92], [93], [56], [94], [65]	Diverse datasets include video streams, marketing data, image-based PV inspection, and attack modeling in smart grids.	Domain-specific pipelines with limited overlap with time-series anomaly detection in energy systems. Often not designed for streaming telemetry.	High performance within specific domains, but evaluation remains context-dependent and not generalizable.	Weak alignment with review scope, limited transferability, and lack of applicability to microgrid anomaly detection scenarios.
Survey / Review Studies	[95], [96], [69], [97], [98], [99]	No primary datasets; focus on synthesis of existing literature across anomaly detection and transformer-based models.	Provide taxonomies, conceptual frameworks, and identification of research gaps rather than implementation.	No direct performance metrics; contribute through qualitative synthesis and comparative insights.	Do not provide empirical validation; findings depend on the quality of included primary studies.

Abbreviations: SCADA Supervisory Control and Data Acquisition; AUC Area Under the Curve; PV Photovoltaic; LSTM Long Short-Term Memory; CNN Convolutional Neural Network, F1 Harmonic Mean of Precision and Recall.

3.6 Reporting Bias

Reporting bias is evident across several dimensions. Studies frequently emphasize favorable performance metrics while omitting operational or robustness-related measures. For example, [55] and [23] report high accuracy without corresponding analysis of false alarm rates or deployment feasibility.

Claims of real-time or streaming performance are also often insufficiently supported. While studies such as [68] and [54] report latency improvements, detailed benchmarking conditions and deployment contexts are rarely provided.

Interpretability claims remain weakly substantiated. Although attention mechanisms are widely used as explanatory tools, their practical interpretability is seldom validated through user studies or domain expert evaluation.

3.7 Certainty of Evidence

The overall certainty of evidence is assessed as low to moderate. While reported performance metrics suggest strong potential for Transformer-based anomaly detection models, limitations identified in risk-of-bias assessment (Section 3.3) and reporting practices (Section 3.6) reduce confidence in these findings.

In particular, the lack of standardized evaluation protocols, limited external validation, and insufficient reporting of deployment constraints the generalizability of results. The absence of robust concept drift handling, combined with reliance on static datasets, further limits applicability to real-world systems.

Consequently, although the body of evidence demonstrates clear methodological progress, its reliability and operational transferability remain uncertain.

4. Discussion

Across all categories, the synthesis highlights key structural limitations. These include limited handling of non-stationary conditions, reliance on supervised learning, inconsistent evaluation practices, and insufficient integration of multimodal telemetry data. These issues collectively constrain the robustness and operational applicability of current approaches.

4.1 General Interpretation of Results in Context of Existing Evidence

This review synthesizes evidence from 113 studies examining anomaly detection in energy and related operational systems, with particular emphasis on Transformer-based and attention-driven approaches. The results demonstrate a clear shift toward Transformer-oriented modelling; however, this shift is neither uniform nor comprehensive across domains, data modalities, or operational settings.

Transformer-based architectures are primarily applied to multivariate time-series anomaly detection, where they effectively capture long-range dependencies in complex telemetry data, as demonstrated in studies such as [8], [12], [2], [50], [59], [6], [62], [5]. Their adoption is strongly associated with telemetry-driven applications, including SCADA-based monitoring, IoT systems, and sensor-based energy environments [7], [23], [73], [47], and [100], indicating that model selection is largely driven by data modality and temporal structure rather than architectural dominance.

However, the dominance of Transformer-based models is context-dependent. In image-driven anomaly detection tasks, particularly photovoltaic inspection, convolutional and object detection architectures remain prevalent. This is evident in studies such as [35], [76], [34], [55], and [101], where spatial feature extraction is prioritized over temporal modelling. This indicates that Transformer architectures are not displacing existing methods but rather complementing them in domains where temporal structure is critical.

4.2 Machine Models and Architectural Characteristics

Model selection across the dataset reflects a task-dependent modelling paradigm, where architectures are chosen based on data characteristics and operational requirements rather than architectural preference alone. Hybrid architecture represents the most prevalent design pattern, combining Transformers with CNNs, LSTMs, or optimization mechanisms to address multiple modelling requirements simultaneously.

Examples include [55], [18], [23], [24], [102], [46] and [75], where Transformer components are integrated with feature extractors or sequential models. This pattern demonstrates that current Transformer implementations are rarely sufficient as standalone models and are instead embedded within broader architectures.

Traditional models continue to play a significant role. LSTM-based approaches remain common in sequential anomaly detection tasks ([1], [20]), while CNN-based approaches dominate vision-based anomaly detection contexts ([35], [76]). Graph-based models appear less frequently but provide additional modelling capacity for structured systems ([21], [45]).

A critical limitation observed across the dataset is the limited exploration of efficiency-aware Transformer architectures. Most studies rely on full-attention models, which introduce high computational and memory overheads, particularly in streaming contexts. Only a small number of studies—such as [12] and [54] explicitly address computational efficiency, and even these remain constrained in scope.

This suggests that the field has prioritized representational capability over practical deployment, leaving a gap in architecture design for sustained, resource-constrained, real-time environments.

4.3 Model Performance and Validation

Table 13 Studies Reporting Key Performance Metrics.

Metric	Representative Studies
F1-score	[8], [59], [6], [102], [50], [47]
Accuracy	[55], [23], [52], [103]
AUROC	[26], [50], [6]
Precision / Recall	[24], [46], [25], [53]
Latency	[12], [54], [55]

Performance reporting across studies (Table 13) consistently indicates high predictive performance, with many models achieving F1-scores exceeding 0.90 (e.g., [8], [23], [46], [67]).

However, these results are primarily obtained under controlled or benchmark conditions, including datasets such as SMAP, SWaT, and SMD ([26], [6], [59]). As such, their applicability to real-world streaming environments remains uncertain.

Evaluation practices are also highly inconsistent. Classification metrics dominate, while probabilistic measures such as AUROC and reliability indicators such as false alarm rate are less frequently reported. Operational metrics, including latency and memory usage—are reported in only a small subset of studies ([12], [54]), limiting understanding of deployment feasibility.

In addition, external validation and reproducibility remain limited, with many studies relying on proprietary or simulated data. This restricts comparability and raises concerns regarding robustness and generalization.

4.4 Evidence Gaps Relative to Research Questions

The synthesis reveals several systemic gaps (Table 14) that directly constrain the development of reliable anomaly detection systems in non-stationary environments.

Table 14 Key Research Gaps Identified.

Gap Area	Evidence from Dataset	Implication
Streaming adaptation	Limited real streaming logic ([12], [22], [53], [10], [104], [79], [82])	Models not deployment-ready
Drift handling	Sparse implementation ([47], [46], [64], [106], [105])	Non-stationarity largely ignored
Architecture efficiency	Full attention dominant ([8], [2], [P121], [42], [80], [81])	High computational cost
Telemetry usage	Limited SCADA/inverter focus ([55], [2], [107], [110])	Reduced operational realism
Metric consistency	Inconsistent reporting across studies ([24], [50], [108])	Poor comparability
Interpretability	No operator validation ([67], [44], [109])	Limited practical utility
Dataset realism & generalization	Use of simulated and benchmark datasets ([1], [33], [92], [69])	Limited real-world transferability
Model generalization across domains	Domain-specific or narrow evaluation settings ([68], [91], [93], [88])	Weak cross-domain robustness

Despite claims of real-time capability, most models rely on static or pseudo-streaming setups with limited continuous adaptation. Concept drift remains poorly addressed, with only a few studies incorporating drift-aware mechanisms despite non-stationary operating conditions.

Data integration and usability gaps persist. Multimodal telemetry remains underutilized [40], [41], while interpretability is largely unvalidated in operational settings, limiting robustness and practical applicability.

4.5 Limitations of Evidence

The available evidence is characterized by substantial heterogeneity in datasets, methodologies, and reporting practices. Many studies rely on simulated or benchmark datasets ([1], [20]), limiting external validity.

Selective reporting is also evident, particularly in the omission of deployment-related metrics. Studies such as [55], [68], [69], [70], [66], [43], [67], and [23] report high performance without corresponding evaluation of computational feasibility or realistic deployment constraints.

The absence of standardized benchmarks further impedes comparison across studies.

4.6 Limitations of the Review Process

This review is limited to English-language studies published between 2023 and 2026. While a structured and reproducible methodology was employed, the reliance on published literature introduces potential publication bias.

Additionally, the heterogeneity of included studies prevented quantitative meta-analysis, limiting direct statistical comparison.

4.7 Implications and Forward Trajectories

4.7.1 Practice

The findings highlight that current Transformer-based models offer strong predictive performance but are not yet fully aligned with real-world deployment requirements [115]. Practitioners should therefore evaluate models not only in terms of accuracy but also in terms of adaptability, efficiency, and interpretability.

4.7.2 Policy

The lack of consistent reporting underscores the need for standardization in anomaly detection research. Reporting should include not only performance metrics but also computational efficiency, deployment conditions, and interpretability validation.

4.7.3 Future Research

The results indicate that future work should focus on:

- development of efficient and adaptive Transformer architectures
- integration of explicit concept drift detection and response mechanisms
- increased use of real-world, high-frequency telemetry data
- design of multimodal anomaly detection frameworks
- establishment of standardized evaluation protocols for streaming systems

5. Conclusion

Transformer-based models demonstrate strong potential for anomaly detection across time-series applications, particularly in capturing complex temporal dependencies. However, their use remains largely confined to controlled evaluation settings, with limited evidence of integration into real-world operational environments.

This review identifies a clear gap between model performance and practical applicability, driven by limited adaptation to non-stationary conditions, weak streaming validation, and inconsistent reporting of efficiency and interpretability. Advancing the field requires adaptive, well-validated approaches supported by standardized evaluation and real-world deployment.

Abbreviations

PRISMA
Preferred Reporting Items for Systematic Reviews and Meta–Analyses
CHARMS
Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling Studies
PROBAST
Prediction Model Risk of Bias Assessment Tool
PV
Photovoltaic
SCADA
Supervisory Control and Data Acquisition

IDS
Intrusion Detection System
IIoT
Industrial Internet of Things
CPS
Cyber–Physical Systems
IoT
Internet of Things
GNN
Graph Neural Network
GCN
Graph Convolutional Network
CNN
Convolutional Neural Network
RNN
Recurrent Neural Network
LSTM
Long Short–Term Memory
VAE
Variational Autoencoder
GAN
Generative Adversarial Network
AUC
Area Under the Curve
AUROC
Area Under the Receiver Operating Characteristic Curve
F1
Harmonic Mean of Precision and Recall
FAR
False Alarm Rate
FPR
False Positive Rate
MAE
Mean Absolute Error
RMSE
Root Mean Square Error
MAPE
Mean Absolute Percentage Error
mAP
Mean Average Precision
OSF
Open Science Framework

Declarations

Other Information

Registration and Protocol

Registration Information

The review protocol was registered on the Open Science Framework (OSF) under an embargo to maintain blinding during the initial screening phase. Registration DOI: 10.17605/OSF.IO/58GXK.

Protocol Access

The protocol is available on the Open Science Framework (see Availability of data and materials).

Amendments

No major amendments were made to the registered protocol. Minor clarifications were added regarding inclusion criteria for hybrid Transformer architectures and interpretability assessment, which were documented in the OSF registration notes.

Availability of data and materials

All data extraction templates, screening documents, search strategies, and scoring frameworks used in this systematic review are available on the Open Science Framework (OSF) at the following persistent identifier: https://doi.org/10.17605/OSF.IO/58GXK. This registration includes the complete protocol, search queries across all six databases (IEEE Xplore, SpringerLink, ScienceDirect, ACM Digital Library, Web of Science, and Google Scholar), data extraction template, and scoring framework with thresholds applied during study selection including:

Data Extraction Template (data extraction and screening) (Additional file 1).

Zotero Screenshots (Additional file 1).

Scoring Framework and Formula in Microsoft Excel (Appendix B).

Competing interests

The authors declare that they have no competing interests.

Funding

No funding was received for this review.

Authors' contributions

ZO conceived and designed the study, developed the search strategy, conducted database searches, screened and selected studies, extracted data, assessed risk of bias, performed the synthesis, and drafted the manuscript.

BS independently screened and selected studies, extracted data, assessed risk of bias, contributed to consensus decisions and interpretation, and critically revised the manuscript. All authors approved the final manuscript.

Acknowledgements

This systematic review was conducted in accordance with PRISMA 2020 guidelines and was informed by the CHARMS checklist for prediction modelling studies and the PROBAST tool for risk of bias assessment.

References

1. Liu X, Li H (2023) Data-Driven Cyberphysical Anomaly Detection for Microgrids With GFM Inverters. IEEE Open J Power Electron 4:498–511. https://doi.org/10.1109/OJPEL.2023.3290900
2. Maldonado-Correa J, Torres-Cabrera J, Martín-Martínez S, Artigao E, Gómez-Lázaro E (2024) Wind turbine fault detection based on the transformer model using SCADA data. Eng Fail Anal 162:108354. https://doi.org/10.1016/j.engfailanal.2024.108354
3. Zhou W, Chen C, Yan Q, Li B, Liu K, Zheng Y, Yang H, Xiao H, Su S (2025) Anomaly Usage Behavior Detection Based on Multi-Source Water and Electricity Consumption Information. IEEE Access 13:12215–12224. https://doi.org/10.1109/ACCESS.2025.3525726
4. Xiao Y, Wang Z, Yan T, Li W, Wu D, Ma J (2025) Time Series Anomaly Detection: A Hybrid Deep Learning Framework with Adaptive Thresholding for Electricity Industry Data. In: 2025 IEEE 12th International Conference on Cyber Security and Cloud Computing (CSCloud), pp 1–6. https://doi.org/10.1109/CSCloud66326.2025.00046
5. Dakheel F, Çevik M (2025) Optimizing Smart Grid Load Forecasting via a Hybrid Long Short-Term Memory-XGBoost Framework: Enhancing Accuracy, Robustness, and Energy Management. Energies 18(11):2842. https://doi.org/10.3390/en18112842
6. Zeng F, Chen M, Qian C, Wang Y, Zhou Y, Tang W (2023) Multivariate time series anomaly detection with adversarial transformer architecture in the Internet of Things. Future Generation Comput Syst 144:244–255. https://doi.org/10.1016/j.future.2023.02.015
7. Yin X, Zhang L, Liu Z, Qiu J, Wang C (2025) A Novel Transformer Model Enhanced by Scaled-CNN and Bi-LSTM Hybrid Networks for Real-Time Threat Identification Within SCADA Systems. IEEE Access 13:196579–196593. https://doi.org/10.1109/ACCESS.2025.3633662
8. Liu X, Li X, Li Y, Tang F, Zhao M (2025) RTdetector: Deep Transformer Networks for Time Series Anomaly Detection Based on Reconstruction Trend. In: IJCAI 2025 Proceedings. https://doi.org/10.24963/ijcai.2025/644
9. Yahya MA, Moya AR, Ventura S (2025) Deep learning for multivariate time series anomaly detection: An evaluation of reconstruction-based methods. Artif Intell Rev 58(12):400. https://doi.org/10.1007/s10462-025-11401-9
10. Neto ECP, Iqbal S, Buffett S, Sultana M, Taylor A (2025) Deep learning for intrusion detection in emerging technologies: A comprehensive survey and new perspectives. Artif Intell Rev 58(11):340. https://doi.org/10.1007/s10462-025-11346-z
11. Mishra AM, Yadav D, Shakya A, Jayesh V, Bala N (2025) A Hybrid Deep Learning Approach for Detecting Anomalies in Real-Time Data Streams. In: 2025 6th International Conference for Emerging Technology (INCET), pp 1–6. https://doi.org/10.1109/INCET64471.2025.11140026
12. Shi Q, Yang Z, Shao S (2026) TransEdge: A lightweight transformer-based anomaly detection framework for edge devices. J Intell Inform Syst. https://doi.org/10.1007/s10844-026-01043-w
13. Xu H, Boyaci A, Lian J, Wilson A (2026) Explainable AI for Multivariate Time Series Pattern Exploration: Latent Space Visual Analytics with Temporal Fusion Transformer and Variational Autoencoders in Power Grid Event

Diagnosis. IEEE Access 11:11744–11767. https://doi.org/10.1109/ACCESS.2025.3602635

14. Page MJ, McKenzie JE, Bossuyt PM et al (2021) The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. BMJ 372:n71. https://doi.org/10.1136/bmj.n71
15. Moons KGM, de Groot JAH, Bouwmeester W et al (2014) Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling Studies: The CHARMS Checklist. PLoS Med 11(10):e1001744. https://doi.org/10.1371/journal.pmed.1001744
16. Wolff RF, Moons KGM, Riley RD et al (2019) PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies. Ann Intern Med 170(1):51–58. https://doi.org/10.7326/M18-1376
17. Elkheir EIA, Abd-Allah MA, Hussien SA, Omar AI (2026) A Unified DWT-CNN-BiLSTM Framework for Integrated Fault Detection, Classification, and Feeder-Level Localization in Microgrids. IEEE Access. https://doi.org/10.1109/ACCESS.2026.3689777
18. Zhang C, Gong Y, Zhang Y, Liu S (2025) An Automatic Multi-Scale Feature Learner and a Transformer Algorithm for Fault Detection. Prot Control Mod Power Syst 10(5):84–102. https://doi.org/10.23919/PCMP.2024.000213
19. Jiang J, Wu H, Zhong C, Song H (2025) Classification of power quality disturbances in microgrids using a multi-level global convolutional neural network and SDTransformer approach. PLoS ONE 20(2):e0317050. https://doi.org/10.1371/journal.pone.0317050
20. Paudel A, Mandal P, Ravikumar G (2025) Forecast-Driven Detection and Mitigation of Cyber Attack in Microgrid Measurement Signal. In: 2025 IEEE Power & Energy Society General Meeting (PESGM), pp 1–5. https://doi.org/10.1109/PESGM52009.2025.11225384
21. Prasad U, Mohanty SR, Singh SP, Jagan A (2025) Robust Centralized Protection Scheme With AI-Based Fault Diagnosis Capabilities for Graph-Structured AC Microgrids. IEEE Trans Smart Grid 16(2):1975–1992. https://doi.org/10.1109/TSG.2024.3515050
22. Altafhusen T (2025) Cloud Sentry: Transformer-Based Anomaly Detection for Scalable and Autonomous Cloud Workload Protection. In: 2025 IEEE First International Conference on Innovations in Engineering and Next-Generation Technologies for Sustainability (ICINVENTS), pp 1–6. https://doi.org/10.1109/ICINVENTS64613.2025.11401421
23. Fraihat S, Yaseen Q, Sanjalawe Y, Abu-Errub A, Makhadmeh SN, Al-Betar MA (2026) Intrusion detection in industrial internet of things network using feature optimization and hybrid deep learning. Discover Internet Things 6(1):34. https://doi.org/10.1007/s43926-026-00284-z
24. Zhang Z, Rahman S (2026) Intrusion detection and mitigation system for smart inverters based on machine learning and environmental sensors. J Mod Power Syst Clean Energy 14(4):1527–1537. https://doi.org/10.35833/MPCE.2025.000713
25. Menaka A, Sakthivel K (2025) Improved Weighted Quantum Whale Optimization with vision transformer for intrusion detection, atmospheric monitoring and recommendation in smart agriculture. Sci Rep 15(1):40526. https://doi.org/10.1038/s41598-025-24425-6
26. Liu Y, Song Y, Shu S, Lin F, Wang J, Guo Y (2026) Adaptive Latent Distribution Modeling for Industrial Time Series Anomaly Detection. IEEE Trans Autom Sci Eng 23:7893–7907. https://doi.org/10.1109/TASE.2026.3674236
27. Nakashima T, Yairi T (2023) Assessing the Performance of Transformer for Time Series Anomaly Detection. In: PHM Society Asia-Pacific Conference. https://doi.org/10.36001/phmap.2023.v4i1.3773
28. Zamanzadeh Darban Z, Webb GI, Pan S, Aggarwal C, Salehi M (2025) Deep Learning for Time Series Anomaly Detection: A Survey. ACM-CSUR 57(1):1–42. https://doi.org/10.1145/3691338
29. Wen C, Wu X, Wang Z, Liu W, Yang J (2024) A novel local feature fusion architecture for wind turbine pitch fault diagnosis with redundant feature screening. Complex Intell Syst 10(6):8109–8125.

https://doi.org/10.1007/s40747-024-01584-z

30. Princz G, Shaloo M, Erol S (2024) Anomaly Detection in Binary Time Series Data: An unsupervised Machine Learning Approach for Condition Monitoring. Procedia Comput Sci 232:1065–1078. https://doi.org/10.1016/j.procs.2024.01.105
31. Song G, Ou J, Cao Y, Wang Y (2025) An energy auditing system based on learnable positional embedding and edge computing for in-situ monitoring of 3D printing processes. Int J Adv Manuf Technol 139(7–8):3493–3509. https://doi.org/10.1007/s00170-025-16098-2
32. Trapani N, Longo L (2024) Enhancing Feature Extraction in Sensor Fault Detection. IFAC-PapersOnLine 58(8):282–287. https://doi.org/10.1016/j.ifacol.2024.08.134
33. Najafzadeh M, Pouladi J, Daghigh A, Beiza J, Abedinzade T (2024) Fault Detection, Classification and Localization Along the Power Grid Line Using Optimized Machine Learning Algorithms. Int J Comput Intell Syst 17(1):49. https://doi.org/10.1007/s44196-024-00434-7
34. Lang D, Lv Z (2024) A PV cell defect detector combined with transformer and attention mechanism. Sci Rep 14(1):20671. https://doi.org/10.1038/s41598-024-72019-5
35. Chen A, Li X, Jing H, Hong C, Li M (2023) Anomaly Detection Algorithm for Photovoltaic Cells Based on Lightweight Multi-Channel Spatial Attention Mechanism. Energies 16(4):1619. https://doi.org/10.3390/en16041619
36. Li C, Liu D, Wang M, Wang H, Xu S (2023) Detection of Outliers in Time Series Power Data Based on Prediction Errors. Energies 16(2):582. https://doi.org/10.3390/en16020582
37. J J, K V (2025) A dual layer secure and energy-efficient model for border surveillance using sea lion inspired strategy in wireless sensor networks. Sci Rep 15(1):24974. https://doi.org/10.1038/s41598-025-07999-z
38. Parameswari M, Jeya Malar PN R (2025) Next generation AI powered framework for autonomous energy optimization and real time anomaly detection in IoT driven wireless sensor networks. Sci Rep 15(1):41104. https://doi.org/10.1038/s41598-025-24968-8
39. Hu J, Zhang G, Zhou C, Zhu H, Huang Y, Ou J (2024) Optimization Design and Performance Research of 5G P-IOT Passive Tags Based on LPWAN and Transformer in Long Range Coverage Transmission. In: 2024 4th International Conference on Computer Science, Electronic Information Engineering and Intelligent Control Technology (CEI), pp 288–294. https://doi.org/10.1109/CEI63587.2024.10871281
40. Pinho LS, Sousa TD, Pereira CD, Pinto AM (2025) Anomaly Detection for PV Modules Using Multi-Modal Data Fusion in Aerial Inspections. IEEE Access 13:88762–88779. https://doi.org/10.1109/ACCESS.2025.3570815
41. Zaboli A, Hong J, Ştefanov A, Liu C-C, Hwang C-S (2025) Large Language Models for Power System Security: A Novel Multi-Modal Approach for Anomaly Detection in Energy Management Systems. IEEE Access 13:203558–203585. https://doi.org/10.1109/ACCESS.2025.3636184
42. Luan S, Zhang J, Li Z, Liu H, Liu J, Li Z (2025) Research on Inter-Turn Short-Circuit Fault Detection and Diagnosis for Dry-Type Air-Core Reactors Based on Impulse Oscillation Method. In: 2025 IEEE 5th International Conference on Electrical Materials and Power Equipment (ICEMPE), pp 1–4. https://doi.org/10.1109/ICEMPE66159.2025.11122925
43. Kohli M, Chhabra I (2026) BALNet: Balanced Attack learning Network for Web Attack Detection via Adaptive Sampling, LDA, and CNNs. Discover Comput 29(1):253. https://doi.org/10.1007/s10791-026-10117-w
44. Ata MM, Osama S, Ibraheem MR, Abas AR (2026) Early prediction of wind turbine anomalies using 1D-CNN and temporal feature engineering on multi-source SCADA data. Sci Rep 16(1):9667. https://doi.org/10.1038/s41598-026-41571-7
45. Meng S, Li C, Tian C, Peng W, Tian C (2023) Transfer learning based graph convolutional network with self-attention mechanism for abnormal electricity consumption detection. Energy Rep 9:5647–5658.

https://doi.org/10.1016/j.egyr.2023.05.006

46. Aly M, Alotaibi NM (2025) SHAP enhanced transformer GWO boosting model for transparent and robust anomaly detection in IIoT environments. Sci Rep 15(1):41037. https://doi.org/10.1038/s41598-025-25033-0
47. Chen B, Huang V, Wang C (2024) Sensor-Drift-Aware Time-Series Anomaly Detection for Climate Stations. In: 2024 IEEE Conference on Artificial Intelligence (CAI), pp 1290–1295. https://doi.org/10.1109/CAI59869.2024.00228
48. Liu S, Jiang P, Wang P, Li L, Chen X, Yin S (2024) An Anomaly Detection Method for Provincial-side Base Operation Logs Based on VAE and Transformer. In: 2024 10th International Conference on Systems and Informatics (ICSAI), pp 1–8. https://doi.org/10.1109/ICSAI65059.2024.10893801
49. Xie Z, Xu H, Chen W, Li W, Jiang H, Su L, Wang H, Pei D (2023) Unsupervised Anomaly Detection on Microservice Traces through Graph VAE. In: Proceedings of the ACM Web Conference 2023, pp 2874–2884. https://doi.org/10.1145/3543507.3583215
50. Maleki S, Pourmoazemi N (2026) Pi-transformer: A prior-informed dual-attention model for multivariate time-series anomaly detection. Appl Soft Comput 195:115029. https://doi.org/10.1016/j.asoc.2026.115029
51. Lv X, Cui S, Wang Y, Lu J, Yu P, Wang K (2026) Patch Time Series Transformer–Based Short–Term Photovoltaic Power Prediction Enhanced by Artificial Fish. Energies 19(1):284. https://doi.org/10.3390/en19010284
52. Hamida IB, Bera PK, Al-Saadi T, Pani SR, Mansouri M (2026) Intelligent Detection of High Impedance Faults in Microgrid Distribution Lines Using Optimized Machine Learning Models. IEEE Access. https://doi.org/10.1109/ACCESS.2026.3691380
53. Cao Y (2025) Intelligent Anomaly Detection with Attention-Based Fusion of Multi-Source Heterogeneous Data. In: Proceedings of the 2025 6th International Conference on Computer Science and Management Technology, pp 590–595. https://doi.org/10.1145/3795154.3795248
54. Alabdulkreem E, Mansouri W, Deepika KK, AlAqil MA, Abaker M, Kavitha M, Priya V, Masih RK (2026) Transformer-based neural optimization for energy-efficient event detection in wide-area power monitoring systems. Int J Inf Secur 25(3):93. https://doi.org/10.1007/s10207-026-01243-z
55. Wang N, Huang S, Liu X, Wang Z, Liu Y, Gao Z (2025) MRA-YOLOv8: A Network Enhancing Feature Extraction Ability for Photovoltaic Cell Defects. Sensors 25(5):1542. https://doi.org/10.3390/s25051542
56. ElMahdy OFM, Hassan ME, Metwalli SM (2025) Machine learning anomaly detection of lost and unaccounted for gas in natural gas networks. J Eng Appl Sci 72(1):123. https://doi.org/10.1186/s44147-025-00677-x
57. Bhattacharya S, Roig Greidanus MD, Gupta S, Sur D, Mazumder SK, Govindarasu M (2024) Machine-Learning-based Anomaly Detection System for Cyber-Physical Attack on a Solid-State Transformer. In: IECON 2024–50th Annual Conference of the IEEE Industrial Electronics Society, pp 1–6. https://doi.org/10.1109/IECON55916.2024.10905322
58. Khawaji A, Martin RJ (2026) ML-Based Threat Severity Detection in Cyber-Physical Systems for Smart Environments Using Behavioral Analysis. Human-Centric Intell Syst 6(1):48–65. https://doi.org/10.1007/s44230-026-00134-9
59. Shin A-H, Kim ST, Park G-M (2023) Time Series Anomaly Detection Using Transformer-Based GAN With Two-Step Masking. IEEE Access 11:74035–74047. https://doi.org/10.1109/ACCESS.2023.3289921
60. Mogi IKA, Shiddiqi AM, Pratomo BA, Za'in C, Ismail M, Manalu AAH (2025) Maelnet: Memorable Anomaly Election Learning Detecting Anomalies in Time Series Data Using Dual-Net Transformer. In: 2025 International Conference on Artificial Intelligence for Sustainable Innovation (AI-SI), pp 1–6. https://doi.org/10.1109/AI-SI66213.2025.11341688
61. Abid T, Ahmim A, Maazouzi F, Chefrour D, Ullah I, Ahmim M, Almukhlifi R (2025) A novel IoT threat detection using GWO feature selection and CNN-enhanced LightGBM. J Cloud Comput 14(1):72. https://doi.org/10.1186/s13677-

025-00785-2

62. Yang X, Howley E, Schukat M (2024) ADT: Time series anomaly detection for cyber-physical systems via deep reinforcement learning. Computers Secur 141:103825. https://doi.org/10.1016/j.cose.2024.103825
63. Nivarthi CP, Huang Z, Gruhl C, Sick B (2024) Multi-Task Representation Learning with Temporal Attention for Zero-Shot Time Series Anomaly Detection. In: 2024 International Joint Conference on Neural Networks (IJCNN), pp 1–10. https://doi.org/10.1109/IJCNN60899.2024.10651344
64. Bhattacharya S, Greidanus MDR, Gupta S, Sur D, Mazumder SK, Govindarasu M (2025) Machine-Learning-Based Adaptive Anomaly Detection for Control Feedback Interferences in Solid-State Transformers. IEEE J Emerg Sel Top Industrial Electron 6(4):1840–1852. https://doi.org/10.1109/JESTIE.2025.3589164
65. Zou Z, Long Z, Xu G, Farmani R, Yu T, Chu S (2025) Learning physics and temporal dependencies: Real-time modeling of water distribution systems via Kolmogorov–Arnold attention networks. Npj Clean Water 8(1):75. https://doi.org/10.1038/s41545-025-00505-y
66. Zhu H, Zhang G, Huang Y, Li S, Xie J, Chen B (2026) Anomaly-Aware Dual-Stream Network for Intelligent Operation Quality Assessment in 5G Wireless Public Networks. In: 2026 6th International Conference on Consumer Electronics and Computer Engineering (ICCECE), pp 345–349. https://doi.org/10.1109/ICCECE69169.2026.11399800
67. Kayode Saheed Y, Ebere Chukwuere J (2025) CPS-IIoT-P2Attention: Explainable Privacy-Preserving With Scaled Dot-Product Attention in Cyber-Physical System-Industrial IoT Network. IEEE Access 13:81118–81142. https://doi.org/10.1109/ACCESS.2025.3566980
68. Ismail A, Aboshady D (2025) A novel lightweight dual-stream recurrent transformer model for anomaly detection in driving dashcam videos. Neural Comput Appl 37(22):18527–18548. https://doi.org/10.1007/s00521-025-11380-6
69. Nevisi MMS, Shoeibi M, Hernando-Gallego F, Martín D, Khatami SS (2025) An Evolutionary Deep Reinforcement Learning-Based Framework for Efficient Anomaly Detection in Smart Power Distribution Grids. Energies 18(10):2435. https://doi.org/10.3390/en18102435
70. Nandhini U, Kumar SVNS (2026) An Enhanced Horse Herding Optimizer-Driven Feature Selection Approach Integrated with CNN for Network Intrusion Detection Systems. Int J Comput Intell Syst 19(1):89. https://doi.org/10.1007/s44196-026-01167-5
71. Zhu L, Gao J, Zhu C, Deng F (2025) Short-term power load forecasting based on spatial-temporal dynamic graph and multi-scale Transformer. J Comput Des Eng 12(2):92–111. https://doi.org/10.1093/jcde/qwaf013
72. Wang C, Erfani S, Alpcan T, Leckie C (2025) OIL-AD: An anomaly detection framework for decision-making sequences. Pattern Recogn 166:111656. https://doi.org/10.1016/j.patcog.2025.111656
73. Liu Z, Wang Y, Wang Q, Hu M (2025) Vision Transformer-Based Anomaly Detection in Smart Grid Phasor Measurement Units Using Deep Learning Models. IEEE Access 13:44565–44576. https://doi.org/10.1109/ACCESS.2025.3549679
74. Aasar T, Sultan K, Khan AU, Czanner S, Abbasi A, Ahmad A (2026) Predicting stator winding short-circuit faults in induction motors using machine learning: A comparative study. Discover Comput 29(1):103. https://doi.org/10.1007/s10791-026-09984-0
75. Guo Y, Tang Y, Pan W (2024) Early fault detection in distribution networks based on wavelet packets and anomaly transformer. In: IEEE 36th Chinese Control and Decision Conference (CCDC). https://doi.org/10.1109/CCDC62350.2024.10588165
76. Ayadi N, Et-Taleby A, Chaibi Y, Salem CE, Benslimane M, Chalh Z (2024) Photovoltaic hotspot fault detection based on detectron2 with faster R-CNN. In: 2024 3rd International Conference on Embedded Systems and Artificial Intelligence (ESAI), pp 1–12. https://doi.org/10.1109/ESAI62891.2024.10913808

77. Lee W, Ye T (2025) Multi-Scale Adaptive Attention for Time Series Anomaly Detection. In: Proceedings of the 2025 7th Asia Conference on Machine Learning and Computing, pp 65–68. https://doi.org/10.1145/3772673.3772696
78. Wang Z, Pei C, Ma M, Wang X, Li Z, Pei D, Rajmohan S, Zhang D, Lin Q, Zhang H, Li J, Xie G (2024) Revisiting VAE for Unsupervised Time Series Anomaly Detection: A Frequency Perspective. In: Proceedings of the ACM Web Conference 2024, pp 3096–3105. https://doi.org/10.1145/3589334.3645710
79. Chen W, Luo Y, Li L, Li J (2024) PV-DETR: A Multimodal Fault Detection Model of PV Arrays based on Parallel Block Attention. In: 2024 IEEE International Conference on Big Data (BigData), pp 3609–3615. https://doi.org/10.1109/BigData62323.2024.10826130
80. Śmiałkowski T, Czyżewski A (2025) Anomaly detection in urban lighting systems using autoencoder and transformer algorithms. Sci Rep 15(1):35972. https://doi.org/10.1038/s41598-025-19414-8
81. Guo Y, Tang Y, Pan W (2025) State Monitoring of Power Distribution Systems Based on ZigBee Technology and Graph Attention Networks. IEEE Access 13:184904–184917. https://doi.org/10.1109/ACCESS.2025.3625976
82. Zhang W, Zhao Z, Chen Y, Wu X (2026) Transformer fault diagnosis in modern power systems using a genetic algorithm-optimized neural network and dissolved gas analysis. Discover Internet of Things 6(1):51. https://doi.org/10.1007/s43926-026-00296-9
83. Zheng X, Jiang Y (2026) SFG: A Universal Time-Frequency Architecture for Versatile Time Series Analysis. IEEE Access 14:64706–64720. https://doi.org/10.1109/ACCESS.2026.3687534
84. Aadithiya BN, K PK KP (2026) S PK, M RS Smart Transformer Phase Failure and Geo-Alert System. In: 2026 5th International Conference on Communication, Computing and Electronics Systems (ICCCES), pp 288–293. https://doi.org/10.1109/ICCCES62661.2026.11436398
85. Wang Y, Wang T, Wang W, Hou Y (2024) Collaborative Application of Deep Learning Models for Enhanced Accuracy and Prediction in Carbon Neutrality Anomaly Detection. J Organizational End User Comput 36(1):1–25. https://doi.org/10.4018/JOEUC.340385
86. Wang J, Zhao Y, Zheng J, Feng Y, Wen L, Huang H, Wang Y (2026) PerSANet: Synergizing Multi-Scale Inception and Energy-Based Attention for Time Series Anomaly Detection. IEEE Access 14:39483–39495. https://doi.org/10.1109/ACCESS.2026.3662023
87. Kabir S, Chowdhury M, Zishan MSR (2025) Proactive Detection of Cyber-Physical Grid Attacks: A Pre-Attack Phase Identification and Analysis Using Anomaly-Based Machine Learning Models. https://doi.org/10.21203/rs.3.rs-5828184/v1. Research Square (preprint)
88. Zhang P, Yang Y, Song R, Wang B, Sheng J, Zhao B (2024) Line Loss Anomaly Perception Method Based on MIC-IF Algorithm for Photovoltaic Low-Voltage Transformer Area. IEEE Access 12:108145–108153. https://doi.org/10.1109/ACCESS.2024.3437368
89. Etemadi A, Farahnak-Ghazani M, Arjmandi H, Mirmohseni M, Nasiri-Kenari M (2023) Abnormality Detection and Localization Schemes Using Molecular Communication Systems: A Survey. IEEE Access 11:1761–1792. https://doi.org/10.1109/ACCESS.2022.3228618
90. Fişne A, Yurtsever MME, Eken S (2024) Energy-efficient buildings with energy-efficient optimized models: A case study on thermal bridge detection. Cluster Comput 27(9):12787–12797. https://doi.org/10.1007/s10586-024-04624-y
91. Liu Y (2025) Intelligent Analysis Methods for Multi-channel Marketing Data Based on Anomaly Detection Algorithms. In: Proceedings of the 2nd International Conference on Image Processing, Machine Learning, and Pattern Recognition, pp 198–206. https://doi.org/10.1145/3759928.3759963
92. Fornás JG, Jaraba EH, Estopiñan AL (2023) TransSiamese: A Transformer-Based Siamese Network for Anomaly Detection in Time Series as Approach for Fault Location in Distribution Grids. IEEE Access 11:103431–103451.

https://doi.org/10.1109/ACCESS.2023.3316600

93. Yu Y, Dizha M, Zhang B, Wen R, Luo F, Guo X, Song J, Wang B, Zhang Z (2025) Computers Mater Continua 85(2):3861–3877. https://doi.org/10.32604/cmc.2025.066097. DH-LDA: A Deeply Hidden Load Data Attack on Electricity Market of Smart Grid
94. Zhang C, Zhang J, Zhao Y, Lu J (2024) Automated data mining framework for building energy conservation aided by generative pre-trained transformers (GPT). Energy Build 305:113877. https://doi.org/10.1016/j.enbuild.2023.113877
95. Zhao J, Chu F, Xie L, Che Y, Wu Y, Burke AF (2026) A survey of transformer networks for time series forecasting. Comput Sci Rev 60:100883. https://doi.org/10.1016/j.cosrev.2025.100883
96. Abdullahi S, Usman Danyaro K, Zakari A, Abdul Aziz I, Amila Wan Abdullah Zawawi N, Adamu S (2025) Time-Series Large Language Models: A Systematic Review of State-of-the-Art. IEEE Access 13:30235–30261. https://doi.org/10.1109/ACCESS.2025.3535782
97. Xu X, Huang X, Bian H, Wu J, Liang C, Cong F (2024) Total process of fault diagnosis for wind turbine gearbox, from the perspective of combination with feature extraction and machine learning: A review. Energy AI 15:100318. https://doi.org/10.1016/j.egyai.2023.100318
98. Zhang J, Yin C, Farbiz F, Jafary-Zadeh M, Sing SL (2025) Advancing machine learning applications for in-situ monitoring and control in laser-based metal additive manufacturing: A state-of-the-art review. Virtual Phys Prototyp 20(1):e2592732. https://doi.org/10.1080/17452759.2025.2592732
99. Zheng R, Sumper A, Aragüés-Peñalba M, Galceran-Arellano S (2024) Advancing Power System Services With Privacy-Preserving Federated Learning Techniques: A Review. IEEE Access 12:76753–76780. https://doi.org/10.1109/ACCESS.2024.3407121
100. Pham LNH, Gonzalez-Longatt F, Riquelme-Dominguez J (2024) Exploring Multiple Functionalities of Cyber-Physical Power System within Digital Substation-Based Laboratory Environment. SSRN. https://doi.org/10.2139/ssrn.5029679
101. Yin W, Lingxin S, Maohuan L, Qianlai S, Xiaosong L (2023) PV-YOLO: Lightweight YOLO for Photovoltaic Panel Fault Detection. IEEE Access 11:10966–10976. https://doi.org/10.1109/ACCESS.2023.3240894
102. John J, Raj RDA, Pallakonda A, Yanamala RMR, Pouresmaeil E, Aghaei J (2026) Deep Learning-Based Anomaly Detection and Authenticated Encryption Framework for PMU Data in Industrial Cyber-Physical Systems. IEEE Trans Industr Inf 1–12. https://doi.org/10.1109/TII.2026.3673230
103. Ibrahim Abo Elkheir E-S, Abd-Allah MA, Aly Hussien S, Omar AI (2026) Real-Time Microgrid Fault Management: Leveraging Hybrid Deep Learning for Superior Detection and Classification Performance. IEEE Access 14:9606–9623. https://doi.org/10.1109/ACCESS.2026.3654266
104. Usmani UA, Abdul Aziz I, Jaafar J, Watada J (2024) Deep Learning for Anomaly Detection in Time-Series Data: An Analysis of Techniques, Review of Applications, and Guidelines for Future Research. IEEE Access 12:174564–174590. https://doi.org/10.1109/ACCESS.2024.3495819
105. Han X, Lan P, Zhao Q, Li G, Zhao S, Sun F (2025) Phase Error Monitoring for CVT Based on Mode Decomposition and Local Outlier Detection. In: 2025 6th International Conference on Smart Grid and Energy Engineering (SGEE), pp 464–469. https://doi.org/10.1109/SGEE68429.2025.11385792
106. Olasehinde DO, Bamisile O, Ejiyi CJ, Zhang G, Cai D, Li J, Wei L, Huang Q (2026) Cybersecurity in cyber-physical power systems: Analyzing vulnerabilities, threats, and control structures. Cluster Comput 29(3):133. https://doi.org/10.1007/s10586-025-05894-w
107. Seman LO, Carolina Rodrigues Klaar A, Henrique Dal Molin Ribeiro M, Frizzo Stefenon S (2025) Enhanced Random Vector Functional Link Networks With Bayesian-Based Hyperparameter Optimization for Wind Speed Forecasting.

IEEE Access 13:208105–208122. https://doi.org/10.1109/ACCESS.2025.3640434

108. Zhang G, Wei C, Jing C, Wang Y (2022) Short-Term Electrical Load Forecasting Based on Time Augmented Transformer. Int J Comput Intell Syst 15(1):67. https://doi.org/10.1007/s44196-022-00128-y
109. Jahromi MZ, Khalaf M, Kassouf M, Kundur D (2024) Towards a more secure reconstruction-based anomaly detection model for power transformer differential protection. Front Energy Res 12:1444697. https://doi.org/10.3389/fenrg.2024.1444697
110. Saravana MK, Roopa MS, Arunalatha JS, Venugopal KR (2026) Transformers for Multivariate Time Series Forecasting: Comprehensive Analysis, Challenges, Research Opportunities, and Future Prospects. IEEE Access 14:11424–11457. https://doi.org/10.1109/ACCESS.2026.3654408
111. Chithra Rani PR, Baalaji K (2025) Deep learning-based ensemble stacking for enhanced intrusion detection in IoT-edge platforms. Discover Appl Sci 7(8):928. https://doi.org/10.1007/s42452-025-06871-z
112. Gao Q, Shen L, Shi J, Gu X, Gu S, Ge Y, Xie Y, Zhu X, Zang B, Zhang M, Nazir MS, Ji J (2025) Transformer-Enhanced Intelligent Microgrid Self-Healing: Integrating Large Language Models and Adaptive Optimization for Real-Time Fault Detection and Recovery. Energy Eng 122(7):2767–2800. https://doi.org/10.32604/ee.2025.065600
113. Liso A, Cardellicchio A, Patruno C, Nitti M, Ardino P, Stella E, Reno V (2024) A Review of Deep Learning-Based Anomaly Detection Strategies in Industry 4.0 Focused on Application Fields, Sensing Equipment, and Algorithms. IEEE Access 12:93911–93923. https://doi.org/10.1109/ACCESS.2024.3424488
114. Piantadosi G, Dutto S, Galli A, De Vito S, Sansone C, Di Francia G (2024) Photovoltaic power forecasting: A Transformer based framework. Energy AI 18:100444. https://doi.org/10.1016/j.egyai.2024.100444
115. Huang C, Zhao F, Xiong H, Luo H (2025) Review of Series Arc Fault Detection Methods in Photovoltaic Systems Based on AI. In: 2025 2nd International Conference on Electronic Engineering and Information Systems (EEISS), pp 1–5. https://doi.org/10.1109/EEISS65394.2025.11085914
116. Satpute K, Maurya SK, Mohapatra A (2024) Unique Transformer Neural Network Based Line Fault Detection Technique in DC Microgrid. In: 2024 23rd National Power Systems Conference (NPSC), pp 1–6. https://doi.org/10.1109/NPSC61626.2024.10987025

Figures

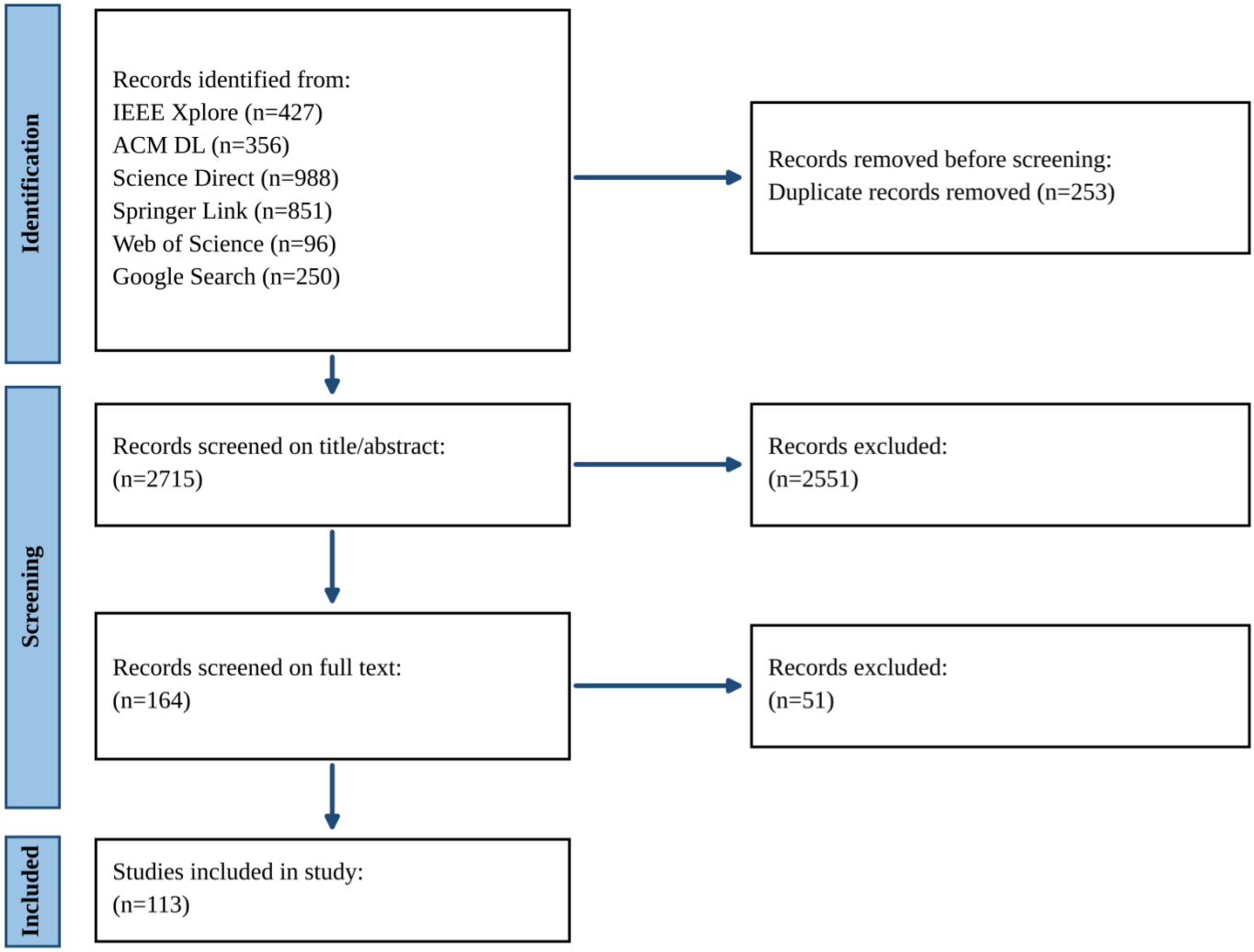


Figure 1

PRISMA flowchart for study selection.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Appendix.docx