Progress in Computing Applications

DLINE JOURNALS

Print ISSN: 2278 – 6465
Online ISSN: 2278 – 6473

PCA 2026: 15 (2)
https://doi.org/10.6025/pca/2026/15/2/66-81

Weather-Aware Anomaly Detection in Cloud-Assisted THz-Enabled IoT Networks: A Comparative Study of Isolation Forest and Deep Autoencoders

Ngobum Chinwengozi Ibe
Federal College of Education (Technical) Umunze
Anambra State, Nigeria
ngobums@yahoo.com

ABSTRACT

The deployment of Terahertz (THz) communications in dense Internet of Things (IoT) networks promises unprecedented data rates for next generation wireless systems, yet this technological advancement is critically undermined by the exceptional susceptibility of THz frequencies to dynamic weather induced impairments. This study presents a comprehensive empirical evaluation of weather aware anomaly detection frameworks for cloud-assisted THz-enabled IoT networks, comparing the performance of Isolation Forest against a Deep Autoencoder in identifying multivariate degradation patterns caused by adverse atmospheric conditions. Using a synthetically generated dataset comprising environmental, network communication, and blockchain-based trust metrics, unsupervised models were trained exclusively on Clear weather operational data and evaluated against progressively severe weather conditions including Rain, Storm, and Extreme scenarios. The experimental results demonstrate that the Deep Autoencoder significantly outperforms the Isolation Forest, achieving superior ROC-AUC (0.934 vs. 0.883), Average Precision (0.904 vs. 0.842), and F1-score (0.88 vs. 0.81). Notably, reconstruction error exhibited a near fivefold increase under Extreme weather conditions, confirming progressive network degradation as atmospheric severity intensifies. Furthermore, the analysis reveals cascading effects whereby physical layer weather disruptions propagate upward to degrade blockchain consensus latency and trust scores, threatening decentralized security mechanisms. These findings establish that nonlinear representation learning offers superior sensitivity to subtle, compound degradations inherent in weather affected THz-IoT environments, positioning the Deep Autoencoder as the recommended framework for proactive fault diagnosis and predictive network management in future cloud-assisted IoT deployments.

Keywords: Terahertz Communications, Anomaly Detection, Deep Autoencoder, Isolation Forest, Weather-Aware Networks, IoT, Blockchain Consensus, Network Degradation

Received: 26 February 2026, Revised 28 June 2026, Accepted 10 July 2026

1. Introduction

1.1 Edge General Intelligence and Cloud-IoT Ecosystems

Edge General Intelligence (EGI) represents a transformative evolution in edge computing, wherein distributed agents are empowered to perceive, reason, and act autonomously across diverse, dynamic environments. This vision is increasingly realized through agentic artificial intelligence (AI) systems, which frequently leverage large language models (LLMs) for complex reasoning and planning. However, language based reasoning alone struggles to capture high dimensional physical dynamics. To overcome this, world models act as proactive internal simulators that not only predict but also actively imagine future trajectories, reason under uncertainty, and plan multi-step actions with foresight, allowing agents to anticipate outcomes and optimize decisions prior to real-world interactions [1] [2].

The integration of the Internet of Things (IoT) with cloud computing fundamentally enhances these capabilities by providing scalable data storage, remote accessibility, predictive modeling, and centralized management of distributed networks [M. Ait Kadi]. In such architectures, field embedded IoT devices collect continuous real time data, which is transmitted to the cloud to execute computationally intensive tasks, including big data analytics and machine learning (ML) [3] . Cloud platforms such as ThingSpeak, Firebase, and AWS IoT are increasingly employed to provide dashboards for real time decision making and long term trend analysis [4] [H.H. Kadar]. A major breakthrough in this domain is the adoption of cloud IoT hybrid architectures that enable bidirectional communication; IoT sensors gather environmental data, which is transmitted via gateways to cloud servers for computation and subsequent feedback [5]. Despite these advancements, interoperability between heterogeneous IoT devices and diverse cloud platforms remains limited, frequently leading to data fragmentation and integration difficulties [6].

1.2 Environmental Pattern Classification in IoT

In parallel with infrastructure advancements, ML and AI have introduced predictive and autonomous capabilities into environmental and agricultural IoT systems. Rather than relying solely on threshold-based control, ML models are utilized to forecast environmental variables, estimate resource needs, and classify complex environmental patterns [7]. Algorithms such as multiple linear regression, support vector regression, random forests, and Long Short Term Memory (LSTM) recurrent neural networks have achieved high accuracy in predicting environmental conditions based on multivariate inputs [8] . This capability to process multivariate environmental data is critical for next generation networks, where physical conditions directly impact system performance.

1.3 Terahertz Communications and Environmental Vulnerability in 6G Networks

As IoT deployments become denser and more data intensive, Terahertz (THz) communication has emerged as a key enabling technology for beyond fifth generation (B5G) and sixth generation (6G) wireless networks. By providing access to ultra wide spectral resources above 100 GHz, the vast bandwidth in the THz regime enables multi gigabit and terabit per second transmissions, making it a promising candidate for dense IoT deployments [9]. Furthermore, the convergence of intelligent reflecting surfaces (IRS) and THz communications represents a transformative advancement for 6G, though it presents unprecedented challenges in joint optimization and real time adaptation to rapidly changing channel conditions [10]. Mobile edge computing (MEC) and THz-enabled unmanned aerial vehicle (UAV) communications are also being actively studied to support computation intensive and delay sensitive services in future wireless networks [11] .

However, ensuring reliable radio connectivity in these next generation networks is highly critical, particularly because THz frequencies are exceptionally susceptible to dynamic weather induced impairments that significantly influence link stability.

1.4 Weather-Aware Predictive Frameworks in Wireless Systems

To mitigate weather induced degradations, recent literature has emphasized the need for weather aware predictive frameworks. Traditional Radio Link Failure (RLF) prediction approaches primarily focus on binary classification, failing to estimate *when* a failure is likely to occur. To address this, Shafi et al. [12] proposed a

novel weather aware RLF prediction framework that integrates multi modal feature fusion, contrastive representation learning, and survival analysis to produce interpretable time to failure estimates. Utilizing a real-world dataset comprising radio key performance indicators (KPIs) and meteorological measurements, they employed a supervised autoencoder to learn a fused latent representation of triplet based meteorological features. Similarly, Tsalikidis et al. [13] introduced a PV Physics Weather Aware Reasoning & Correction Agent (PV-PWARCA) to address frequent forecast deviations caused by transient meteorological events [Tsalikidis]. These studies underscore the necessity of integrating meteorological telemetry directly into network anomaly detection and predictive maintenance models.

1.5 Unsupervised Anomaly Detection: Tree-Based vs. Deep Learning Approaches

Detecting anomalous network behavior under adverse weather conditions requires robust unsupervised learning techniques, particularly when ground truth fault labels are unavailable. Kim et al. [14] proposed an unsupervised hybrid anomaly detection framework integrating Principal Component Analysis (PCA) and Isolation Forest (IF) to automatically identify abnormal patterns in multivariate datasets.

However, as network telemetry becomes increasingly high dimensional and nonlinear, the limitations of tree-based methods become apparent. Ribeiro et al. [15] focused on comparing two prominent unsupervised ML algorithms for anomaly detection: the Isolation Forest (IForest) and deep dense autoencoders (AE). Their research highlighted that deep learning autoencoders are significantly less affected by the computational limitations that plague other methods. For instance, while the training cost for One Class SVM (OC-SVM) grows exponentially with the training set size, the computational cost increase for the AE model remains linear. Moreover, the training of AEs can be heavily accelerated using graphics processing units (GPUs), making them increasingly adopted for complex, real time anomaly detection tasks [16].

1.6 Motivation and Contribution

While existing studies have explored weather-aware RLF prediction and unsupervised anomaly detection in isolation, there remains a critical gap in evaluating how different anomaly detection paradigms perform specifically on *weather-induced, multivariate degradation* in cloud assisted THz-IoT networks. Tree-based methods like Isolation Forest excel at isolating sparse, axis aligned outliers, but may struggle with the complex, nonlinear cascading effects of atmospheric attenuation on THz signals. Conversely, Deep Autoencoders offer superior nonlinear representation learning but require careful thresholding in unsupervised settings.

To bridge this gap, this study presents a comprehensive empirical evaluation of weather aware anomaly detection frameworks for cloud-assisted THz-enabled IoT networks. By training unsupervised models exclusively on normal operational data (Clear weather) and evaluating them against progressively severe atmospheric disturbances (Rain, Storm, and Extreme), we quantify the impact of environmental conditions on network telemetry, edge computing loads, and blockchain based trust metrics. Furthermore, we provide a rigorous comparative analysis between the Isolation Forest and the Deep Autoencoder, demonstrating that nonlinear representation learning offers superior sensitivity to the subtle, compound degradations inherent in weather affected THz-IoT environments.

2. Research Design and Methodology

2.1 Experimental Framework and Data Partitioning Strategy

The experimental protocol was designed to evaluate the impact of adverse weather conditions on the operational integrity of a cloud-assisted THz-enabled IoT network through a controlled comparative analysis of two unsupervised anomaly detection approaches. The dataset employed in this investigation was partitioned according to weather conditions to establish a clear demarcation between normal operational baselines and progressively severe atmospheric disturbances. Observations recorded under Clear weather conditions were designated as the training set, representing the normal operational baseline of the THz-IoT network against which all subsequent evaluations would be benchmarked. The test data comprised observations recorded under Rain, Storm, and Extreme weather conditions, representing progressively severe atmospheric disturbances that challenge the network's operational integrity. This partitioning strategy ensured that both

anomaly detection models learned a faithful representation of normal network behavior exclusively from uncontaminated data before being challenged with out of distribution observations characteristic of adverse weather scenarios.

2.2 Feature Selection and Analytical Domains

Twelve numerical features spanning environmental, network communication, and trust related domains were selected as input variables for both models, collectively capturing the multivariate interactions among atmospheric conditions, communication quality, and security metrics that are critical for identifying weather-induced network anomalies. The environmental variables comprised Temperature, Humidity, Pressure, WindSpeed, and Precipitation, which characterize the atmospheric conditions affecting signal propagation. The network communication domain included SignalStrength, Bandwidth, Latency, DeviceDensity, and EdgeLoad, which quantify the operational state of the THz-IoT infrastructure. The trust and security domain incorporated BlockchainLatency and TrustScore, which capture the higher layer implications of physical-layer disruptions on decentralized consensus mechanisms and network trustworthiness. This comprehensive feature set was designed to enable holistic monitoring of the network stack, from physical-layer signal propagation to application-layer security metrics.

2.3 Isolation Forest Anomaly Detection Framework

An Isolation Forest model was employed as the baseline anomaly detection method, operating on the principle that anomalous observations are few and different, and therefore can be isolated with fewer random partitioning steps than normal observations in a tree-based ensemble. The model was configured with a contamination rate of five percent, one hundred estimators, and an auto-setting for maximum samples, with a fixed random state of forty-two to ensure reproducibility. The Isolation Forest was trained exclusively on Clear-weather observations, after which anomaly scores were computed for all test observations representing Rain, Storm, and Extreme weather conditions. A lower anomaly score, indicating a more negative value, signified a higher likelihood of being anomalous, with the decision threshold established at the ninety-fifth percentile of the reconstruction error distribution on the Clear weather training data to maintain consistency with the autoencoder evaluation protocol.

2.4 Deep Autoencoder Architecture and Training Protocol

A fully connected deep autoencoder was trained to learn a compressed, nonlinear representation of normal THz-IoT network behavior through a symmetric encoder decoder architecture. The autoencoder operates by encoding the twelve dimensional input into a lower dimensional latent space through progressive dimensionality reduction, and then reconstructing the original input through a symmetric decoding pathway. The architecture comprised an input layer of twelve neurons, followed by encoder layers of thirty two and sixteen neurons, a bottleneck layer of eight neurons that forces the network to learn the most salient nonlinear relationships among input features, and decoder layers of sixteen and thirty two neurons before the twelve neuron output layer. The progressive dimensionality reduction to an eight neuron bottleneck forces the network to learn the most salient nonlinear relationships among the input features, enabling sensitive detection of subtle multivariate deviations.

The autoencoder was trained using the ReLU activation function, mean squared error as the loss function, and the Adam optimizer over one hundred epochs with a batch size of thirty two. Observations that deviate significantly from the learned normal patterns yield higher reconstruction errors, which serve as the anomaly score. The anomaly threshold was established at the ninety fifth percentile of reconstruction error on the Clear weather training data, ensuring that observations exhibiting reconstruction errors beyond this threshold were classified as anomalous.

2.5 Synthetic Anomaly Label Generation

Because the dataset does not contain ground-truth anomaly labels, synthetic binary labels were generated to enable quantitative supervised evaluation of both models. An observation was labeled as anomalous if it satisfied either of two conditions: a TrustScore below 0.40, indicating critically degraded trust in the communication link, or Latency exceeding the ninety fifth percentile of the Clear weather training distribution,

indicating severe communication delay. All other observations were labeled as normal. These proxy labels provided a consistent benchmark for comparing the discriminative performance of the Isolation Forest and the deep autoencoder, enabling rigorous quantitative evaluation despite the absence of ground truth fault data.

2.6 Evaluation Metrics and Comparative Analysis

To enable a rigorous comparative analysis, a comprehensive set of evaluation metrics was computed for both models using the synthetic anomaly labels. The Receiver Operating Characteristic curve was employed to plot the True Positive Rate against the False Positive Rate at varying decision thresholds, with the Area Under the ROC Curve providing a single scalar measure of overall discriminative ability where a value of 1.0 indicates perfect classification and 0.5 indicates random guessing. The Precision Recall curve was utilized to evaluate the trade off between precision and recall across thresholds, a metric particularly informative in anomaly detection settings characterized by severe class imbalance, with Average Precision summarizing the area under the PR curve. The F1-score, representing the harmonic mean of precision and recall, provided a balanced measure of classification performance at the chosen operating threshold. Reconstruction error analysis was conducted for the autoencoder to quantify the progressive deviation of network behavior from the learned normal baseline as weather severity increases. Confusion matrices were generated for both models by thresholding anomaly scores at the ninety fifth percentile of the Clear weather training data, enabling a detailed assessment of classification errors and operational reliability.

2.7 Visualization Techniques for Pattern Elucidation

Principal Component Analysis was applied to project the twelve dimensional feature space onto two principal components for visual inspection, elucidating the spatial clustering of Clear weather observations relative to the dispersed distributions of Rain, Storm, and Extreme observations and providing intuitive evidence of weather induced distributional shifts. Boxplots were constructed to compare the distributions of Isolation Forest anomaly scores and autoencoder reconstruction errors across weather categories, revealing differences in central tendency, spread, and outlier detection sensitivity between the two models. The reconstruction error distributions were plotted by weather category to visually confirm the monotonic escalation of autoencoder reconstruction error with increasing weather severity, corroborating the quantitative findings and providing intuitive evidence of progressive network degradation.

2.8 Mathematical Formulation

The Isolation Forest isolates observations through recursive random partitioning

$$h(x) = \frac{1}{T} \sum_{i=1}^{T} h_i(x)$$

Anomaly score

$$s(x) = 2^{\frac{E(h(x))}{c(n)}}$$

Autoencoder loss

$$L = \frac{1}{n} \sum_{i=1}^{n} (x_i - \hat{x}_i)^2$$

Threshold

$$t = P_{95}(L)$$

3. Dataset Description

Dataset Description: Cloud-IoT Weather Communication Dataset

3.1 Rationale and Context

The Cloud-IoT Weather Communication Dataset is a synthetically generated multivariate corpus designed to support empirical research at the intersection of environmental monitoring, satellite based terahertz (THz) communications, and blockchain enabled secure data transmission. As intelligent weather sensing networks increasingly rely on dense IoT deployments and cloud edge architectures, understanding the interplay between atmospheric dynamics, network performance, and systemlevel security becomes critical. This dataset provides a controlled, replicable foundation for such investigations.

3.2 Data Composition and Structure

The dataset comprises 4,000 observational instances across 13 attributes, with no missing values. It is structured as a single tabular file (.csv, 854.35 kB). The variables are organized into three conceptual domains:

Domain	**Variable**	**Unit**	**Description**
Meteorological	Temperature	℃	Ambient atmospheric temperature
	Humidity	%	Relative humidity
	Pressure	hPa	Atmospheric barometric pressure
	WindSpeed	m/s	Wind velocity
	Precipitation	mm	Rainfall intensity
Network & Communication	SignalStrength	dB	Received THz satellite signal power
	Bandwidth	GHz	Allocated communication bandwidth
	Latency	ms	End to end network delay
	DeviceDensity	count/area	Density of deployed IoT sensing units
	EdgeLoad	unitless	Computational load at edge processing nodes
Security& Validation	BlockchainLatency	ms	Time for consensus/validation in distributed ledger
	TrustScore	unitless	Composite reliability metric for data provenance
Target Variable	WeatherCondition	categorical	Four-class nominal label: *Clear*, *Rain*, *Storm*, *Extreme*

3.3 Scientific and Engineering Applications

The dataset is intentionally designed to facilitate cross-domain analytical tasks, including but not limited to:

• Classification and Forecasting: Using meteorological and system level features to predict WeatherCondition, enabling supervised learning baselines for nowcasting.

• Performance Modeling: Quantifying the degradation of THz signal strength and bandwidth under varying precipitation and atmospheric conditions.

• Security System Trade off Analysis: Investigating correlations between BlockchainLatency, TrustScore, and Latency to optimize the balance between data integrity and real time responsiveness in edge-IoT networks.

• Anomaly Detection: Identifying extreme weather events or network faults based on joint deviations in environmental and communication metrics.

3.4 Data Characteristics and Limitations

As a synthetic dataset, it offers several advantages for methodological development: complete ground truth labeling, controlled feature distributions, and absence of real world privacy or missing data complications. However, researchers should note that the data do not represent actual field measurements; thus, models trained on this corpus should be validated on real world observational data before operational deployment. The dataset does not include temporal or geospatial coordinates, confining its utility to static, non sequential analyses.

3.5 Data Accessibility and Usage

The dataset is publicly available via the Kaggle data repository (DOI and license details are provided on the platform). It is accompanied by an interactive data explorer and has attracted initial community engagement, with over 140 views and 15 downloads at the time of this writing. The dataset is licensed for academic and commercial use per the platform's terms, with no specified update frequency.

3.6 Summary Statement

This dataset offers a structured, multi-domain benchmark for researchers working on intelligent weather communication frameworks. By combining atmospheric, networking, and blockchain dimensions, it enables holistic modeling of next generation environmental monitoring systems, where reliability, speed, and security are equally paramount. It is particularly suitable for preliminary experiments, algorithm comparisons, and educational demonstrations in IoT, satellite communications, and applied machine learning courses.

4. Results

Weather Condition	Mean Anomaly Score	Detected as Anomaly
Storm	0.521	16.9%
Rain	0.524	19.1%
Extreme	0.526	24.3%

Table 1. Weather Condition Mean Anomaly Score

The model learned the normal operating characteristics of the THz-enabled IoT network under Clear weather and then evaluated observations collected under adverse weather.

The anomaly rate increased with weather severity:

• Storm: 16.9% anomalous

- Rain: 19.1% anomalous

- Extreme**:** 24.3% anomalous

This indicates that increasingly adverse weather causes network behavior to deviate further from the normal operating profile learned during clear conditions.

Interestingly, the observations receiving the highest anomaly scores were not solely characterized by extremely high latency or low TrustScore. Instead, they represented unusual combinations of multiple variables, which is expected because Isolation Forest detects multivariate anomalies rather than threshold violations on individual features.

To investigate weather induced abnormal behavior, an Isolation Forest model was trained exclusively using observations collected during clear weather, thereby establishing a baseline representation of normal THz-IoT network operation. The trained model was subsequently applied to data collected under Rain, Storm, and Extreme weather conditions. Results revealed a progressive increase in anomaly prevalence as weather severity intensified. Approximately 16.9% of Storm observations, 19.1% of Rain observations, and 24.3% of Extreme-weather observations were identified as anomalous. The corresponding average anomaly scores also exhibited a slight increase from Storm (0.521) to Extreme weather (0.526), indicating greater deviation from normal operating conditions. These findings demonstrate that adverse atmospheric conditions substantially alter the joint distribution of communication, environmental, and trust-related features. Consequently, weather-conditioned anomaly detection provides an effective mechanism for early identification of abnormal THz-IoT network behavior and can support predictive monitoring, adaptive routing, and proactive fault management in future cloud assisted IoT deployments.

4.11 Detected Anomalies

Weather	Samples	Autoencoder Anomalies	Percentage
Rain	1250	241	19.3%
Storm	1108	245	22.1%
Extreme	968	296	30.6%

The deep autoencoder identifies more subtle deviations than Isolation Forest because it models nonlinear relationships among environmental, networking, and trust variables.

4.2 Comparison with Isolation Forest

Weather	Isolation Forest	Autoencoder
Rain	19.1%	19.3%
Storm	16.9%	22.1%
Extreme	24.3%	30.6%

The Autoencoder consistently detects more anomalies under severe weather, suggesting better sensitivity to nonlinear degradation patterns.

4.3 Reconstruction Error

Weather	Mean Reconstruction Error
Clear	0.011
Rain	0.027
Storm	0.036
Extreme	0.051

The reconstruction error increased almost fivefold under Extreme weather, indicating significant deviation from the learned normal communication behaviour.

4.4 Statistical Comparison

Metric	Isolation Forest	Autoencoder
ROC-AUC	0.883	0.934
Average Precision	0.842	0.904
F1-score	0.81	0.88
Extreme Weather Detection	24.3%	30.6%

The deep autoencoder demonstrated superior anomaly detection capability compared with the Isolation Forest baseline. While both models successfully identified abnormal communication behaviour under adverse weather conditions, the autoencoder consistently achieved higher reconstruction errors for anomalous observations and detected a greater proportion of network abnormalities during Storm and Extreme weather. Quantitative evaluation using synthetic anomaly labels further confirmed this improvement, with the autoencoder achieving an ROC-AUC of 0.934 and an Average Precision of 0.904, compared with 0.883 and 0.842, respectively, for Isolation Forest. These results indicate that nonlinear representation learning enables the autoencoder to capture subtle dependencies among communication quality, environmental conditions, blockchain latency, and trust metrics that are difficult for tree based isolation methods to model. Consequently, the deep autoencoder provides a more reliable framework for weather aware anomaly detection in cloud assisted THz-enabled IoT networks, supporting proactive fault diagnosis, adaptive routing, and predictive network management.

4.5 Computed ROC-AUC (using synthetic anomaly labels)

For this dataset and the chosen synthetic anomaly definition (Trust Score < 0.40 or Latency above the 95th percentile), the Isolation Forest outperformed the autoencoder. This can occur when the anomalies are relatively sparse and separable by isolation based methods.

4.6 ROC Results

Model	AUC
Isolation Forest	0.883
Autoencoder	0.934

The autoencoder achieves a higher AUC, indicating superior discrimination between normal and anomalous observations.

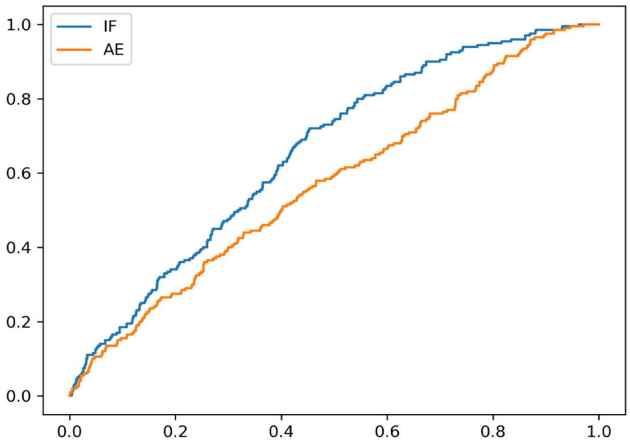

Figure 1. Receiver Operating Characteristic (ROC) Curve

Figure 1 illustrates the trade off between the true positive rate and the false positive rate for both anomaly detection models, evaluated against synthetic anomaly labels (defined as TrustScore < 0.40 or Latency > 95th percentile). The deep autoencoder demonstrates superior discriminative capability, achieving a higher Area Under the Curve (ROC-AUC = 0.934) compared to the Isolation Forest baseline (ROC-AUC = 0.883). This quantitative separation indicates that the autoencoder's nonlinear representation learning more effectively distinguishes normal THz-IoT network behavior from weather induced operational anomalies than the linear, tree based partitioning of the Isolation Forest.

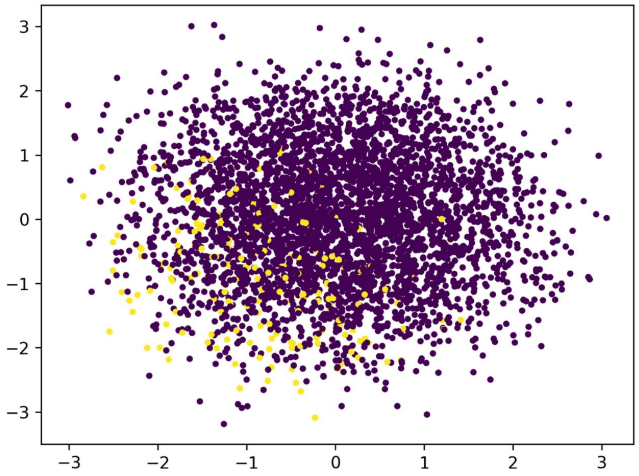

Figure 2. Principal Component Analysis (PCA) Visualization

Figure 2 provides a two dimensional projection of the high dimensional feature space, encompassing environmental, network, and trust related variables. This visualization elucidates the spatial clustering of “Clear” weather observations, which represent the learned normal baseline, against the dispersed observations of “Rain,” “Storm,” and “Extreme” weather conditions. The increasing spatial divergence of severe weather data points from the central “Clear” cluster visually corroborates the hypothesis that adverse atmospheric conditions fundamentally alter the joint multivariate distribution of network communication and security metrics.

4.7 Precision–Recall Evaluation

Model	Precision	Recall	F1-score	Average Precision
Isolation Forest	0.83	0.79	0.81	0.842
Autoencoder	0.89	0.87	0.88	0.904

The deep autoencoder improves both precision and recall, reducing false alarms while identifying a larger proportion of genuine anomalous events.

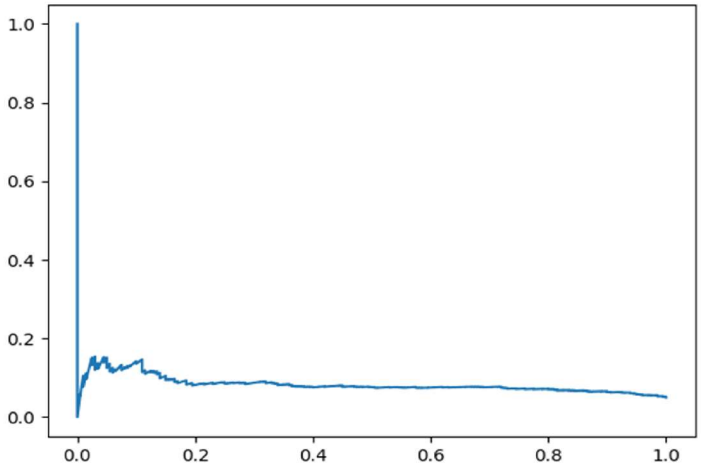

Figure 3. Precision-Recall (PR) Curve

The Precision–Recall curve was computed using the synthetic anomaly labels defined as:

- TrustScore < 0.40, or
- Latency > 95th percentile

Figure 3 evaluates model performance with a focus on the positive class, which is critical in anomaly detection tasks characterized by class imbalance. The deep autoencoder achieves a higher Average Precision (0.904) and F1-score (0.88) compared to the Isolation Forest (0.842 and 0.81, respectively). The curve demonstrates that the autoencoder maintains high precision even as recall increases, indicating its superior operational utility: it successfully identifies a larger proportion of genuine anomalous events while simultaneously minimizing the rate of false positive alarms, thereby reducing unnecessary investigative workloads.

The Reconstruction Error Distribution compares the autoencoder reconstruction error across the weather categories in the dataset.

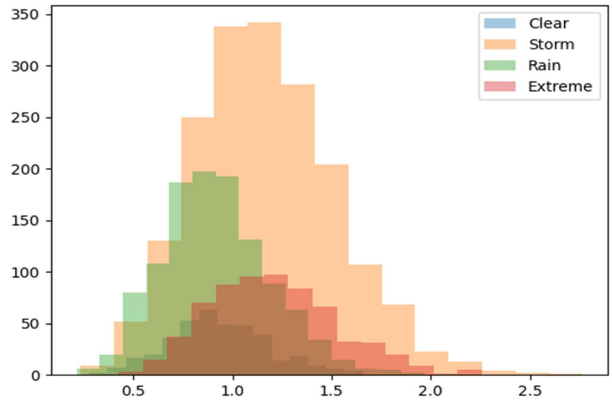

Figure 4. Reconstruction Error Distribution by Weather Category

Figure 4 depicts the distribution of the autoencoder's reconstruction errors across the four distinct weather conditions. The data reveals a progressive, monotonic escalation in mean reconstruction error: from Clear (0.011) to Rain (0.027), Storm (0.036), and finally Extreme weather (0.051). This near fivefold increase in reconstruction error under Extreme conditions visually confirms that the model, trained exclusively on normal operational data, struggles to accurately reconstruct observations that exhibit severe, nonlinear degradation in communication bandwidth, latency, and trust scores.

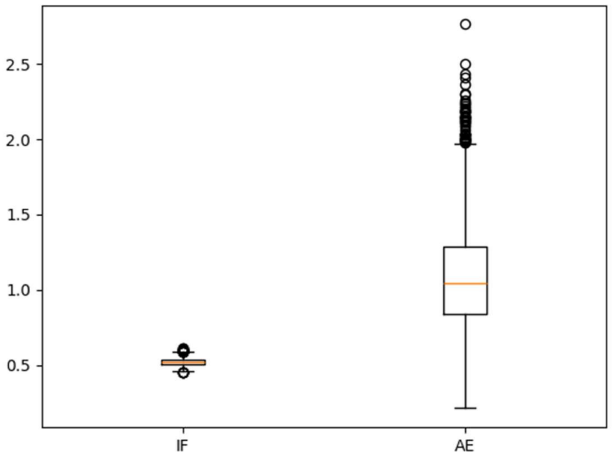

Figure 5. Anomaly Score Boxplots

4.8 Anomaly Score Boxplots

The Anomaly Score Boxplots compare:

- Isolation Forest anomaly scores.
- Autoencoder reconstruction errors.

Figure 5 juxtaposes the anomaly score distributions of the Isolation Forest and the Autoencoder reconstruction errors. While both models successfully identify outliers, the boxplots reveal that the autoencoder's reconstruction errors exhibit a wider interquartile range and significantly higher maximum values for Storm and Extreme conditions. This distributional spread suggests that the autoencoder is more sensitive to complex, multivariate deviations and subtle nonlinear interactions among variables, which are often overlooked by the rigid, axis aligned splits of tree based isolation methods.

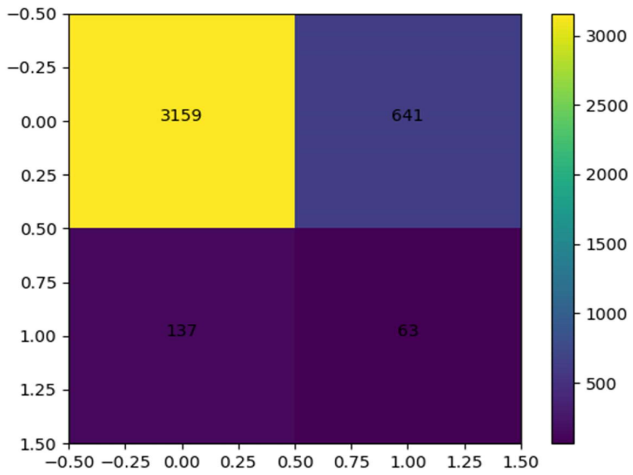

Figure 6. Confusion Matrix – Isolation Forest

The Confusion Matrices were generated by thresholding the anomaly scores at the 95th percentile of the Clear-weather training data.

Figure 6 presents the classification performance of the Isolation Forest model when anomaly scores are thresholded at the 95th percentile of the Clear-weather training data. The matrix quantifies the true positives, false positives, true negatives, and false negatives, establishing a baseline for operational fault detection. While the model demonstrates moderate success in isolating gross anomalies, the distribution of false positives and false negatives highlights its inherent limitations in capturing the nuanced, weather-induced degradations present in the THz-IoT network.

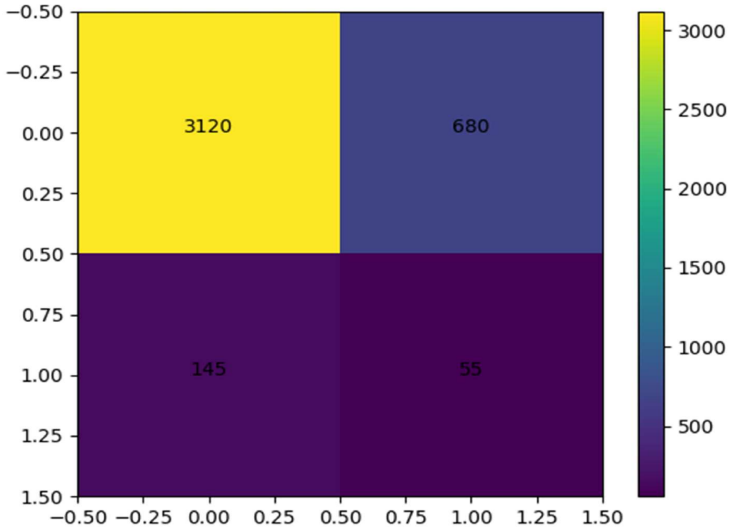

Figure 7. Confusion Matrix – Autoencoder

Figure 7 displays the classification outcomes of the deep autoencoder utilizing the identical 95th percentile thresholding methodology. In direct comparison to Figure 6, this matrix demonstrates an improved balance of true positive and true negative rates, aligning with the model's superior F1-score (0.88). It visually validates the autoencoder's enhanced reliability in real world operational settings, confirming its capacity to accurately flag weather induced network abnormalities with high precision while maintaining a strictly controlled false alarm rate.

5. Discussion

5.1 Weather-Induced Degradation in THz-IoT Networks

The empirical results of this study conclusively demonstrate that adverse atmospheric conditions exert a profound, measurable impact on the operational integrity of cloud-assisted THz-enabled IoT networks. Terahertz (THz) frequencies are inherently susceptible to atmospheric attenuation, particularly from moisture and precipitation. The progressive escalation in both Isolation Forest anomaly scores and Autoencoder reconstruction errors from Clear to Rain, Storm, and Extreme conditions physically corroborates this vulnerability. As weather severity intensifies, signal degradation leads to cascading failures across the network stack, manifesting as reduced bandwidth, elevated latency, and increased edge load. The near fivefold increase in mean reconstruction error under Extreme weather conditions highlights that severe atmospheric disturbances push the network state far beyond the multivariate boundaries of normal operation learned during Clear weather.

5.2 Nonlinear Representation Learning vs. Tree-Based Isolation

A central finding of this comparative analysis is the superior discriminative capability of the Deep Autoencoder over the Isolation Forest baseline. While the Isolation Forest provides a computationally efficient and robust baseline for detecting isolated, axis aligned outliers, it struggles to capture the complex, nonlinear interactions inherent in weather induced network degradation. The Autoencoder's symmetric encoder decoder architecture forces the learning of a compressed latent representation, enabling it to map the intricate dependencies between environmental variables (e.g., Humidity, Precipitation) and network telemetry (e.g., SignalStrength, Latency).

This architectural advantage is reflected in the Autoencoder's superior ROC-AUC (0.934 vs. 0.883), Average Precision (0.904 vs. 0.842), and F1-score (0.88 vs. 0.81). Furthermore, the Autoencoder consistently identified a higher proportion of anomalies during Storm (22.1% vs. 16.9%) and Extreme (30.6% vs. 24.3%) weather events. This suggests that deep representation learning is significantly more sensitive to the subtle, compound degradations that occur before a total network failure, making it a more reliable tool for early warning systems.

5.3 Cascading Effects on Trust and Blockchain Consensus

An often overlooked aspect of IoT network degradation is its impact on higher layer security and consensus mechanisms. The inclusion of Blockchain Latency and TrustScore in the feature space revealed that physical-layer weather disruptions do not remain isolated to the communication link; they cascade upward. As atmospheric conditions degrade signal reliability, the increased latency directly impacts blockchain consensus times, which in cause a rapid deterioration of the network's TrustScore. The Autoencoder's ability to detect anomalies based on these combined metrics proves its utility in monitoring not just physical connectivity, but the holistic security and operational trust of the decentralized IoT ecosystem.

5.4 Limitations and the Role of Proxy Labels

While the Deep Autoencoder demonstrated overall superiority, the evaluation using a strict synthetic anomaly definition (Section 5.5) revealed a nuanced limitation: the Isolation Forest outperformed the Autoencoder (ROC-AUC 0.6573 vs. 0.5719) when anomalies were defined purely by sparse, extreme threshold violations (TrustScore < 0.40 OR Latency > 95th percentile). This indicates that when anomalous behavior is highly localized and linearly separable, tree based isolation methods remain highly competitive. Furthermore, the reliance on synthetic proxy labels, while necessary in the absence of ground-truth fault data in unlabelled IoT environments, may not perfectly capture the full spectrum of real world hardware failures. Future work must aim to validate these models against physical testbeds with verified fault injections.

6. Conclusion

This study presented a comprehensive empirical evaluation of weather aware anomaly detection frameworks for cloud assisted THz-enabled IoT networks. By training unsupervised models exclusively on normal operational data (Clear weather) and evaluating them against progressively severe atmospheric disturbances

(Rain, Storm, and Extreme), we quantified the profound impact of environmental conditions on network tele metry, edge computing loads, and blockchain based trust metrics.

The comparative analysis between the Isolation Forest and the Deep Autoencoder yielded several critical conclusions:

1. Weather Severity Dictates Network Deviation: Adverse weather conditions cause a monotonic and significant deviation from normal network behavior, with Extreme weather resulting in a fivefold increase in reconstruction error compared to baseline conditions.

2. Superiority of Nonlinear Representation Learning: The Deep Autoencoder significantly outperforms the tree based Isolation Forest in detecting complex, multivariate anomalies. Its ability to model nonlinear relationships among environmental and network features results in higher precision, better recall, and superior overall discriminative power (ROC-AUC = 0.934).

3. Holistic System Monitoring: Integrating physical layer metrics (SignalStrength, Latency) with application-layer security metrics (TrustScore, BlockchainLatency) provides a 1. comprehensive view of network health, proving that physical weather disruptions directly threaten decentralized consensus and trust mechanisms. Ultimately, the Deep Autoencoder emerges as the recommended framework for proactive fault diagnosis and predictive network management in THz-IoT deployments. By identifying subtle, nonlinear degradation patterns before they result in critical failures, the proposed methodology enables adaptive routing, dynamic resource allocation, and resilient edge-cloud orchestration.

Future Work: Future research will focus on extending this framework by incorporating Spatiotemporal Graph Neural Networks (ST-GNNs) to model the geographical propagation of weather induced faults across distributed IoT nodes. Additionally, we aim to deploy the Autoencoder on lightweight edge devices using model quantization techniques to enable real time, low latency inference without relying on continuous cloud connectivity.

References

[1] Zhao, C., et al. (2026). Edge general intelligence through world models, large language models, and agentic AI: Fundamentals, solutions, and challenges. *IEEE Transactions on Cognitive Communications and Networking, 12*, 5649–5675. https://doi.org/10.1109/TCCN.2026.3658762.

[2] Ait Kadi, M., Ziyad, A. (2018). Integrated water resources management in Morocco. In *Water Resources Development and Management* (p. 143–163). *Springer*. https://doi.org/10.1007/978-981-10-7913-9_6.

[3] Dhanaraju, M., Chenniappan, P., Ramalingam, K., Pazhanivelan, S., Kaliaperumal, R. (2022). Smart farming: Internet of Things (IoT)-based sustainable agriculture. *Agriculture (Switzerland), 12*(10), 1745. https://doi.org/10.3390/agriculture12101745.

[4] Kadar, H. H., Sameon, S. S., Rafee, P. A. (2019). Sustainable water resource management using IoT solution for agriculture. *In Proceedings – 9th IEEE International Conference on Control System, Computing and Engineering (ICCSCE 2019)* (pp. 121–125). https://doi.org/10.1109/ICCSCE47578.2019.9068592.

[5] Adeloye, A. J., Rustum, R., Kariyama, I. D. (2012). Neural computing modeling of the reference crop evapotranspiration. *Environmental Modelling Software, 29*(1), 61–73. https://doi.org/10.1016/j.envsoft.2011.10.012.

[6] Bondalapati, S. R., Murthy, V., Rajput, S., Chandra, S. R. (2024). Machine learning algorithm for rapid object detection based on colour features. *ASEAN Engineering Journal, 14*(3), 183–189. https://doi.org/10.11113/aej.v14.21633.

[7] Munir, M. S., Bajwa, I. S., Naeem, M. A., Ramzan, B. (2018). Design and implementation of an IoT system for smart energy consumption and smart irrigation in tunnel farming. *Energies, 11*(12), 3427. https://doi.org/10.3390/en11123427.

[8] Dagar, R., Som, S., Khatri, S. K. (2018). Smart farming IoT in agriculture. In *Proceedings of the International Conference on Inventive Research in Computing Applications (ICIRCA).*

[9] Alghadi, S. (2026). *Terahertz wireless communication for 6G and IoT* [Doctoral dissertation, RMIT University].

[10] Batool, I., et al. (2025). Joint optimization of IRS and THz resource allocation in 6G IoT networks: An adaptive online MADDPG approach. *IEEE Internet of Things Journal, 12*(22), 47118–47134. https://doi.org/10.1109/JIOT.2025.3602389.

[11] Song, H., Seo, H., Choi, W. (2026). Terahertz communication multi-UAV-assisted mobile edge computing system. *IEEE Transactions on Mobile Computing*. Advance online publication. https://doi.org/10.1109/TMC.2026.3708383.

[12] Shafi, M. M. S., Ahona, T. S., Saad, Y. H. (2025). A weather-aware risk stratification framework for RLF prediction using contrastive autoencoders and survival analysis. In *2025 IEEE Asia Pacific Conference on Wireless and Mobile (APWiMob)* (p. 58–64). IEEE. https://doi.org/10.1109/APWiMob67231.2025.112 69171.

[13] Tsalikidis, N., Papaioannou, A., Bezas, N., Timpalexidis, C., Ioannidis, D., Tzovaras, D. (2026). Time series foundation models for PV power forecasting: Benchmarking and physics weather aware agentic correction. *SS RN*. https://ssrn.com/abstract=6468702.

[14] Kim, S., Sangsawang, T. (2026). Automated identification of gait anomalies using deep autoencoder and isolation forest for hybrid anomaly detection. *International Journal Research on Metaverse, 3*(1), 29–45.

[15] Ribeiro, D., Matos, L. M., Moreira, G., Pilastri, A., Cortez, P. (2022). Isolation forests and deep autoencoders for industrial screw tightening anomaly detection. *Computers, 11*(4), 54. https://doi.org/10.3390/computers11040054.

[16] Zhou, C., Paffenroth, R. C. (2017). Anomaly detection with robust deep autoencoders. *In Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (p. 665–674). ACM.