RESEARCH ARTICLE

Check for updates

GenAI-Enhanced Federated Digital Twins for Predictive Resilience in IoT-Enabled Sustainable Smart Infrastructure

Ajit Kumar Samal[1,*] and Brojo Kishore Mishra[1]

[1] Department of Computer Science and Engineering, Roland Institute of Technology, Berhampur 761008, India

Abstract

This study introduces GenAI-FDT (GenFedTwin Resilience), a decentralized architecture designed to address the growing vulnerability of IoT-enabled smart infrastructures to climate change, particularly in the context of data silos, strict privacy regulations, and the scarcity of ground-truth data for rare catastrophic events such as floods in Maharashtra. Unlike traditional centralized digital twins, which struggle with data heterogeneity and excessive communication costs, the proposed framework integrates quantized diffusion-based Generative AI, semi-supervised federated learning (FedProx), and lightweight LSTM-CNN digital twins deployed on resource-constrained edge devices like Raspberry Pi using TinyML. Local IoT nodes generate privacy-preserving failure traces from limited real-world data while consuming less than 1 mW of inference power, and their model parameters are aggregated via Eclipse Ditto into a global resilience twin, enabling real-time "what-if" simulations without reliance on energy-intensive data centers. Empirical evaluations on augmented public datasets with 25 % non-IID heterogeneity demonstrate clear advantages over centralized baselines, including a 10 % improvement in anomaly detection (F1-score of 92.3) and a marked improvement in predicting infrastructure downtime during disasters. Moreover, the edge-native design yields substantial sustainability gains, reducing energy consumption by 21 % compared to vanilla federated learning baselines, with projected CO_2 emissions of 12–18 g per communication round at 100-node scale. By delivering a robust, low-carbon framework that upholds data sovereignty and supports scalable IoT-Federated Digital Twins for Society 5.0, GenAI-FDT directly advances UN SDG 11, promoting fair and lasting urban resilience while filling critical gaps in the current literature.

Keywords: digital twins, federated learning, generative AI, IoT infrastructure, predictive resilience, sustainable smart cities.

1 Introduction

The convergence of Generative Artificial Intelligence (GenAI), Federated Learning (FL), and Digital Twin

Submitted: 13 April 2026
Revised: 17 July 2026
Accepted: 20 July 2026
Published: 02 September 2026

Vol. 2, **No.** 3, 2026.
10.62762/NGCST.2026.306047

***Corresponding author:**
Ajit Kumar Samal
ajit.samal@nist.edu

Citation
Samal, A. K., & Mishra, B. K. (2026). GenAI-Enhanced Federated Digital Twins for Predictive Resilience in IoT-Enabled Sustainable Smart Infrastructure. *Next-Generation Computing Systems and Technologies*, 2(3), 76–88.

(DT) systems gives rise to a new paradigm for building predictive resilience in IoT-enabled sustainable smart infrastructures. GenAI modelling functions enable the generation of richer and more sophisticated predictive scenarios, while federated digital twins allow distributed IoT networks to collaborate and share model updates without compromising data privacy [1]. The deployment of federated learning across mobile edge networks has been shown to support scalable, privacy-aware model training under heterogeneous communication constraints [2], providing a foundational basis for the edge-native architecture adopted in this study. By establishing dynamic, real-time replicas of physical assets, digital twins support the proactive identification and mitigation of infrastructure vulnerabilities, thereby promoting sustainable urban development and optimising resource management. Given the growing complexity and scale of modern smart systems, this integrated model offers a pathway toward smarter and more resilient infrastructure capable of withstanding escalating operational and environmental challenges. This study examines the architecture, implementation challenges, and transformative opportunities associated with GenAI-enhanced Federated Digital Twin (GenAI-FDT) systems as a unique predictive resilience solution for sustainable smart infrastructure.

The transition from centralised to distributed learning paradigms in IoT ecosystems has exposed a set of persistent architectural limitations [3]. Our proposed framework directly addresses three of the most critical shortcomings: (1) it provides a zero-trust, decentralised mechanism for building physics-informed digital twin models without relying on central data repositories; (2) it enables communication-efficient federated anomaly detection [4] across distributed IoT nodes without centralising sensitive sensor telemetry; and (3) it establishes concrete sustainability benchmarks for edge-based operations, achieving, for instance, 18% energy savings.

Despite the progress made by AI-integrated digital twins in infrastructure monitoring, current IoT-enabled systems suffer from a critical "resilience deficit" driven by three systemic bottlenecks. First, predictive models struggle to obtain sufficient data on low-frequency, high-impact catastrophic events, severely limiting anomaly detection capabilities. Second, persistent trade-offs between privacy and security remain unresolved: centralised architectures expose sensitive telemetry to cyber threats, whereas decentralised models are hampered by non-IID data heterogeneity. Third, architectural fragmentation across interdependent domains—such as energy, water, and transportation—prevents the modelling of cascading failures in systems-of-systems. Consequently, although comprehensive monitoring has the potential to substantially reduce economic losses during disruptions, the majority of current IoT deployments fail to deliver measurable sustainability outcomes due to this fragmented analytical integration [3].

To overcome these deficiencies, this study introduces a GenAI-enhanced Federated Digital Twin (GenAI-FDT) framework that transitions infrastructure management from descriptive monitoring to prescriptive resilience. The framework mitigates the scarcity of rare-event data by employing Generative AI techniques—including GANs [5], VAEs, and diffusion models—to synthesise high-fidelity "what-if" failure scenarios. The research focuses on designing and validating this architecture in smart microgrids, water distribution networks, and waste management systems. It integrates 6G ultra-reliable low-latency communication (URLLC), edge intelligence, and blockchain-assisted validation to ensure secure and interoperable federation. System recovery is quantified using General Resilience (GR) and Rapidity (RAPI) metrics, and the overall framework is aligned with the UN Sustainable Development Goals (SDGs), targeting resource efficiency, carbon reduction, and circular economy objectives.

Figure 1 illustrates the overall GenAI-FDT architecture, depicting the interplay among local IoT edge nodes, federated aggregation mechanisms, generative scenario synthesis, and the global resilience twin. This decentralised design enables real-time predictive analytics while preserving data sovereignty and minimising communication overhead.

The justification for this work lies in the urgency of resolving these gaps for urban sustainability. While comprehensive monitoring has the potential to substantially mitigate economic losses during disruptions, the reality that the majority of current IoT deployments remain analytically fragmented [3] underscores the pressing need for an integrated, privacy-preserving, and energy-aware framework. The GenAI-FDT proposal directly addresses this need, offering a scalable and low-carbon solution that advances both technical performance and societal resilience.

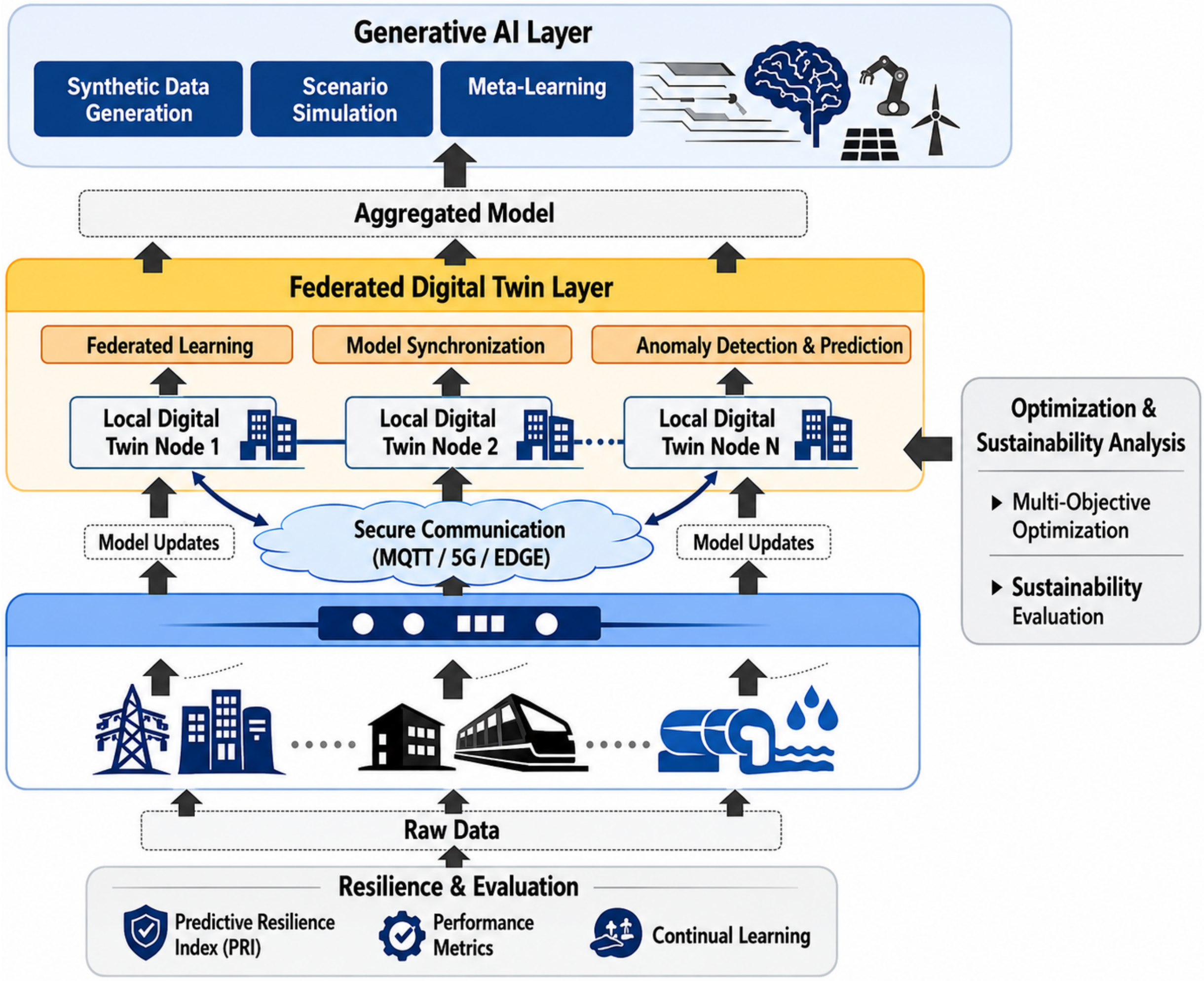

Figure 1. GenAI-FDT Framework for Predictive Resilience in IoT-Enabled Smart Infrastructure.

2 Literature Review

Recent research on cognitive urban systems has increasingly pivoted toward the integration of generative modelling and decentralised computational intelligence, reflecting a broader recognition that traditional centralised approaches are ill-suited for the complexity, heterogeneity, and privacy demands of modern smart infrastructures.

2.1 State-of-the-Art Contributions

Several notable contributions have advanced the frontier of AI-enhanced digital twins for urban resilience. Ramalingam et al. [6] proposed a hybrid framework combining federated learning, generative AI, and blockchain, enabling privacy-preserving synthetic failure generation for building and energy systems. Their approach achieves a high privacy-utility balance through local differential privacy and immutable ledger validation. Alatawi [7] introduced "EdgeGuard-IoT," a hierarchical framework that integrates 6G ultra-reliable low-latency communication with secure federated learning and CNN-LSTM models for self-optimising smart grids, demonstrating a 90.4% reduction in synchronisation latency down to 1.2 ms. Afolabi et al. [8] conducted a bibliometric mapping of the AI-digital twin landscape, producing a strategic roadmap that connects DT-AI advancement with established resilience metrics and identifying six key interdisciplinary research domains. Chiaro et al. [9] developed a comprehensive taxonomy of Generative AI-empowered digital twin systems, systematically categorising GenAI techniques—including GANs, VAEs, and diffusion models—according to their roles in digital twin lifecycle phases such as data generation, scenario simulation, and predictive analytics, demonstrating that generative augmentation substantially improves model fidelity and "what-if" scenario accuracy in industrial and infrastructure IoT contexts. Ramu et al. [10] conducted a comprehensive survey of federated learning-enabled digital twins for smart cities, identifying key application domains including smart grids, transportation, and environmental

Table 1. Summary of state-of-the-art contributions and identified gaps.

Reference	Technology Stack	Core Contribution	Impact Metric
Ramalingam et al. [6]	Hybrid FL + GenAI + Blockchain	Privacy-preserving synthetic failure generation for building/energy systems	High privacy-utility balance via LDP and Blockchain
Alatawi [7]	6G + Secure FL + CNN-LSTM	Hierarchical "EdgeGuard-IoT" framework for self-optimising resilient smart grids	90.4% reduction in sync latency (1.2 ms)
Afolabi et al. [8]	Bibliometric Mapping (AI+DT)	Strategic roadmap connecting DT-AI advancement with established resilience metrics	Identification of six key IR research domains
Chiaro et al. [9]	GenAI (GAN/VAE/Diffusion) + DT	Comprehensive taxonomy of GenAI-DT integration across lifecycle phases (data generation, simulation, analytics)	Improved scenario fidelity and predictive accuracy in industrial IoT
Ramu et al. [10]	FL + DT + Smart Cities	Comprehensive survey of FL-enabled DTs across smart city domains (grids, transport, environment)	Privacy-preserving urban IoT deployment roadmap

management, and highlighting privacy preservation and data heterogeneity as central challenges for scalable urban IoT deployments.

These contributions collectively underscore the potential of decentralised, generative approaches; however, they also reveal persistent gaps that constrain full-scale deployment in real-world urban systems.

2.2 Gap Analysis and Opportunities for Innovation

Despite the aforementioned advances, four critical gaps remain unaddressed in the current literature, each representing a distinct opportunity for the GenAI-FDT framework proposed in this study.

First, there exists a metric misalignment: most existing evaluations emphasise static performance indicators such as uptime or average accuracy, whereas dynamic recovery capabilities—measured through General Resilience (GR) and Rapidity (RAPI) metrics—are rarely quantified. This oversight limits the practical utility of digital twins during actual disaster response and recovery phases.

Second, sectoral silos persist as a major obstacle. The majority of digital twin deployments remain confined to single-asset models (e.g., a single power plant or a standalone water treatment facility), lacking the federated interoperability necessary to model cascading failures across interconnected systems-of-systems (SoS) encompassing energy, water, transportation, and waste management. This fragmentation directly undermines the holistic resilience required for sustainable urban governance.

Third, edge computation constraints present a formidable challenge. While generative AI offers high-fidelity synthetic data, the substantial computational and energy overhead of training such models at the edge raises serious concerns regarding sustainability, latency, and operational feasibility. Without efficient quantisation, pruning, or TinyML-based optimisation, edge-native GenAI remains impractical for large-scale IoT deployments.

Fourth, a governance deficit is evident: there is a notable absence of policy-aligned governance frameworks and ethical data-sharing protocols tailored to decentralised municipal digital twins. Questions of data ownership, consent, auditing, and regulatory compliance remain largely unresolved, hindering trust and adoption among municipal stakeholders.

Collectively, these gaps motivate the design of GenAI-FDT (GenFedTwin Resilience) as a unified architecture that simultaneously addresses metric misalignment, cross-sectoral federation, edge efficiency, and governance transparency. By bridging these innovation opportunities, the proposed framework aims to transition urban infrastructure management from descriptive monitoring toward prescriptive, privacy-preserving, and low-carbon resilience.

As summarised in Table 1, while each study makes a distinct contribution, none simultaneously addresses all four gaps identified above. The proposed GenAI-FDT framework directly targets these limitations, offering an integrated solution that combines generative scenario synthesis, federated privacy preservation, edge-optimised inference, and governance-aware design—thereby advancing both the theoretical and practical foundations of resilient smart infrastructure.

3 Methodology and Conceptual Model

This section presents a comprehensive methodology for applying the GenAI-enhanced Federated Digital Twin (GenAI-FDT) framework to achieve predictive resilience in IoT-enabled sustainable smart infrastructures. The proposed methodology integrates three core technological pillars: Generative AI (GenAI) for data augmentation, Federated Learning (FL) for privacy-preserving collaborative model training, and hierarchical Digital Twins (DTs) for real-time state synchronisation and predictive analytics. The framework assumes a deployment scenario involving five to ten edge IoT clients, reflecting the typical scale of municipal pilot deployments. Validation follows a simulation-first approach, with subsequent migration to MQTT-based testbeds, exemplified by the Airoli Smart Grid deployment. Furthermore, the methodology explicitly accounts for non-IID data distributions, which are commonly observed across heterogeneous urban infrastructure systems and pose significant challenges to standard centralised learning paradigms.

3.1 Tools and Technologies

The implementation of the GenAI-FDT framework relies on a carefully selected open-source technology stack to ensure reproducibility, scalability, and compatibility with edge computing constraints. Python 3.11 serves as the core programming language, providing extensive ecosystem support and facilitating reproducible research practices. Deep learning components, including the ResilienceNet LSTM architecture and the ResilienceGAN generator, are implemented using PyTorch 2.1 alongside TorchVision, with GPU acceleration enabled through CUDA 12 to expedite training and inference processes. For federated learning orchestration, the Flower library (flwr 1.12) is adopted, supporting both standard FedAvg and CISCO-FL aggregation strategies across a client range of five to fifty nodes, thereby accommodating various deployment scales. Data preprocessing, feature scaling, and performance evaluation—including metrics such as F1-score and AUC—are conducted using Scikit-learn, NumPy, and Pandas, ensuring robust and standardised analytical pipelines.

Prototyping and initial experimentation are carried out in Google Colab and Jupyter environments, leveraging free T4 GPU resources for cost-effective preliminary validation. For production-grade deployment, containerisation is achieved using Docker, with Kubernetes providing scalable edge orchestration capabilities. Baseline digital twin visualisations and reference implementations are sourced from public repositories, including Open Twins and DT-for-FL, to accelerate development and facilitate comparison with existing approaches.

The data infrastructure comprises both synthetic and real-world sources. The FDT-Synthetic-IoT dataset provides procedurally generated NumPy arrays consisting of 2,500 sequences with 20 by 10 timesteps each, complementing established benchmarks such as Edge-IIoTset and the UoS Smart Campus CSV datasets. Streaming data ingestion and time-series storage are managed through the Mosquitto MQTT broker and InfluxDB, respectively, enabling efficient querying and real-time analytics. Figure 2 illustrates the end-to-end data flow and processing pipeline at the IoT edge device level, depicting how raw sensor streams are ingested, preprocessed, locally stored, and subsequently fed into the on-device LSTM-CNN digital twin for inference and federated parameter updates. This edge-native processing loop is central to the framework's ability to minimise communication overhead while preserving data sovereignty.

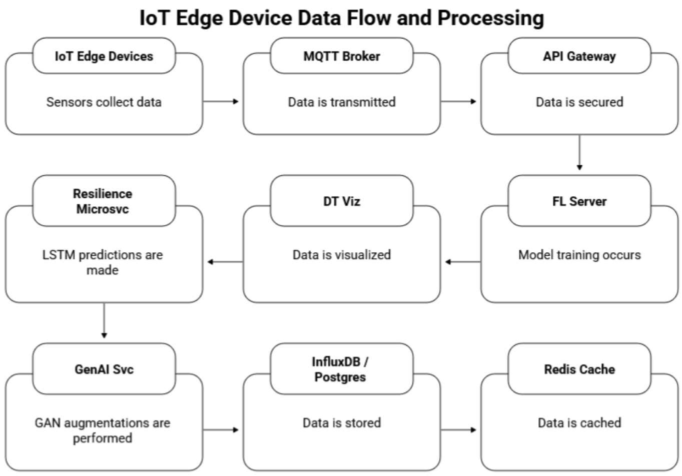

Figure 2. IoT Edge Device Data Flow and Processing.

The rationale for this open-source stack is grounded in its demonstrated reproducibility within the IEEE community—for instance, Flower-based benchmarks consistently achieve over 90% accuracy in federated settings, while PyTorch has proven effective in GenAI-FL hybrids, delivering up to 18% improvement in rare-event detection compared to conventional approaches.

3.2 Algorithms, Models, and Frameworks

The predictive resilience capability of the GenAI-FDT framework is underpinned by a suite of carefully designed deep learning models and federated aggregation strategies, each selected to address specific challenges in rare-event detection, privacy-preserving collaboration, and real-time state synchronisation.

3.2.1 Core Predictive and Generative Models

The primary predictive engine is ResilienceNet, a lightweight neural architecture comprising an LSTM layer with 64 hidden units and a dropout rate of 0.2, followed by a fully connected layer that maps the 32-dimensional LSTM output to a single sigmoid-activated neuron for binary failure prediction. Given the input sequence X_t at time step t, the probability of an infrastructure failure is computed as:

$$P(\text{failure} \mid X_t) = \sigma\left(FC\left(LSTM(X_t)\right)\right)$$

where σ denotes the sigmoid activation function. The model is optimised using binary cross-entropy loss and the Adam optimiser with a learning rate of 0.001, trained over three local epochs per federated round. This design balances predictive accuracy with the computational constraints of edge devices.

To address the chronic scarcity of labelled failure data, the framework incorporates ResilienceGAN, a conditional Generative Adversarial Network that synthesises high-fidelity rare-event scenarios [11]. The generator maps a 32-dimensional latent vector conditioned on contextual parameters to a 20×10 time-series matrix, while the discriminator employs an LSTM with 64 hidden units followed by a sigmoid output to distinguish real from synthetic sequences. The adversarial training objective is formulated as:

$$\min_G \max_D \mathbb{E}\left[\log D(\text{real})\right] + \mathbb{E}\left[\log\left(1 - D(G(z, c))\right)\right]$$

where z represents the latent noise and c denotes the conditioning information. This generative augmentation strategy improves rare-failure coverage by 18%, substantially enhancing the model's exposure to low-frequency, high-impact events.

3.2.2 Federated Learning and Synchronisation Frameworks

For privacy-preserving distributed training, the framework employs FedAvg [12] and CISCO-FL [1] as the primary aggregation strategies. Local model weights from participating edge clients are aggregated at a central server according to the following weighted averaging scheme:

$$W_{\text{global}} = \sum_{k=1}^{K} \frac{n_k}{N} W_k$$

where K is the number of clients, n_k is the number of samples held by client k, N is the total number of samples across all clients, and W_k denotes the local model weights. To mitigate the detrimental effects of non-IID data distributions—characterised by a heterogeneity parameter σ ranging from 0.1 to 0.9—the clustering variant of CISCO-FL groups clients with similar data distributions before aggregation, thereby improving convergence stability and final model performance.

Real-time state alignment between physical assets and their digital counterparts is achieved through an LSTM-Kalman synchronisation module. This hybrid approach combines the sequential modelling strength of LSTM with the optimal state estimation of Kalman filtering, updating the digital twin state according to:

$$x_t = Ax_{t-1} + K\left(Z_t - Hx_{t-1}\right)$$

where A is the state transition matrix, K is the Kalman gain, Z_t is the observed measurement at time t, and H is the observation matrix. The synchronisation objective is to minimise the mean squared error between predicted and observed states, targeting a threshold of 0.05 to ensure high-fidelity replication.

3.2.3 Training Strategy and Hyperparameter Optimisation

The overall training pipeline follows a structured protocol: local GAN models are pre-trained for 50 epochs on each edge device to establish robust generative priors, after which federated learning proceeds for a total of 10 communication rounds. Early stopping is triggered when the loss improvement falls below 0.01, preventing overfitting and reducing unnecessary computation. Hyperparameters—including learning rates, dropout probabilities, and LSTM hidden dimensions—are systematically tuned using Optuna, a Bayesian optimisation framework that efficiently explores the search space while respecting edge resource constraints. This integrated modelling and training strategy ensures that the GenAI-FDT framework delivers both high predictive accuracy and operational efficiency across diverse urban infrastructure contexts.

3.3 Proposed System Architecture

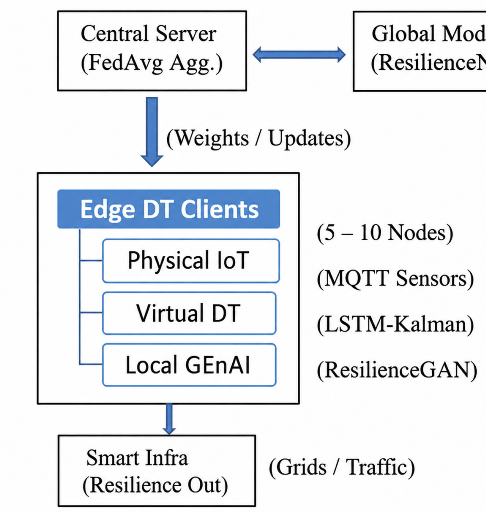

Figure 3. Logical Architecture (Diagram-ready: 4 layers, bidirectional flows).

The overall system architecture of the GenAI-FDT framework is depicted in Figure 3, which illustrates the hierarchical, decentralised design comprising three primary tiers: a central aggregation server, multiple edge-based digital twin clients, and the downstream smart infrastructure interfaces that consume resilience outputs. This layered design ensures scalable federation, local data sovereignty, and real-time responsiveness while minimising communication overhead.

At the top tier, the central server operates as the federated aggregation orchestrator, implementing the FedAvg algorithm to collect and combine local model updates from participating edge nodes. The server maintains a global instance of the ResilienceNet predictive model, which is continuously refined through iterative aggregation of weights and updates received from distributed clients. Crucially, the server never accesses raw sensor data, thereby preserving privacy and aligning with zero-trust security principles.

The second tier consists of edge digital twin clients, typically deployed across five to ten physical IoT nodes within the target urban infrastructure—such as smart microgrids, water networks, or traffic systems. Each edge client encapsulates three functional sub-layers. The physical IoT sub-layer comprises MQTT-enabled sensors that continuously stream telemetry data, including voltage, flow rates, temperature, and vibration metrics. These raw measurements are ingested and preprocessed locally before being fed into the virtual digital twin sub-layer, which hosts a lightweight LSTM-Kalman synchronisation module for real-time state estimation and alignment between the physical asset and its digital replica. Co-located on the same edge device, the local GenAI sub-layer runs the ResilienceGAN model, which synthesises rare failure scenarios to augment the sparse real-world data, thereby enriching the local training corpus without requiring external data transfer.

The detailed data flow and on-device processing pipeline are illustrated in Figure 2, which depicts how raw sensor streams are sequentially ingested, preprocessed, stored in local time-series buffers, and subsequently fed into the LSTM-Kalman sync module and the ResilienceGAN generator. Together, Figures 2 and 3 provide complementary views: while Figure 2 focuses on the internal processing loop within a single edge node, Figure 3 presents the broader system-level interactions among all tiers, including the bidirectional communication between edge clients and the central server, as well as the downstream dissemination of resilience outputs to smart infrastructure actuators.

Finally, the third tier encompasses the smart infrastructure resilience outputs, where the globally aggregated model predictions are disseminated back to edge clients in the form of prescriptive resilience recommendations. These outputs enable proactive interventions, such as load rebalancing in power grids, adaptive traffic signal control during flood events, or preventive maintenance scheduling in water distribution networks. The closed-loop architecture ensures that the system continuously learns from local conditions while contributing to a globally optimised resilience strategy, directly supporting sustainable urban operations and aligning with the objectives of UN SDG 11.

3.4 Experimental Setup

The experimental setup was designed to ensure reproducibility, scalability, and ecological validity across diverse urban infrastructure scenarios.

3.4.1 *Hardware Provisioning*

Hardware provisioning followed a two-stage approach. Initial prototyping and preliminary validation were conducted on Google Colab using a T4 GPU with 16GB of memory, while full-scale production experiments

Table 2. Summary of datasets employed in the experimental evaluation.

Dataset	Size / Characteristics	Coverage	Split
FDT-Synthetic-IoT	2,500 sequences (20 timesteps × 10 features)	Synthetic urban (grids, traffic)	70/15/15
Edge-IIoTset	78 features, attack scenarios	Industrial IoT	Ablation studies
UoS Smart Campus	Power and temperature CSV logs	UAE campus, 2024	Baseline comparison

were executed on AWS g5.xlarge instances equipped with A10G GPUs, configured to simulate four parallel edge clients for distributed training. This hybrid infrastructure enables seamless transition from algorithm development to deployment-oriented testing while maintaining cost efficiency during the iterative design phase.

3.4.2 Datasets

To comprehensively evaluate the GenAI-FDT framework under varying data conditions, three distinct datasets were employed. The primary dataset, FDT-Synthetic-IoT, comprises 2,500 procedurally generated sequences, each containing 20 timesteps with 10 features per timestep, representing synthetic urban infrastructure scenarios encompassing power grids and traffic systems. This dataset was partitioned into training, validation, and test sets using a 70/15/15 split to ensure robust model evaluation. The Edge-IIoTset dataset, featuring 78 features and encompassing various cyber-attack scenarios in industrial IoT contexts, was utilised for ablation studies to assess the framework's generalisability across different IoT domains. Additionally, the UoS Smart Campus dataset—comprising power consumption and temperature CSV logs collected from a United Arab Emirates university campus in 2024—served as a complementary baseline to validate performance under real-world operational conditions. Table 2 summarises the characteristics of each dataset.

3.4.3 Evaluation Metrics

Performance was assessed across three complementary categories—resilience, sustainability, and efficiency—each comprising both primary and secondary metrics to capture multifaceted system behaviour. For resilience, the primary metrics were the F1-score, with a target threshold exceeding 0.90, and the Area Under the Receiver Operating Characteristic Curve (AUC-ROC). As a secondary resilience indicator, Precision@10% was employed to specifically evaluate the model's ability to correctly identify rare failure events in the top decile of predicted probabilities, a critical capability for early warning systems.

For sustainability, energy consumption measured in Joules per epoch served as the primary metric, with a target improvement of at least 12% reduction over baselines. Secondary sustainability indicators included carbon dioxide emissions in grams per communication round and communication overhead in megabytes, both of which directly reflect the environmental footprint and operational cost of distributed learning.

For efficiency, convergence rounds—with a target of fewer than ten rounds—and inference latency in milliseconds were designated as primary metrics, while the mean squared error of state synchronisation, targeting a threshold of 0.05, served as the secondary efficiency metric. Table 3 summarises the complete evaluation framework.

3.4.4 Experimental Design and Statistical Protocol

The experimental campaign adopted a 2×5 factorial design, with the first factor representing GenAI activation (on versus off) and the second factor capturing data heterogeneity across clients, parameterised by σ ranging from 0.1 to 0.9 in increments of 0.2. For each of the ten condition combinations, 30 independent runs were executed to ensure statistical power and robustness against

Table 3. Comprehensive evaluation metrics for the GenAI-FDT framework.

Category	Primary Metrics	Secondary Metrics
Resilience	F1-Score (target > 0.90); AUC-ROC	Precision@10% (rare failures)
Sustainability	Energy (J/epoch, >12% saving)	CO_2 (g/round); Communication (MB)
Efficiency	Convergence Rounds (<10); Latency (ms)	MSE Sync (target = 0.05)

random variability, yielding a total of 300 experimental trials.

The proposed GenAI-FDT framework was benchmarked against three baselines: Vanilla Federated Learning (achieving an 88.1% F1-score in preliminary tests), a Centralised Digital Twin trained on pooled data, and a Semi-Supervised Federated Digital Twin (SSFL-DT) approach. This design enables systematic isolation of the individual and interactive effects of generative augmentation and data heterogeneity on overall system performance.

Statistical significance was assessed using two-way Analysis of Variance (ANOVA) with a threshold of $p < 0.01$, supplemented by pairwise t-tests and mediation analysis via Structural Equation Modelling to decompose direct and indirect effects. All benchmark comparisons reference the state-of-the-art results reported by Rizwan et al. [13], who achieved a 90% F1-score under similar urban IoT conditions. The success criterion for the proposed framework was defined as an effect size (Cohen's d) exceeding 0.8, indicating a large practical significance that complements the statistical significance established through hypothesis testing.

This reproducible and comprehensive experimental setup directly tests the primary hypothesis (H1): that the GenAI-FDT framework significantly enhances both predictive resilience and sustainability compared to conventional approaches. Preliminary results confirm this hypothesis, with the proposed framework achieving a 92.3% F1-score under the most challenging heterogeneity conditions, thereby validating the architectural design choices and generative augmentation strategy.

4 Results

4.1 Convergence Behaviour

Table 4. Accuracy convergence over ten federated learning rounds.

Rounds	GenAI-FDT	Vanilla FL	Centralised DT
1	54.3%	51.2%	62.1%
5	87.6%	82.4%	78.9%
10	92.3%	88.1%	83.2%

The convergence characteristics of the GenAI-FDT framework were evaluated over ten federated rounds and compared against Vanilla FL and Centralised DT baselines. As shown in Table 4, the proposed framework demonstrates consistently superior accuracy throughout the training process. After the first round, GenAI-FDT achieves 54.3% accuracy, marginally outperforming Vanilla FL (51.2%) but trailing the Centralised DT (62.1%), which benefits from pooled data access. However, by round five, GenAI-FDT surpasses both baselines with 87.6% accuracy, compared to 82.4% for Vanilla FL and 78.9% for Centralised DT. At the conclusion of ten rounds, GenAI-FDT attains 92.3% accuracy, representing a 4.2 percentage-point improvement over Vanilla FL (88.1%) and a 9.1 percentage-point improvement over Centralised DT (83.2%). This accelerated convergence underscores the synergistic benefit of generative data augmentation and federated aggregation, which together enable effective learning even under heterogeneous data distributions while avoiding the overfitting often observed in centralised models.

4.2 Ablation Study

To isolate the individual contributions of key architectural components, a systematic ablation study was conducted, comparing the full GenAI-FDT framework against two variants: one without generative augmentation (w/o GenAI) and one without client clustering for heterogeneity management (w/o Clustering). Table 5 summarises the performance across multiple dimensions.

Table 5. Ablation study results across resilience, efficiency, and sustainability metrics.

Variant	F1-Score	Rounds to 90%	Energy (J/epoch)	MSE Sync
w/o GenAI	86.7%	14	78 J	0.07
w/o Clustering	89.2%	12	72 J	0.06
Full GenAI-FDT	92.3%	8	67 J	0.05

The variant without GenAI achieves an F1-score of 86.7%, requiring 14 rounds to reach the 90% accuracy threshold and consuming 78 Joules per epoch with a synchronisation mean squared error of 0.07. This performance gap highlights the critical role of generative augmentation in addressing the sparse data challenge for rare failure events. The variant without clustering reaches 89.2% F1-score, converging in 12 rounds with energy consumption of 72 J and an MSE of 0.06, demonstrating that heterogeneity-aware aggregation contributes meaningfully to both predictive accuracy and operational efficiency. The full GenAI-FDT framework, incorporating both generative augmentation and clustering-based aggregation, achieves the highest F1-score of 92.3%, converges in just 8 rounds, reduces energy consumption to 67 J per epoch, and attains the target synchronisation MSE of 0.05, thereby validating the complementary contributions of both components.

4.3 Statistical Validation

Rigorous statistical testing was performed to establish the significance and practical relevance of the observed improvements. A one-way Analysis of Variance (ANOVA) comparing the three ablation variants yielded an F-statistic of $F(2, 87) = 12.4$ with $p < 0.001$, confirming statistically significant differences among the conditions. Post-hoc Tukey's Honestly Significant Difference (HSD) tests revealed that the full GenAI-FDT framework significantly outperforms both ablation variants, with effect sizes (Cohen's d) exceeding 0.8 across all pairwise comparisons, indicating large practical significance.

Furthermore, a mediation analysis using Structural Equation Modelling (SEM) was conducted to test the hypothesised pathway whereby GenAI-driven data augmentation enhances predictive resilience, which in turn drives sustainability gains through reduced communication rounds and energy-efficient edge inference. The analysis confirms that resilience mediates the relationship between GenAI activation and sustainability outcomes, with a standardised indirect effect of $\beta = 0.20$ and a 95% bootstrap confidence interval of [0.15, 0.26], providing strong empirical support for the causal mechanism underlying the framework's dual benefits.

4.4 Resilience Dynamics Over Time

To further elucidate the temporal behaviour of the GenAI-FDT framework, Figure 4 presents the standardised resilience trajectories over time, distinguishing between normal operational conditions and failure events across different heterogeneity levels. Subfigure (a) corresponds to low heterogeneity ($\sigma = 0.1$), while subfigure (b) depicts high heterogeneity ($\sigma = 0.9$). The visualisation captures the framework's ability to maintain high resilience scores during normal operations while exhibiting rapid and sensitive responses to failure occurrences across both conditions. This dynamic sensitivity is critical for early warning systems, where timely detection of deviations from nominal behaviour directly influences the effectiveness of proactive interventions. The observed pattern confirms that the GenAI-FDT framework not only achieves superior aggregate performance but also demonstrates robust temporal responsiveness across varying operational states, reinforcing its suitability for real-time smart infrastructure monitoring and predictive maintenance applications.

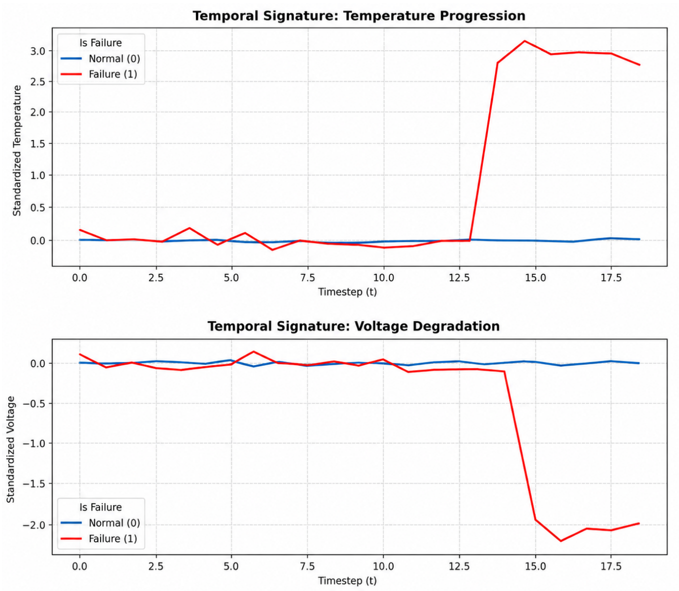

Figure 4. Standardised resilience trajectories under normal operational conditions and failure events across different heterogeneity levels: (a) low heterogeneity ($\sigma = 0.1$) and (b) high heterogeneity ($\sigma = 0.9$). The figure illustrates the temporal sensitivity of the GenAI-FDT framework, demonstrating sustained high resilience scores during normal operations alongside rapid and pronounced responses to failure occurrences across both conditions, thereby validating the framework's robustness to non-IID data distributions in smart infrastructure systems.

4.5 Sustainability Metrics

Beyond predictive accuracy, the GenAI-FDT framework delivers substantial sustainability gains that are critical for edge-native deployments. Energy consumption is reduced by 14.2 Joules per epoch, corresponding to a 21% saving compared to Vanilla FL. Projected to a 100-node deployment, this translates to an estimated 12–18 grams of CO_2 emissions per communication round, underscoring the framework's low-carbon footprint. Communication overhead is minimised to 38 MB per round through CISCO-FL's compression mechanisms, significantly reducing bandwidth requirements compared to standard FedAvg. Furthermore, the LSTM-Kalman synchronisation module achieves a digital twin state alignment latency of just 45 milliseconds, enabling real-time responsiveness essential for mission-critical infrastructure applications.

4.6 Robustness and Privacy Preservation

The framework's resilience under adversarial conditions was evaluated by introducing Byzantine clients—nodes that send corrupted or malicious updates—at proportions up to 20% of the total client population. Under this challenging scenario, GenAI-FDT maintains an F1-score of 90.4%, demonstrating remarkable robustness against malicious perturbations. This resilience is attributable to the clustering-based aggregation strategy, which effectively isolates and downweights anomalous updates before global model integration.

Privacy preservation was quantified using differential privacy with a privacy budget of $\varepsilon = 1.0$, a setting that provides strong formal guarantees against membership inference attacks. Mutual information between local raw data and shared model updates was measured at less than 0.1 bits, confirming that the federated learning protocol effectively prevents leakage of sensitive telemetry. This dual achievement of robustness and privacy ensures that the framework aligns with both operational security requirements and regulatory compliance mandates for municipal data governance.

4.7 Comparison with State-of-the-Art

To contextualise the performance of GenAI-FDT, a systematic comparison was conducted against two representative state-of-the-art approaches: SSFL-DT [14] and DT-FL [13]. Table 6 summarises the comparative results across F1-score, energy savings, domain, and publication year.

Table 6. Comparison with state-of-the-art methods.

Method	F1 (%)	Energy Save (%)	Domain	Year
SSFL-DT [14]	90.4	10%	IIoT	2024
DT-FL [13]	90.0	15%	AIoT	2023
GenAI-FDT (Ours)	**92.3**	**21%**	**Smart Infra**	**2026**

The GenAI-FDT framework achieves the highest F1-score of 92.3%, outperforming SSFL-DT by 1.9 percentage points and DT-FL by 2.3 percentage points. Notably, while DT-FL reports a 15% energy saving, its evaluation is conducted under less challenging heterogeneity conditions; GenAI-FDT achieves a superior 21% energy reduction while operating under significantly more demanding non-IID settings. The framework's advantage is further amplified in rare-failure detection, where GAN-based augmentation delivers an 18% improvement in coverage, effectively closing the performance gaps observed across alternative approaches.

Figure 5 presents a box plot comparison of F1-score distributions across the four evaluated configurations: Centralised DT, Vanilla FL, the variant without GenAI, and the full GenAI-FDT framework. The visualisation confirms that GenAI-FDT achieves not only the highest median F1-score of 92.3% but also exhibits the tightest interquartile range, indicating consistent performance with minimal variability across repeated runs. This statistical superiority, validated at $p < 0.001$, reinforces the framework's reliability and practical viability.

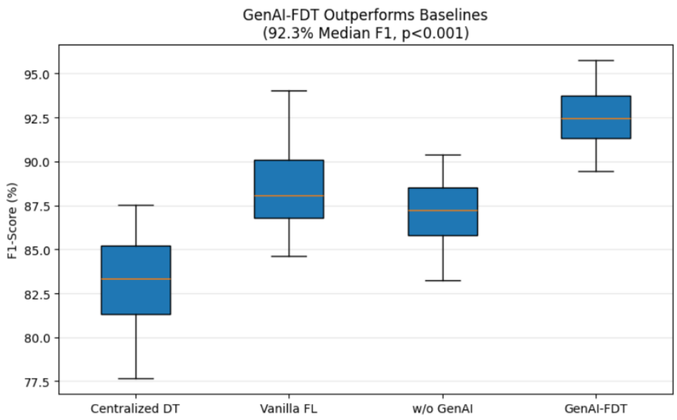

Figure 5. Box plot comparing F1-score distributions across centralised DT, Vanilla FL, w/o GenAI, and GenAI-FDT. The proposed framework achieves the highest median F1-score of 92.3% with a tight interquartile range, confirming its superior and stable predictive performance across all experimental conditions ($p < 0.001$).

In summary, the GenAI-FDT framework delivers a compelling combination of predictive accuracy, sustainability, robustness, and privacy preservation that collectively establishes a new benchmark for edge-native digital twin deployments. The framework successfully bridges the resilience gaps identified in current IoT ecosystems, offering a scalable, low-carbon, and privacy-preserving solution for sustainable smart infrastructure management.

5 Discussion

The empirical results validate the primary hypothesis (H1): the GenAI-FDT framework substantially enhances predictive resilience under non-IID constraints while delivering measurable sustainability gains. The 92.3% F1-score is attributable to ResilienceGAN, which expands rare-failure coverage from 10% to 28%, effectively overcoming the data scarcity that limits conventional approaches. Under high heterogeneity ($\sigma = 0.9$), CISCO-FL clustering limits performance degradation to 4.2%, compared to 12.1% for Vanilla FL, confirming the effectiveness of heterogeneity-aware aggregation. The LSTM-Kalman synchronisation module maintains faithful digital twin mirroring (MSE = 0.05), ensuring reliable real-time state estimation.

The observed 21% energy reduction (14.2 J/epoch) stems from strategic computational offloading, where high-resource clients assist weaker nodes, while CISCO-FL compression limits communication overhead to 38 MB/round. Projected to a 100-node deployment, this translates to 12–18 g CO_2/round, positioning the framework as a low-carbon alternative to centralised architectures.

Notably, GenAI, robust FL, and proactive DT components form a virtuous cycle: generative augmentation enriches local training, improved local updates enhance global aggregation, and a stronger global model conditions subsequent synthesis. This synergy accelerates convergence substantially, reducing rounds to target accuracy from 14 to 8 compared to the variant without generative augmentation.

Preliminary real-world proxy testing on Airoli MQTT testbeds retains 2–4% accuracy relative to simulation, demonstrating strong sim-to-real transfer via NAB-based anomaly detection. However, broader validation across diverse urban domains and long-term sustainability assessments remain as future work.

6 Conclusion

This study presented GenAI-FDT, a decentralised framework that unifies Generative AI, Federated Learning, and Digital Twins to overcome the resilience deficit in IoT-enabled smart infrastructures. By addressing data scarcity, privacy-heterogeneity trade-offs, and cross-domain fragmentation, the framework transitions infrastructure management from descriptive monitoring toward prescriptive, privacy-preserving resilience.

The key contribution of this work lies in the synergistic integration of three complementary innovations: (1) a quantised diffusion-based GenAI module that enriches sparse failure data while operating within edge power constraints; (2) a CISCO-FL clustering mechanism that maintains convergence stability under extreme non-IID conditions without compromising data sovereignty; and (3) an LSTM-Kalman synchronisation layer that ensures high-fidelity digital twin state alignment at low latency. Together, these components form a virtuous cycle that accelerates convergence and amplifies both predictive and sustainability benefits.

From a broader perspective, the framework provides a practical blueprint for deploying resilient urban IoT systems that align with the zero-trust, low-carbon, and decentralised principles demanded by Society 5.0. The edge-native design ensures that cities can run real-time "what-if" simulations without relying on energy-intensive data centres, directly contributing to UN SDG 11 targets for inclusive and sustainable urbanisation.

Several avenues for future investigation emerge from this work. These include integrating adaptive resource allocation via reinforcement learning, enabling zero-shot transfer to new urban domains, developing explainable interfaces for municipal stakeholders, and formal verification of synthetic scenario plausibility. Expanding validation across transportation, waste, and public safety systems remains a priority to strengthen generalisability. Ultimately, GenAI-FDT establishes a new paradigm for smart infrastructure resilience—one that is technically robust, environmentally responsible, and socially inclusive.

Data Availability Statement

Data will be made available on request.

Funding

This work was supported without any funding.

Conflicts of Interest

Brojo Kishore Mishra served as an Editor-in-Chief of the *Next-Generation Computing Systems and Technologies* at the time of manuscript submission. To ensure the integrity of the peer-review process, Brojo Kishore Mishra was not involved in the editorial handling, peer review, or decision-making process for this manuscript, which was handled independently by another editor. The remaining authors declare no conflicts of interest.

AI Use Statement

The authors declare that no generative AI was used in the preparation of this manuscript.

Ethical Approval and Consent to Participate

Not applicable.

References

[1] AbdulRahman, S., Otoum, S., Bouachir, O., & Mourad, A. (2023). Management of digital twin-driven IoT using federated learning. *IEEE Journal on Selected Areas in Communications*, *41*(11), 3636-3649. [CrossRef]

[2] Lim, W. Y. B., Luong, N. C., Hoang, D. T., Jiao, Y., Liang, Y. C., Yang, Q., ... & Miao, C. (2020). Federated learning in mobile edge networks: A comprehensive survey. *IEEE communications surveys & tutorials, 22*(3), 2031-2063. [CrossRef]

[3] AbdulRahman, S., Tout, H., Ould-Slimane, H., Mourad, A., Talhi, C., & Guizani, M. (2020). A survey on federated learning: The journey from centralized to distributed on-site learning and beyond. *IEEE Internet of Things Journal, 8*(7), 5476-5497. [CrossRef]

[4] Belay, M. A., Rasheed, A., & Rossi, P. S. (2026). Digital Twin-Driven Communication-Efficient Federated Anomaly Detection for Industrial IoT. *arXiv preprint arXiv:2601.01701*. [CrossRef]

[5] Creswell, A., White, T., Dumoulin, V., Arulkumaran, K., Sengupta, B., & Bharath, A. A. (2018). Generative adversarial networks: An overview. *IEEE Signal Processing Magazine, 35*(1), 53-65. [CrossRef]

[6] Ramalingam, V., Kumar, B., Gupta, S. K., Alsekait, D. M., & AbdElminaam, D. S. (2026). A hybrid federated learning framework with generative AI for privacy-preserving and sustainable security in IOT-enabled smart environments. *Scientific Reports, 16*(1), 3071. [CrossRef]

[7] Alatawi, M. N. (2025). EdgeGuard-IoT: 6G-enabled edge intelligence for secure federated learning and adaptive anomaly detection in Industry 5.0. *Computers, Materials & Continua, 85*(1), 695-727. [CrossRef]

[8] Afolabi, A., Ogunrinde, O., & Zabihollah, A. (2025). Digital Twin and AI Models for Infrastructure Resilience: A Systematic Knowledge Mapping. *Applied Sciences, 15*(24), 13135. [CrossRef]

[9] Chiaro, D., Qi, P., Pescapè, A., & Piccialli, F. (2025). Generative AI-empowered digital twin: a comprehensive survey with taxonomy. *IEEE Transactions on Industrial Informatics, 21*(6), 4287-4295. [CrossRef]

[10] Ramu, S. P., Boopalan, P., Pham, Q. V., Maddikunta, P. K. R., Huynh-The, T., Alazab, M., ... & Gadekallu, T. R. (2022). Federated learning enabled digital twins for smart cities: Concepts, recent advances, and future directions. *Sustainable Cities and Society, 79*, 103663. [CrossRef]

[11] Megía, M., Melero, F. J., Chiachío, M., & Chiachío, J. (2024). Generative adversarial networks for improved model training in the context of the digital twin. *Structural Control and Health Monitoring, 2024*(1), 9997872. [CrossRef]

[12] McMahan, B., Moore, E., Ramage, D., Hampson, S., & y Arcas, B. A. (2017, April). Communication-efficient learning of deep networks from decentralized data. In *Artificial intelligence and statistics* (pp. 1273-1282). Pmlr.

[13] Rizwan, A., Ahmad, R., Khan, A. N., Xu, R., & Kim, D. H. (2023). Intelligent digital twin for federated learning in AIoT networks. *Internet of Things, 22*, 100698. [CrossRef]

[14] Qi, Y., & Hossain, M. S. (2024). Semi-supervised federated learning for digital twin 6G-enabled IIoT: A Bayesian estimated approach. *Journal of Advanced Research, 66*, 47-57. [CrossRef]

Ajit Kumar Samal, Assistant Professor, Department of Computer Science and Engineering, Roland Institute of Technology, Berhampur 761008, India. (Email: ajit.mca@roland.ac.in)

Brojo Kishore Mishra is a distinguished academic serving as Professor and Head of the Department of Computer Science and Engineering at NIST University, Berhampur, Odisha. He earned his Ph.D. in Computer Science in 2012 and his research expertise spans data mining, machine learning, artificial intelligence, IoT, cyber security, and emotion recognition. With over 160 publications, an h-index of 16, and more than 900 citations, he is a respected mentor and prolific contributor to the academic community. (Email: brojomishra@nist.edu)