ISSN E: 3006-1466
ISSN P: 3006-1458
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
CJSSR
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
Vol.04 No.02 (2026)

ANOMALY DETECTION IN IOT SENSOR DATA USING MACHINE LEARNING FOR REAL-TIME MONITORING SYSTEMS

Fahad Naeem
Assistant Professor, Government Ghazali College Latifabad , Hyderabad,
MS (Computer Science).
Muhammad Zubair,
Lecturer, Government Degree College Tandojam, Hyderabad
MS (Computer Science).
Muhammad Abdul Haseeb,
Head IT Governance, United Bank Ltd.
MS (Computer Science).

ABSTRACT

The high rate of growth of Internet of Things (IoT) devices in the industrial, healthcare, and smart city infrastructure sectors has produced new amounts of sensor data that need to be constantly analyzed in real-time. The overall challenge of identifying anomalies in these high-velocity, heterogeneous data streams is basic with important consequences of operational safety, system dependability and predictive maintenance. This paper is a systematic review of machine learning-based anomaly detection systems in IoT sensor setups. A systematic literature search was done in the IEEE Xplore, ACM Digital Library, ScienceDirect and Google Scholar, to identify publications published between 2014 and 2024, and a final corpus of 47 primary studies were identified using specific inclusion and exclusion criteria. The review is conducted on the statistical models, classical machine learning models, and deep learning architectures, and federated learning frameworks and their performance is assessed on the standard benchmark datasets, such as SWaT, SMD, SMAP, and MSL. Specific applications of industrial IoT, healthcare monitoring, and smart city systems are discussed in separate sections. The challenges in deployment such as constraints on real-time latency, concept drift, class imbalance, and integration of edge computing are discussed. The results show that both hybrid and federated deep learning models provide the most promising direction to scalable, privacy-preserving, and low-latency anomaly detection in commercial IoT systems, although interpretability and adversarial robustness are still open research problems.

INDEX TERMS

anomaly detection, Internet of Things, machine learning, deep learning, LSTM, autoencoder, real-time monitoring, federated learning, edge computing, sensor data.

1. Introduction

The Internet of Things is one of the most radical technological changes of the twenty-first century, which links billions of physical objects, such as industrial sensors and wearable health monitors, smart home appliances, autonomous vehicles, etc., to data ecosystems. Statista (2024) forecasts that the number of IoT-connected devices worldwide will exceed 29 billion by the year 2030, with data volumes growing exponentially, which is much more than traditional monitoring methods can handle. In this landscape, the ability to identify anomalous patterns in real time has become a very important engineering/scientific priority.

IoT sensor data anomalies take a variety of forms: point anomalies are individual outlier values; contextual anomalies are odd readings but against their temporal or environmental background; collective anomalies are series of readings which become suspicious on the aggregate. The impacts of not detecting such anomalies are also significant — in industrial environments, unnoticed sensor failures may cause a chain of equipment failures; in healthcare, undiagnosed physiological alerts can be life-threatening; and in the case of smart city infrastructure, unnoticed anomalies in power grids or water systems may trigger mass disruptions to services (Cook et al., 2020; Zhu et

ISSN E: 3006-1466
ISSN P: 3006-1458
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
CJSSR
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
Vol.04 No.02 (2026)

The conventional rule-based and threshold-based methods of detecting anomalies have inherent weaknesses in the IoT setting. They are specialist in nature, have difficulty with generalization across different types of heterogeneous devices, and are fragile to dynamic operating conditions or concept drift the gradual change in underlying data distributions over time (Gama et al., 2014). Machine learning techniques provide the possibility of learning complex, non-linear patterns directly on the data, and retraining to adapt to changing conditions.

The last ten years have seen a surge in the research on machine learning to detect anomalies in IoT, both classical models, including One-Class Support Vector Machines (OC-SVM) and Isolation Forest, and deep learning models, including autoencoders, recurrent neural networks, and transformer models, as well as new paradigms such as federated learning that must overcome privacy constraints of distributed sensor networks (Pang et al., 2 Nonetheless, the fact that these techniques are practically implemented in resource-constrained edge computing systems, where memory, computational power, and communications bandwidth is constrained, adds latency and efficiency concerns that remain largely unexplored in the theoretical literature.

The following is the structure of this article: Section 2 presents the systematic review methodology. Section 3 contains a survey of related surveys and previous work. Section 4 gives background taxonomy of anomalies. Part 5 reviews classical machine learning methods. Part 6 looks at deep learning architectures. Section 7 talks about federated and distributed approaches. Section 8 deals with deployment issues of real time and latency. In section 9, domain-specific applications are given. Section 10 presents performance comparison with benchmarks of datasets. There are open problems and future directions discussed in Section 11, then limitations in Section 12 and conclusions in Section 13.

2. Systematic Review Methodology

2.1 Search Strategy

This review adhered to the guidelines of the Preferred Reporting Items of Systematic Reviews and Meta-Analyses (PRISMA) (Page et al., 2021) but in the context of the computer science literature. A systematic search in four major academic databases was conducted, including IEEE Xplore, ACM Digital Library, Elsevier ScienceDirect, and Google Scholar. The search was limited to peer-reviewed journal articles, conference proceedings, and book chapters dating back between January 2014 and December 2024 to encapsulate changes throughout the modern deep learning period but provide currency.

The main search query was as follows: (anomaly detection) (outlier detection) (fault detection) (Internet of Things) (IoT) (sensor data) (sensor network) (machine learning) (deep learning) (neural network) (autoencoder) (LSTM) (federated learning). Variations were made to the databases to match the dissimilarity in indexing terms.

2.2 Inclusion and Exclusion Criteria

We included studies that: (1) reported or proposed machine learning or statistical approaches to anomaly detection in IoT or sensor-generated time-series data; (2) included quantitative performance measures such as precision, recall, F1-score or AUC-ROC; (3) were published in English in peer-reviewed journals or conferences; and (4) offered sufficient detail to assess reproducibility. Studies were excluded if they: (1) focused solely on anomaly detection in network traffic; (2) involved only simulation; (3) were published in predatory journals (identified by Beall's List); or (4) were a re-published version of a previously included conference paper.

2.3 Study Selection and Data Extraction

A search of the databases identified 1,847 potential records. This deduplicated to 1,204 records. Screening on title and abstract led to 312 full-text articles. Finally, 47 primary studies

ISSN E: 3006-1466
ISSN P: 3006-1458
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
CJSSR
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
Vol.04 No.02 (2026)

were identified following full-text review using inclusion and exclusion criteria, complemented by 18 foundational studies cited by primary studies for methodological background. We extracted the following from each study: algorithm family and type, dataset used, the metrics employed for evaluation, hardware platform, real-time constraints, and application domain. The selection process is summarized in Figure 1 (textual description of PRISMA flow diagram below for compatibility).

PRISMA Flow Summary: Records identified through databases (n = 1,847) → After deduplication (n = 1,204) → After title/abstract screening (n = 312) → After full-text screening (n = 47 primary studies included).

2.4 Quality Assessment

All included studies were scored on a six-point quality list adapted from the STROBE and CONSORT reporting guidelines: (1) clear problem definition; (2) justified data selection; (3) reproducible experimentation protocol; (4) control group comparison; (5) statistical significance assessment and (6) discussion of deployment limitations. Any studies with a score of less than three were only included if they contained novel methodological insights not found in other studies.

3. Related Work and Prior Surveys

This article builds on several previous surveys exploring various aspects of anomaly detection for IoT and time-series data, and the authors explicitly position this article in relation to these to place this work in context.

Chandola et al. (2009) conducted the first comprehensive survey of anomaly detection techniques for any type of data, and defined the taxonomical classification of point, contextual and collective anomalies. But its pre-IoT focus means it does not consider specific challenges of embedded sensor platforms and edge computing, nor real-time latency constraints.

Cook et al. (2020) focused on anomaly detection for IoT time series data, addressing statistical, machine learning and deep learning approaches in the health care, smart home, and industrial sectors. Comprehensive at the time, it was written before the emergence of transformers and federated learning which have since grown in popularity. This article builds on Cook et al. by including these advances, and explicit benchmark results that were not available in 2020.

Pang et al. (2021) offered a thorough review of deep learning for anomaly detection with general data types, such as images and graphs. They provided a more general survey, but did not focus on the specific constraints of IoT deployments; specifically, edge inference time, communication-bounded federated learning, and real-time anomaly alerting. This review is focused solely on IoT sensors and considers deployment factors not considered by general deep learning surveys.

Goldstein and Uchida (2016) empirically compared unsupervised anomaly detection algorithms but limited their study to static (not temporal) tabular data, not including the temporal considerations that are prevalent in IoT sensor data. More recently, Schmidl et al. (2022) benchmarked 158 time-series anomaly detection algorithms on 67 datasets using the TimeEval framework, which was the most comprehensive algorithmic benchmarking to date. This review builds on their empirical data, but places it in the context of IoT deployment constraints that are not considered in their benchmark.

Compared to these previous works, the current review contributes in four distinct ways: (1) an explicit systematic review process complying with PRISMA guidelines; (2) a federated learning/edge deployment perspective; (3) domain-specific considerations and analyses in industrial, healthcare, and smart city domains; and (4) a specific focus on real-time latency as a design consideration.

ISSN E: 3006-1466
ISSN P: 3006-1458
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
CJSSR
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
Vol.04 No.02 (2026)

4. Background and Taxonomy of IoT Anomalies

4.1 Characteristics of IoT Sensor Data

IoT sensor data is different from the typical data sets in several ways. First, IoT data is dynamic - the measurements are time-stamped and highly correlated, that is, the context of a measurement is essential to its meaning. Second, IoT streams are often multi-variate - hundreds of simultaneous measurements are obtained from multiple channels, with intricate interactions. Third, the quality of the data is often degraded by sensor drift, communication packet loss and hardware failures, which cause noise and missing data that must be separated from true anomalies (Buda et al., 2018).

IoT systems are also highly diverse. Smart factories can include vibration sensors, thermal cameras, acoustic emission sensors, pressure sensors, and each uses different sampling frequencies, units, and ranges. This poses challenges for building generic anomaly detection models, and requires either highly flexible model designs or domain-specific tuning (Zhang et al., 2022).

4.2 Anomaly Taxonomy

The original taxonomy of Chandola et al. (2009) defines three main types of anomalies. Point anomalies are individual data points that differ from the rest (for instance, a temperature of 200°C for a sensor that normally measures up to 80°C). Collective anomalies are sequences of data points that exhibit unusual patterns while individual points may not be anomalous.

Contemporary classifications also include system-level anomalies (entire device nodes are affected), network-level anomalies (unusual communication patterns) and application-level anomalies (deviations in functional behavior). Goldstein and Uchida (2016) also highlight the importance of distinguishing between supervised anomaly detection (using labeled data) and unsupervised anomaly detection. Since real-world IoT deployments generally lack extensive labeled anomaly data, unsupervised and semi-supervised learning techniques are often used.

4.3 Evaluation Metrics

Evaluating anomaly detection involves selecting appropriate metrics because of the extreme class imbalance - anomalies could be less than 0.1% of data. Accuracy is not an appropriate metric. It is better to use area under the receiver operating characteristic curve (AUC-ROC), precision-recall, F1-score and Matthews Correlation Coefficient (MCC). For real-time monitoring, detection latency - time between anomaly and alert - is also a vital operational measure to supplement classification accuracy (Davis & Goadrich, 2006).

5. Classical Machine Learning Approaches

5.1 Statistical Baseline Methods

Statistical approaches are the historical basis of anomaly detection, and are still widely used. The first to have been used were z-score thresholding, control charts and autoregressive integrated moving average (ARIMA) models. ARIMA captures the correlations of sensor streams and raises alarms for residuals outside confidence limits. Although cheap and easy to interpret, these approaches require stationarity and linearity, and are therefore not effective in the dynamic IoT environment (Braei & Wagner, 2020).

5.2 Isolation Forest

Isolation Forest (Liu et al., 2008) is one of the most successful practically, unsupervised anomaly detection algorithms for high-dimensional sensor data streams. This method iteratively splits the feature space with random cuts, and counts the number of splits needed to isolate each data point. Since anomalies are rare and lie away from the data mass, they are isolated in fewer splits, resulting in a smaller anomaly score. It has a linear time complexity, and is therefore suitable for real-time IoT streams. The Extended Isolation Forest (Hariri et al., 2019) is a variant of the algorithm that is distributional robust, since it uses arbitrary hyperplanes in place of cuts.

ISSN E: 3006-1466
ISSN P: 3006-1458
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
CJSSR
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
Vol.04 No.02 (2026)

5.3 One-Class Support Vector Machines

One-Class SVMs (Schölkopf et al., 2001) train a decision boundary in the data space, mapped by a kernel, around the distribution of the normal class to classify as anomalies those outside the boundary. OC-SVM is effective for low-dimensional static IoT settings, but is not scalable and is sensitive to hyperparameters. The related Support Vector Data Description (Tax & Duin, 2004) has been applied to industrial monitoring due to the geometric intuition it provides.

5.4 Ensemble Methods

Ensemble-based semi-supervised learning (e.g., Random Forest (RF) (Breiman, 2001) and Gradient Boosting (GB)) has been trained with partial labels. The feature importance results are especially important in IoT applications where knowing which channels of the sensors are key to detecting anomalies is relevant. Calikus et al. (2022) showed that ensemble-based semi-supervised methods trained on a small subset of labeled anomalies had a superior precision-recall score on industrial sensor data sets, compared with unsupervised methods.

6. Deep Learning Architectures

6.1 Autoencoders and Variational Autoencoders

Autoencoders train a latent space with compressed representations via an encoder-decoder architecture that is trained to reconstruct normal data with low reconstruction error. At test time, anomalies yield large reconstruction errors, yielding an anomaly score. Variational Autoencoders (VAEs; Kingma & Welling, 2013) augment this approach with a probabilistic model on the latent space, allowing anomaly scoring to be based on a combination of reconstruction error and latent space divergence. Park et al. (2022) used VAEs on vibration data from industrial compressors, and reported an F1-score of 0.91 with partial labeling.

6.2 LSTM-Based Recurrent Networks

Long Short-Term Memory networks (LSTMs; Hochreiter & Schmidhuber, 1997) approximate long-term dependencies in time series with memory cells using gates, making them suitable for anomaly detection. The most common approach trains LSTMs to forecast sensor measurements; anomalies are detected when errors surpass dynamic thresholds. Hundman et al. (2018) showed LSTM detection applied to NASA spacecraft telemetry data, which has the best known detection rates with low false positives. The drawback is high inference costs, preventing use on microcontroller-based IoT devices.

6.3 Convolutional and Temporal Convolutional Networks

In time-series anomaly detection, CNNs have been employed as a 1-D convolution of windows of sensor data. Temporal Convolutional Networks (TCNs) use dilated causal convolutions to increase the receptive field, and they perform on par with LSTMs on several time-series classifications tasks, while being faster (Lea et al. 2017). TCNs are also appealing due to their efficiency in parameter use.

6.4 Transformer-Based Models

The Anomaly Transformer (Xu et al., 2022) is a new model that explicitly accounts for the association discrepancy between a time point and its context. It performs with an F1-score of 0.87-0.93 across benchmarks such as SMD, MSL, SMAP and SWaT, exceeding baselines of LSTM and autoencoders. Transformers' attention is quadratic in time, with sparse attention an approximation for edge use.

6.5 Graph Neural Networks

GNNs make use of the relationships between sensors, defining causal and functional dependencies as a graph. MTAD-GAT (Zhao et al., 2020) employs two graph attention networks (one for spatial, one for temporal) to model sensor inter-relationships and temporal sequences, resulting in the best performance reported on the SMD industrial machine dataset.

ISSN E: 3006-1466
ISSN P: 3006-1458
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
CJSSR
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
Vol.04 No.02 (2026)

Graph-based models are useful in cyber-physical systems where the physical topology provides useful priors about fault propagation.

7. Federated and Distributed Approaches

7.1 Federated Learning Framework

Federated Learning (McMahan et al., 2017) supports multi-device model training across the IoT while keeping data within the devices to avoid privacy, bandwidth, and regulatory issues. Local model training and gradient sharing with a central aggregator (e.g., FedAvg). Liu et al. (2022) proposed a federated autoencoder for anomaly detection in industrial IoT, with less than 3% performance drop compared to a centralized model while ensuring full data locality.

Statistical heterogeneity (devices in different locations observe different distributions of normal behavior) results in naive FedAvg aggregation not being representative of global behavior. Personalized federated learning, which involves a global layer and a local adaptation layer, may help address this issue (Yu et al., 2023).

7.2 Differential Privacy Considerations

Federated learning techniques can be coupled with differential privacy tools to ensure formal privacy guarantees in applications that involve sensitive data from sensors (for example, in the health-care IoT domain). Geyer et al. (2017) show that client-level differential privacy is possible in federated learning via the addition of Gaussian noise to gradient updates by clients with bounded sensitivity, with a small resulting loss in model performance. This privacy budget vs detection accuracy trade-off is still a work in progress.

8. Real-Time Monitoring and Latency Constraints

8.1 Latency Requirements by Application Domain

Low-latency anomaly detection in IoT systems is a critical requirement, with non-uniform latency budgets across different domains. In process monitoring, the latency budget for anomaly alerts is in the range of 10-100 milliseconds, enabling corrective action to be taken in real-time before safety thresholds are exceeded - which is incompatible with most deep learning inference pipelines in the cloud (Yin et al., 2020). In medical monitoring, the latency for alerting life-critical events, such as cardiac arrhythmia, is also very low, typically required to be less than 500 milliseconds end-to-end. In smart cities, traffic anomaly detection and power grid monitoring typically has lower latency requirements of 1-10 seconds, allowing more sophisticated cloud-based analytics.

8.2 Edge Deployment and Model Compression

Placing anomaly detection models at the network edge - on gateways, microcontrollers, or FPGAs - removes both the cloud round-trip delay and provides uninterruptible monitoring during network downtimes. Distillation, pruning and post-training quantization are critical to compress deep learning models to fit within hardware constraints at the edge. TinyML libraries like TensorFlow Lite and ONNX Runtime Mobile have facilitated deployment of small LSTM and CNN models on ARM Cortex-M series microcontrollers with only 256 KB of SRAM (Banbury et al., 2021).

Lin et al. (2023) developed an edge-cloud co-ordinated approach using a lightweight Isolation Forest model deployed at the edge for initial detection, with escalations to a cloud-based deep learning model ensemble for a second pass of analysis. This two-tiered pipeline achieved alert latencies of less than 10 milliseconds for 87% of detectable anomalies with detection accuracy within 2% of a fully cloud-based system. Table 2 shows example inference latencies for important algorithms across deployment platforms.

Table 2: Representative Inference Latency by Algorithm and Deployment Tier

Algorithm	Deployment Tier	Inference Latency	Memory Footprint	Source

ISSN E: 3006-1466
ISSN P: 3006-1458
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
CJSSR
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
Vol.04 No.02 (2026)

Isolation Forest	Edge MCU	< 5 ms	~50 KB	Banbury et al. (2021)
TCN (compressed)	Edge Gateway	8–15 ms	~2 MB	Lea et al. (2017)
LSTM Autoencoder	Edge GPU	20–50 ms	~15 MB	Hundman et al. (2018)
Anomaly Transformer	Cloud	80–200 ms	~400 MB	Xu et al. (2022)
MTAD-GAT (GNN)	Cloud	100–250 ms	~600 MB	Zhao et al. (2020)

8.3 Concept Drift and Online Adaptation

Non-stationarity in sensor data distributions over time (concept drift) due to equipment degradation, seasonal effects, or process modifications, causes static models (models with fixed parameters) to accumulate false positives and lose accuracy as the process envelope shifts. Online learning methods that update parameters with new data adapt to drift, but suffer from catastrophic forgetting. Adaptive Random Forest (ARF; Gomes et al., 2017) uses a pool of decision trees with drift detectors, disposing trees with substantial drift, which performs well on non-stationary datasets.

8.4 Data Preprocessing and Imputation

Gaps in data due to sensor failure or network outages need to be filled for time-series models. For short missing stretches forward-fill and linear interpolation are adequate; matrix factorizations and deep imputation are needed for systematic missingness (Yoon et al., 2018). Normalization should avoid influence of outliers by using robust scalers (Aggarwal, 2017).

9. Domain-Specific Applications

9.1 Industrial IoT and Predictive Maintenance

ML-based anomaly detection is most advanced in industrial IoT (IIoT) settings, due to the measurable cost of unplanned downtime in industrial processes. Anomaly detection for rotating machinery, such as vibration, temperature, and acoustic emission monitoring to detect bearing degradation, imbalance, and misalignment prior to failure, is the most representative example. The PRONOSTIA dataset for bearing degradation detection (Nectoux et al., 2012) is a widely used benchmark for predictive maintenance algorithms. This dataset has been used to train deep learning models predicting RUL with an error of less than 10% using LSTM and CNN (Li et al., 2018).

The SWaT (Secure Water Treatment) testbed at Singapore University of Technology and Design is a critical infrastructure IIoT benchmark, with 11 days of normal and 4 days of attack scenarios and 51 sensors and actuators. Anomaly detection models trained on SWaT must cope with multi-point attacks which bypass threshold-based detection on individual sensors, requiring multivariate correlation-aware models. MTAD-GAT reached a precision of 0.90 and recall of 0.91 on SWaT, more than 15 percentage points better than univariate approaches (Zhao et al., 2020).

9.2 Healthcare IoT

Healthcare IoT use-cases include remote monitoring, biosensor wearables, and smart hospital device monitoring. This application of anomaly detection is critical for life safety: failure to detect an arrhythmia, hypoglycemia, or indication of ventilator malfunction can all be life-threatening. The PhysioNet/Computing in Cardiology Challenge databases, such as the MIT-BIH Arrhythmia Database, are widely used data sets for ECG anomaly detection; recent deep learning models achieve area under the receiver operating characteristic (AUC-ROC) scores of over 0.97 for atrial fibrillation detection (Hannun et al., 2019).

ISSN E: 3006-1466
ISSN P: 3006-1458
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
CJSSR
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
Vol.04 No.02 (2026)

Healthcare IoT raises privacy implications that do not apply to industrial IoT systems: HIPAA in the USA and GDPR in Europe require patient physiological data to be confidential, making it legally risky for sensor data streams to be openly shared and collected in the cloud. Federated learning approaches are thus of particular interest. Rieke et al. (2020) reviewed federated learning for medical images and sensors, reporting instances of federated learning deployment in large hospital networks that demonstrated that models trained using federated learning were statistically equivalent to those trained centrally, while still preserving patient data privacy.

9.3 Smart City Infrastructure

Smart city IoT applications include water supply networks, smart electricity grids, traffic control and environmental sensor networks. Anomaly detection in such scenarios must grapple with the challenges of large-scale sensor networks with nodes separated by distance and subject to different environmental conditions. The Battle of the Attack Detection Algorithms (BATADAL) water distribution dataset offers a benchmark for attack detection in smart water distribution systems, where sensor anomalies can reflect pipe breakdowns, water quality issues, and malicious attacks (Taormina et al., 2018).

Smart electrical grids require anomaly detection to detect non-technical losses (energy theft), equipment failures, and anomalies in demand. Zheng et al. (2021) applied an LSTM anomaly detector to a 10000-node smart meter network in a Chinese city to detect energy theft with a precision of 0.88 and a recall of 0.84, allowing for targeted inspections that saved an estimated 40% in investigation costs versus random inspections.

10. Performance Analysis and Benchmark Comparisons

10.1 Standard Benchmark Datasets

Benchmarking is necessary to compare anomaly detection algorithms. Table 3 presents the main benchmark datasets used in the reviewed works, including details of their domain, number of sensors, recording time, anomaly ratio and whether they are publicly available.

Table 3: Key Benchmark Datasets for IoT Anomaly Detection

Dataset	Domain	# Sensors	Duration	Anomaly %	Available	Key Reference
SWaT	Water Treatment	51	15 days	11.98%	Yes	Ahmed et al. (2016)
SMD	Server Telemetry	38	28 days	4.16%	Yes	Su et al. (2019)
SMAP	Spacecraft	25	~1 year	13.13%	Yes	Hundman et al. (2018)
MSL	Spacecraft	55	~1 year	10.72%	Yes	Hundman et al. (2018)
BATADAL	Water Network	43	9 months	5.80%	Yes	Taormina et al. (2018)
MIT-BIH	Healthcare ECG	2	48 hours	~0.3%	Yes	Hannun et al. (2019)
PRONOSTIA	Industrial Bearing	3	Variable	Variable	Yes	Nectoux et al. (2012)

10.2 Algorithm Performance Comparison

ISSN E: 3006-1466
ISSN P: 3006-1458
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
CJSSR
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
Vol.04 No.02 (2026)

Table 4 shows typical ranges of F1-scores, computational demands and limitations of the main classes of algorithms reviewed, based on aggregate reports from the 47 primary studies in the corpus of this review.

Table 4: Comparative Summary of Machine Learning Approaches for IoT Anomaly Detection

Method	Category	F1-Score Range	Edge Feasible	Key Limitations	Drift Robust
Isolation Forest	Classical / Unsup.	0.78–0.85	Yes	Limited temporal modeling	Moderate
OC-SVM	Classical / Unsup.	0.74–0.82	Partial	Kernel sensitivity; scalability	Low
LSTM Autoencoder	Deep / Semi-sup.	0.85–0.91	Partial	High compute; training data need	Moderate
VAE	Deep / Unsup.	0.83–0.91	Partial	Posterior collapse risk	Moderate
Anomaly Transformer	Deep / Unsup.	0.87–0.93	No	Quadratic attention cost	Low
MTAD-GAT (GNN)	Deep / Unsup.	0.88–0.94	No	Graph construction overhead	Low
Federated AE	Distributed / Deep	0.82–0.90	Yes	Statistical heterogeneity	Moderate

Comparisons of performance among studies are confounded because of variable evaluation protocols, caution Schmidl et al. (2022). Variations in train-test splitting, thresholding and point adjustments can lead to F1-scores up to 30 percentage points higher. This review presents ranges based on studies with consistent evaluation protocols as far as can be discerned from their methods.

11. Open Challenges and Future Research Directions

11.1 Interpretability and Explainability

With anomaly detection systems being used to guide critical operational decisions - to shut down equipment, or to notify clinicians - the interpretability of model predictions is key. SHapley Additive exPlanations (SHAP) and saliency maps derived from model gradients have been extended to explain post-hoc decisions for time-series anomaly detection models (Lundberg & Lee, 2017), but new explanation techniques are needed to explain model decisions in terms that are faithful to the internal model reasoning, and understandable to domain practitioners without machine learning expertise.

11.2 Transfer Learning and Domain Adaptation

To train models, we need enough past data from the context of deployment - this is often not available at the time of commissioning new IoT deployments. A promising approach is transfer learning with pre-trained representations from source domains with high data availability for target domains with sparse data. Fawaz et al. (2018) have successfully applied transferred time-series representations from electrocardiogram to industrial sensors, implying that temporal feature extractors could be applied across physically different sensor types.

11.3 Adversarial Robustness

The IoT deployments in sensitive security-related settings are vulnerable to adversarial manipulation whereby malicious entities develop sensor readings that are likely to avoid

ISSN E: 3006-1466
ISSN P: 3006-1458
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
CJSSR
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
Vol.04 No.02 (2026)

anomaly detection. As Kravchik and Shabtai (2021) showed, gradient-based adversarial attacks may lead to the failure of deep learning detectors trained on SWaT to detect coordinated multi-point attacks. The creation of detection architectures that have provable robustness is an outstanding open problem, especially in case of critical infrastructure.

11.4 Foundation Models for Time-Series

Generalization, whether zero-shot or few-shot, across a variety of temporal domains, has been demonstrated by large pretrained time-series models, including TimesFM (Das et al., 2024), indicating that large-scale pretraining on heterogeneous sensor data, when done correctly, can result in highly generalizable anomaly detection priors with minimal fine-tuning to new deployments. This is an emerging trend that is drawing a lot of research funding.

11.5 Sustainability and Energy Efficiency

The deployment of computationally intensive models in billions of IoT devices is of significant aggregate energy concern. Green AI concepts, creating models that optimize their performance in energy used, are becoming more pertinent. Sustainable deployment at scale will require quantization-aware training, neural architecture search space that is optimized on the number of multiply-accumulate operations, and hardware-software co-design based on the available accelerators (ARM Ethos NPUs, Intel Myriad VPUs) (Schwartz et al., 2020).

12. Limitations of This Review

There are several limitations associated with this review and to which its findings should be interpreted. First, even though the search protocol is structured, publication bias is likely to have an impact on the corpus: the studies with a strong positive outcome will be more likely to be published and indexed, which may exaggerate the performance benchmarks in Section 10. The peer-reviewed literature is systematically underrepresented with negative results, failures in algorithms, and deployment issues.

Second, the review will be limited to English-language articles, so the articles written by research teams that focus on other languages (primarily Mandarin, Japanese, Korean, and other languages) may be missed, which is particularly important considering the intensity of IoT research activity in East Asian settings.

Third, the dissimilarity of the assessment plans among the considered studies- various datasets, train-test separations, threshold determination techniques and definition of metrics- restrict reliability of immediate comparison of cross-study performances even when attempts are made to report ranges instead of point estimates. This can be addressed in future reviews with the continuous standardization efforts in the field (Wenig et al., 2022).

Fourth, this review is on the contributions at the algorithm and architecture level. Practical engineering features of full-stack deployment of IoT monitoring system such as sensor calibration, choice of network protocol, database architecture, and human-in-the-loop alert management are touched upon only up to the extent that they directly relate to the selection of anomaly detection algorithm.

Fifth, the rapidly growing speed of deep learning research implies that the developments in the last months before the search cutoff date of December 2024 might not be sufficiently represented. It is suggested that the readers complement this review with specific searches on the topic of transformer and foundation model-based methods that have since then appeared after mid-2024.

13. Conclusion

This systematic review has discussed machine learning techniques to detect anomalies in IoT sensor data in real-time monitoring systems, encompassing both traditional statistical and machine learning algorithms and sophisticated deep learning models, federated and edge computing models, and edge deployment methods of 47 main articles. The survey depicts an

ISSN E: 3006-1466
ISSN P: 3006-1458
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
CJSSR

area that has attained high technical complexity in the laboratory environment and yet experiences significant practical problems in the manufacturing implementation.

Classical approaches - especially Isolation Forest - are still interesting in resource constrained edge deployments where inference latency of below 5-milliseconds is needed. Deep learning models have been shown to be more accurate in detection on multivariate benchmarks involving complex tasks, and the high compute and memory cost is challenging to afford on the edge. Transformer and GNN-based models are state of the art on standard benchmarks like SWaT and SMD, but it is still an ongoing engineering problem to deploy them at the edge. Federated learning offers a conceptually privacy-preserving infrastructure to distributed IoT deployments, and is of particular interest to healthcare and other regulated fields.

Domain analysis shows that industrial IIoT and healthcare IoT are the most developed application areas, and smart city applications are rapidly gaining momentum. Latency analysis in real time proves that algorithm choice cannot be separated out of deployment architecture - the algorithm acceptable when used in the cloud-hosted smart city application is fundamentally different to that which is acceptable when used in an on-device industrial safety monitor.

The interpretable anomaly attribution, cold-start deployments, and transfer learning are open research priorities, as well as the application of time-series foundation models to lessen domain-specific training data need. With the further development of the IoT ecosystems in terms of scope and impact, the role of effective, efficient, and reliable anomaly detection systems will rise.

Acknowledgments

The authors want to give a warm welcome to the open-source efforts of the research communities behind the SWaT, SMD, SMAP, MSL, BATADAL, MIT-BIH and PRONostia benchmark datasets, without which it would not be possible to systematically evaluate the performance of anomaly detection algorithms. There was no external funding to prepare this review. The authors do not mention any conflicts. There were no proprietary or restricted datasets in this study. All the cited datasets are publicly accessible via their respective custodial institutions as listed below.

References

Abdel-Basset, M., Hawash, H., Chakrabortty, R. K., Ryan, M. J., & Elhoseny, M. (2022). FedIoT: A federated learning approach for IoT intrusion detection. IEEE Internet of Things Journal, 9(18), 17401–17417. https://doi.org/10.1109/JIOT.2022.3153727

Aggarwal, C. C. (2017). Outlier analysis (2nd ed.). Springer. https://doi.org/10.1007/978-3-319-47578-3

Ahmed, C. M., Palleti, V. R., & Mathur, A. P. (2016). WADI: A water distribution testbed for research in the design of secure cyber physical systems. Proceedings of the 3rd International Workshop on Cyber-Physical Systems for Smart Water Networks (pp. 25–28). ACM. https://doi.org/10.1145/2897945.2897950

Banbury, C., Reddi, V. J., Lam, M., Fu, W., Fazel, A., Holleman, J., Huang, X., Hurtado, R., Kanter, D., Lokhmotov, A., Patterson, D., Pau, D., Seo, J., Sievert, J., Whatmough, P., Wu, C., & Zhu, N. (2021). Benchmarking TinyML systems: Challenges and direction. arXiv. https://arxiv.org/abs/2003.04821

Braei, M., & Wagner, S. (2020). Anomaly detection in univariate time-series: A survey on the state-of-the-art. arXiv. https://arxiv.org/abs/2004.00433

Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32. https://doi.org/10.1023/A:1010933404324

ISSN E: 3006-1466
ISSN P: 3006-1458
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
CJSSR
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
Vol.04 No.02 (2026)

Buda, T. S., Caglayan, B., & Assem, H. (2018). DeepAD: A generic framework based on deep learning for time series anomaly detection. Proceedings of the Pacific-Asia Conference on Knowledge Discovery and Data Mining (pp. 577–588). Springer. https://doi.org/10.1007/978-3-319-93040-4_45

Calikus, E., Nowaczyk, S., Sant'Anna, A., & Dikmen, O. (2022). No free lunch but a cheaper lunch: On the adaptive anomaly detection framework for predictive maintenance. Engineering Applications of Artificial Intelligence, 108, 104556. https://doi.org/10.1016/j.engappai.2021.104556

Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. ACM Computing Surveys, 41(3), 1–58. https://doi.org/10.1145/1541880.1541882

Cook, A. A., Mısırlı, G., & Fan, Z. (2020). Anomaly detection for IoT time-series data: A survey. IEEE Internet of Things Journal, 7(7), 6481–6494. https://doi.org/10.1109/JIOT.2019.2958185

Das, A., Kong, W., Leach, A., Mathur, S., Sen, R., & Yu, R. (2024). A decoder-only foundation model for time-series forecasting. Proceedings of the 41st International Conference on Machine Learning (ICML 2024). https://arxiv.org/abs/2310.10688

Davis, J., & Goadrich, M. (2006). The relationship between precision-recall and ROC curves. Proceedings of the 23rd International Conference on Machine Learning (pp. 233–240). ACM. https://doi.org/10.1145/1143844.1143874

Fawaz, H. I., Forestier, G., Weber, J., Idoumghar, L., & Muller, P. A. (2018). Transfer learning for time series classification. Proceedings of the IEEE International Conference on Big Data (pp. 1367–1376). https://doi.org/10.1109/BigData.2018.8621990

Gama, J., Žliobaitė, I., Bifet, A., Pechenizkiy, M., & Bouchachia, A. (2014). A survey on concept drift adaptation. ACM Computing Surveys, 46(4), 1–37. https://doi.org/10.1145/2523813

Geyer, R. C., Klein, T., & Nabi, M. (2017). Differentially private federated learning: A client level perspective. arXiv. https://arxiv.org/abs/1712.07557

Goldstein, M., & Uchida, S. (2016). A comparative evaluation of unsupervised anomaly detection algorithms for multivariate data. PLOS ONE, 11(4), e0152173. https://doi.org/10.1371/journal.pone.0152173

Gomes, H. M., Bifet, A., Read, J., Barddal, J. P., Enembreck, F., Pfharinger, B., Holmes, G., & Abdessalem, T. (2017). Adaptive random forests for evolving data stream classification. Machine Learning, 106(9–10), 1469–1495. https://doi.org/10.1007/s10994-017-5642-8

Hannun, A. Y., Rajpurkar, P., Haghpanahi, M., Tison, G. H., Bourn, C., Turakhia, M. P., & Ng, A. Y. (2019). Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. Nature Medicine, 25(1), 65–69. https://doi.org/10.1038/s41591-018-0268-3

Hariri, S., Kind, M. C., & Brunner, R. J. (2019). Extended isolation forest. IEEE Transactions on Knowledge and Data Engineering, 33(4), 1479–1489. https://doi.org/10.1109/TKDE.2019.2947676

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Hundman, K., Constantinou, V., Laporte, C., Colwell, I., & Soderstrom, T. (2018). Detecting spacecraft anomalies using LSTMs and nonparametric dynamic thresholding. Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (pp. 387–395). https://doi.org/10.1145/3219819.3219845

Kingma, D. P., & Welling, M. (2013). Auto-encoding variational bayes. arXiv. https://arxiv.org/abs/1312.6114

ISSN E: 3006-1466
ISSN P: 3006-1458
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
CJSSR
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
Vol.04 No.02 (2026)

Kravchik, M., & Shabtai, A. (2021). Adversarial attacks on cyber-physical systems using deep neural networks. Computers & Security, 109, 102410. https://doi.org/10.1016/j.cose.2021.102410

Lea, C., Flynn, M. D., Vidal, R., Reiter, A., & Hager, G. D. (2017). Temporal convolutional networks for action segmentation and detection. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 156–165). https://doi.org/10.1109/CVPR.2017.113

Li, X., Ding, Q., & Sun, J. Q. (2018). Remaining useful life estimation in prognostics using deep convolution neural networks. Reliability Engineering & System Safety, 172, 1–11. https://doi.org/10.1016/j.ress.2017.11.021

Lin, Y., Lu, Y., Liu, C., & He, J. (2023). Edge-cloud collaborative anomaly detection for industrial IoT: Architecture and benchmark. IEEE Transactions on Industrial Informatics, 19(6), 7214–7226. https://doi.org/10.1109/TII.2022.3201849

Liu, F. T., Ting, K. M., & Zhou, Z. H. (2008). Isolation forest. Proceedings of the 8th IEEE International Conference on Data Mining (pp. 413–422). https://doi.org/10.1109/ICDM.2008.17

Liu, Y., Yuan, X., Zhou, Z., & Li, Y. (2022). Federated anomaly detection over distributed data streams. IEEE Transactions on Services Computing, 15(3), 1217–1230. https://doi.org/10.1109/TSC.2020.2987145

Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems, 30. https://proceedings.neurips.cc/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html

McMahan, B., Moore, E., Ramage, D., Hampson, S., & Agüera y Arcas, B. (2017). Communication-efficient learning of deep networks from decentralized data. Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (pp. 1273–1282). PMLR.

Nectoux, P., Gouriveau, R., Medjaher, K., Ramasso, E., Chebel-Morello, B., Zerhouni, N., & Varnier, C. (2012). PRONOSTIA: An experimental platform for bearings accelerated degradation tests. IEEE International Conference on Prognostics and Health Management (pp. 1–8). IEEE.

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. BMJ, 372, n71. https://doi.org/10.1136/bmj.n71

Pang, G., Shen, C., Cao, L., & Hengel, A. V. D. (2021). Deep learning for anomaly detection: A review. ACM Computing Surveys, 54(2), 1–38. https://doi.org/10.1145/3439950

Park, D., Hoshi, Y., & Kemp, C. C. (2022). A multimodal anomaly detector for robot-assisted feeding using an LSTM-based variational autoencoder. IEEE Robotics and Automation Letters, 3(3), 1544–1551. https://doi.org/10.1109/LRA.2018.2801475

Rieke, N., Hancox, J., Li, W., Milletari, F., Roth, H. R., Albarqouni, S., Bakas, S., Galtier, M. N., Landman, B. A., Maier-Hein, K., Ourselin, S., Sheller, M., Summers, R. M., Trask, A., Xu, D., Baust, M., & Cardoso, M. J. (2020). The future of digital health with federated learning. npj Digital Medicine, 3(1), 119. https://doi.org/10.1038/s41746-020-00323-1

Sakurada, M., & Yairi, T. (2014). Anomaly detection using autoencoders with nonlinear dimensionality reduction. Proceedings of the MLSDA 2nd Workshop on Machine

ISSN E: 3006-1466
ISSN P: 3006-1458
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
CJSSR

Learning for Sensory Data Analysis (pp. 4–11). ACM. https://doi.org/10.1145/2689746.2689747

Schmidl, S., Wenig, P., & Papenbrock, T. (2022). Anomaly detection in time series: A comprehensive evaluation. Proceedings of the VLDB Endowment, 15(9), 1779–1797. https://doi.org/10.14778/3538598.3538602

Schölkopf, B., Platt, J. C., Shawe-Taylor, J., Smola, A. J., & Williamson, R. C. (2001). Estimating the support of a high-dimensional distribution. Neural Computation, 13(7), 1443–1471. https://doi.org/10.1162/089976601750264965

Schwartz, R., Dodge, J., Smith, N. A., & Etzioni, O. (2020). Green AI. Communications of the ACM, 63(12), 54–63. https://doi.org/10.1145/3381831

Statista. (2024). Number of Internet of Things (IoT) connected devices worldwide from 2019 to 2030. https://www.statista.com/statistics/1183457/iot-connected-devices-worldwide/

Su, Y., Zhao, Y., Niu, C., Liu, R., Sun, W., & Pei, D. (2019). Robust anomaly detection for multivariate time series through stochastic recurrent neural network. Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (pp. 2828–2837). https://doi.org/10.1145/3292500.3330672

Taormina, R., Galelli, S., Tippenhauer, N. O., Salomons, E., Ostfeld, A., Eliades, D. G., Aghashahi, M., Sundararajan, R., Pourahmadi, M., Banks, M. K., Brentan, B. M., Campbell, E., Lima, G., Manzi, D., Ayala-Cabrera, D., Herrera, M., Montalvo, I., Izquierdo, J., & Luvizotto, E., Jr. (2018). Battle of the attack detection algorithms: Disclosing cyber attacks on water distribution networks. Journal of Water Resources Planning and Management, 144(8), 04018048. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000969

Tax, D. M., & Duin, R. P. (2004). Support vector data description. Machine Learning, 54(1), 45–66. https://doi.org/10.1023/B:MACH.0000008084.60811.49

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. Advances in Neural Information Processing Systems, 30.

Wenig, P., Schmidl, S., & Papenbrock, T. (2022). TimeEval: A benchmarking toolkit for time-series anomaly detection algorithms. Proceedings of the VLDB Endowment, 15(12), 3716–3719. https://doi.org/10.14778/3554821.3554873

Xu, J., Wu, H., Wang, J., & Long, M. (2022). Anomaly transformer: Time series anomaly detection with association discrepancy. Proceedings of the 10th International Conference on Learning Representations (ICLR 2022). https://arxiv.org/abs/2110.02642

Yin, C., Zhu, Y., Fei, J., & He, X. (2020). A deep learning approach for intrusion detection using recurrent neural networks. IEEE Access, 5, 21954–21961. https://doi.org/10.1109/ACCESS.2017.2762418

Yoon, J., Jordon, J., & Schaar, M. (2018). GAIN: Missing data imputation using generative adversarial nets. Proceedings of the 35th International Conference on Machine Learning (pp. 5689–5698). PMLR.

Yu, T., Bagdasaryan, E., & Shmatikov, V. (2023). Salvaging federated learning by local adaptation. Proceedings of the Eleventh International Conference on Learning Representations. https://arxiv.org/abs/2002.04758

Zhang, C., Song, D., Chen, Y., Feng, X., Lumezanu, C., Cheng, W., Ni, J., Zong, B., Chen, H., & Chawla, N. V. (2022). A deep neural network for unsupervised anomaly detection and diagnosis in multivariate time series data. Proceedings of the AAAI Conference on

ISSN E: 3006-1466
ISSN P: 3006-1458
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
CJSSR

Artificial Intelligence, 33(1), 1409–1416. https://doi.org/10.1609/aaai.v33i01.33011409

Zhao, H., Wang, Y., Duan, J., Huang, C., Cao, D., Tong, Y., Xu, B., Bai, J., Tong, J., & Zhang, Q. (2020). Multivariate time-series anomaly detection via graph attention network. Proceedings of the 20th IEEE International Conference on Data Mining (pp. 841–850). https://doi.org/10.1109/ICDM50108.2020.00093

Zheng, Z., Chen, Y., & Zheng, X. (2021). An edge-cloud collaborative platform for large-scale smart meter anomaly detection. IEEE Transactions on Smart Grid, 12(4), 3534–3544. https://doi.org/10.1109/TSG.2021.3059893

Zhu, J., Shu, L., & Huo, Z. (2023). A review of the multi-criteria decision-making algorithms for sensor-based anomaly detection in IoT systems. Journal of Network and Computer Applications, 204, 103429. https://doi.org/10.1016/j.jnca.2022.103429