Climate Vulnerability Assessment of Infrastructure Using Edge-AI Integrated IoT Systems: A Survey

[1]Raj Mehta, [2]Dr. Varsha Shah

[1] Independent Researcher, Arizona State University, Columbia, SC, USA
[2] Principal, University of Mumbai, Mumbai, India

Abstract: Climate change is amplifying extremes that directly threaten critical infrastructure. Timely, spatially resolved vulnerability assessment is indispensable for adaptation planning and operational resilience. Cloud-first analytics alone struggle with bandwidth, latency, privacy, and continuity constraints in fast-evolving hazards. This survey synthesizes advances at the intersection of climate vulnerability assessment, internet-of-things (IoT) sensing, and edge artificial intelligence (edge-AI). We ground the assessment problem in contemporary climate risk evidence and definitions, propose an end-to-end framework linking hazard–exposure–vulnerability constructs to IoT/edge data flows, and review methods spanning sensing architectures, communication standards, on-device learning (TinyML, model compression, federated learning), spatio-temporal learning over sensor networks, and digital-twin integration. Representative deployments in flood monitoring, structural health monitoring, and wildfire detection illustrate how edge-AI reduces detection latency, preserves operation under degraded connectivity, and improves data stewardship—capabilities aligned with the needs of climate adaptation and risk-informed asset management [1]–[3], [9], [10].

IndexTerms **- climate risk, infrastructure resilience, IoT, Edge-AI, TinyML, federated learning, digital twins, spatio-temporal learning.**

I. INTRODUCTION

This survey focuses on edge-resident analytics for climate-hazard vulnerability of physical infrastructure (transport, water, power) where inference occurs on devices or gateways with intermittent backhaul. We include IoT sensing, messaging/data standards, embedded/near-edge ML, and digital-twin integration insofar as the edge is required for latency, privacy, or continuity. We exclude purely cloud-centric pipelines, satellite-only products without edge handshake, and grid-scale market operations beyond sensing/nowcasting. Selection prioritizes peer-reviewed deployments (2019–2025) and standards with active adoption.

Anthropogenic warming has reached approximately 1.1 °C for the 2011–2020 decade relative to 1850–1900 baselines, increasing the likelihood and severity of compound and cascading hazards. Infrastructure systems face escalating risks at every increment of warming [1]. In the United States, the Fifth National Climate Assessment documents rising heavy precipitation, heat extremes, wildfire weather, and coastal flooding—stressors that degrade transport, energy, water, and communications systems and amplify cascading disruptions [2]. Globally, a high-resolution multi-hazard analysis estimates expected annual direct damages to road and rail infrastructure on the order of US$14.6 billion, with roughly 73% attributable to surface and river flooding [3]. These figures underscore an operational imperative: vulnerability assessment that is both fast and situationally aware during hazard evolution.
Traditional cloud-centric analytics are powerful but often misaligned with urgency, bandwidth, privacy, and continuity requirements encountered during hazards. Edge computing and edge-AI address these gaps by executing inference close to data sources, thereby reducing end-to-end latency and dependence on wide-area backhaul while enabling context-preserving data minimization [4], [5]. In climate-exposed infrastructure, where degraded connectivity is common during storms or wildfires, the ability to reason at the edge without constant cloud access is not merely an optimization but a resilience prerequisite.

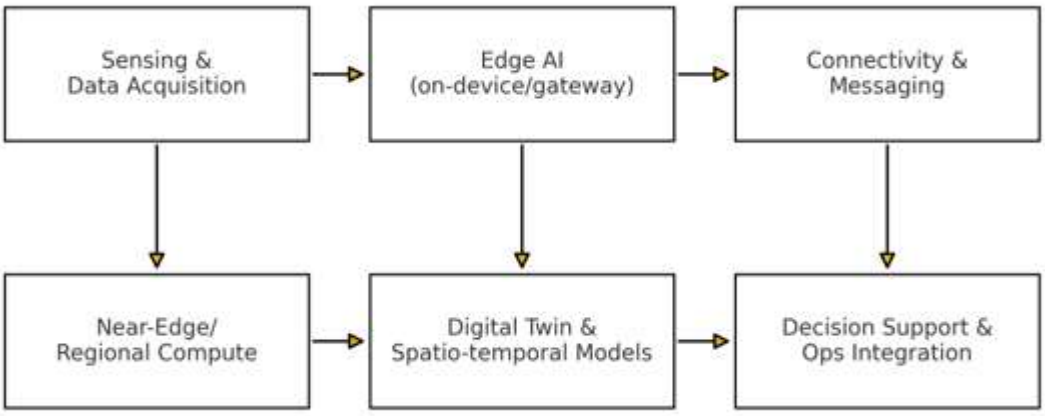

Figure 1. Layered edge–cloud architecture for climate vulnerability assessment using IoT and Edge-AI.

II. CONCEPTS AND DEFINITIONS FOR CLIMATE VULNERABILITY

NEED OF THE STUDY.

We adopt the IPCC framing in which risk emerges from the interaction of hazards (physical climate events or trends), exposure (assets and populations in harm's way), and vulnerability (susceptibility and capacity to cope and adapt). Vulnerability integrates sensitivity of infrastructure to hazard intensities and the adaptive capacity embodied in design standards, maintenance regimes, redundancies, and institutional supports [1], [2]. A climate vulnerability assessment for infrastructure therefore requires reliable

hazard nowcasts/forecasts and impact proxies, granular exposure data (asset locations, types, and condition), fragility or performance functions mapping intensities to damage or service loss, and decision-oriented outputs (alerts, risk scores, and prioritization) under uncertainty.

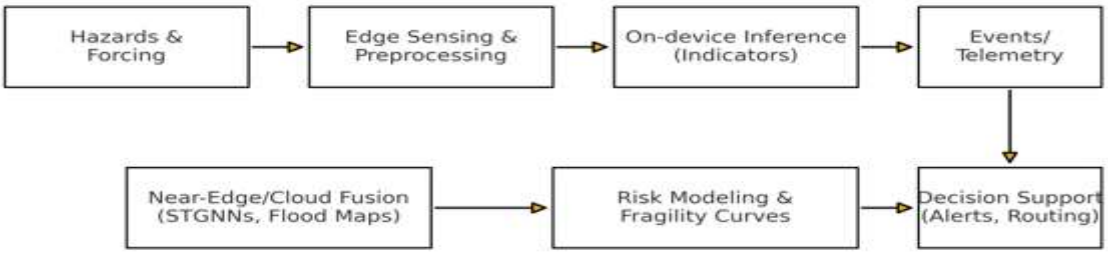

Figure 2. Data flow from hazard sensing to risk-calibrated decision outputs

Takeaway: Calibrated risk scores (with explicit uncertainty) reduce false reassurance under novel conditions and improve decision thresholds.

III. System Architecture: From Sensing to Decision

Vulnerability assessment draws on multi-modal signals: hydrometeorological measurements (stage, flow, rainfall), thermal and humidity conditions, wind and pressure, soil moisture and suction, vibration/strain/acoustic emissions for structural health, imaging for smoke or debris, and power-quality indicators for grid stress. To normalize heterogeneous observations, the OGC SensorThings API provides an open, geospatially enabled, resource-constrained standard for sensor metadata and time-series exchange, mapping observations to well-defined entities and enabling interoperable queries [7].

Connectivity options must align with terrain, density, and power envelopes. MQTT v5 is a lightweight publish/subscribe protocol widely used for telemetered sensor networks; its small header overhead, topic-based routing, and retained messages are appropriate for bandwidth-sensitive or intermittently connected devices [8]. LPWANs such as LoRaWAN and 3GPP NB-IoT deliver low-bit-rate, long-range links for battery-powered nodes, whereas 5G/URLLC and time-sensitive networking support deterministic latencies for low-delay workloads. Standards-led messaging and data models serve two goals: interoperability across heterogeneous asset owners and auditability of vulnerability indicators over time.

Edge endpoints span microcontrollers with kilobytes of SRAM (TinyML class) to micro-servers with dedicated NPUs/GPUs. Compression, quantization, distillation, and neural architecture search produce models that fit constrained envelopes while preserving task accuracy [5]. For near-edge aggregation, gateways coordinate local fusion, run heavier spatio-temporal models, and maintain store-and-forward queues for resilience.

Table 1 summarizes sensor modalities and edge-computable indicators relevant to climate vulnerability assessment.

Sensor Modality	Infrastructure Use	Edge-Computable Feature	Vulnerability/Impact Proxy	Example Citation
Pressure/Ultrasonic Stage	Culverts, rivers, urban drainage	Rate-of-rise; threshold exceedance	Rapid inundation and overtopping risk	[13], [14]
Accelerometers/Strain	Bridges, buildings, towers	Modal features; anomaly scores	Damage progression and serviceability loss	[11]
RGB/IR Cameras	Wildland-urban interface, rights-of-way	Smoke/flame detection; debris detection	Ignition/obstruction early warning	[12]
Thermo-Hygrometers	Railways, pavements, substations	Heat-index and dew-point trends	Thermal stress and buckling risk	[2]
Soil Moisture/Tensiometers	Slopes, embankments, pipelines	Saturation trends; change points	Shallow landslide susceptibility	[2]

Table 2 highlights messaging and data standards and their roles in edge-to-cloud operations.

Standard/Protocol	Role in the Stack	Why It Matters	Typical Placement
OGC SensorThings (2016)	Data model & API	Interoperable geospatial time-series exchange	Device, gateway, and data hubs [7]
MQTT v5 (2019)	Messaging (pub/sub)	Lightweight, topic-based, retained messages	Device and gateway brokers [8]
LoRaWAN	LPWAN access	Low power, long range for battery nodes	Device to gateway backhaul [13]
NB-IoT	Cellular LPWAN	Managed spectrum and coverage	Device to cellular core
5G/URLLC	Low-latency access	Deterministic latency for critical links	Field backhaul and campus networks

IV. EDGE-AI METHODS FOR VULNERABILITY INDICATORS

Calibration and Uncertainty at the Edge

Edge classifiers and regressors SHOULD be calibrated (e.g., temperature scaling) and paired with lightweight uncertainty estimates (e.g., conformal prediction) to avoid over-confident false reassurance under domain shift. Emphasize false-negative risk and calibrated thresholds in evaluation.

At the sensor or gateway, lightweight classifiers and regressors transform raw signals into hazard-relevant indicators—fast stage-rise detection from stage sensors, thermal anomaly scores for heat-sensitive assets, or vibration-based swing and tilt changes for bridges. Edge-adapted convolutional and temporal models—via pruning and integer quantization—operate within tight latency and energy envelopes on embedded SoCs, while TinyML models enable milliwatt-scale inference on microcontrollers [5]. Federated or transfer-learning protocols support local adaptation without centralizing sensitive data [6].

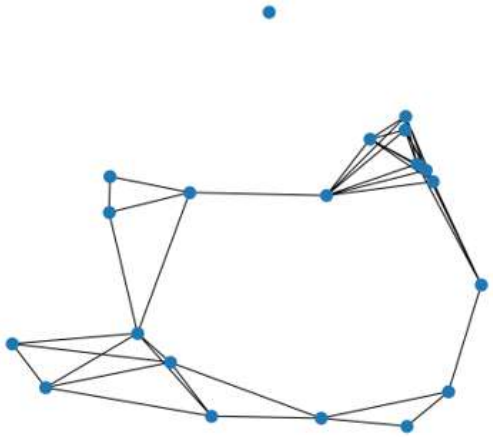

Figure 3. Conceptual sensor network graph underpinning spatio-temporal learning and propagation of vulnerability signals.

Takeaway: STGNNs exploit physical/structural coupling for propagation, but sparse or incorrect topology can harm predictions.

Complex interdependencies among infrastructure assets and sensors motivate spatio-temporal graph neural networks (STGNNs), which respect network topology and temporal dynamics. Such models have been widely applied to traffic and environmental sensor networks, and the same formulations extend to flood-routing sensors and distributed structural instrumentation to produce nowcasts or to condition fragility curves [15], [16]. For vision-based tasks, compact detectors based on MobileNet or the YOLO family demonstrate real-time performance on embedded platforms; quantized implementations on Jetson-class hardware have reported tens of frames per second for smoke/flame detection while maintaining acceptable accuracy [12].

V. DIGITAL TWINS AND DECISION SUPPORT

Twin Integration Notes

Co-simulation vs. surrogate: physics co-simulation offers fidelity but higher latency; surrogate models at the edge/gateway enable fast 'what-if' screening. Handoff latency from edge event → twin analysis SHOULD be budgeted explicitly in the design.

Digital twins synthesize live streams with physics-based or data-driven models to track infrastructure state and project consequences of hazard evolution and interventions. For flood-prone corridors, a twin can ingest stage sensors, radar rainfall, and image-based debris detection to produce segment-level risk trajectories and diversion recommendations. Edge inference operates as the data and event curation layer, while the twin hosts heavier scenario runs and what-if analyses. In urban resilience assessments, these capabilities enable real-time data integration and interactive decision support, yet governance, validation, and model fidelity remain central challenges to avoid maladaptation [2].

VI. SECURITY, PRIVACY, AND GOVERNANCE

Edge-AI infrastructures must conform to AI-specific governance and IoT device security baselines. The NIST AI Risk Management Framework (AI RMF 1.0) articulates outcomes for Govern, Map, Measure, and Manage, along with trustworthiness characteristics—validity, reliability, safety, security/resilience, explainability, privacy, and fairness—that should be operationalized across the lifecycle [9]. For device-level safeguards, NIST SP 800-213A catalogs IoT cybersecurity requirements spanning identification, configuration, data protection, interface access, software update, state awareness, and secure execution—capabilities directly relevant to safety-critical sensing and control [10]. Federated and transfer learning reduce central aggregation of sensitive data but must be hardened against poisoning and backdoor attacks when deployed for field adaptation [6].

VII. REPRESENTATIVE APPLICATION DOMAINS

Domain-Specific Field Constraints and Targets

• Flood nodes: payload ≤ 12 bytes/event; radio duty-cycle ≤ 1%; enclosure ≥ IP67; update 1–5 min; examples include LoRaWAN deployments and urban street-level networks [17], [20].

• Structural health monitoring (SHM): MCU-class TinyML viable for selected tasks; target energy/inference ≤ 20 mJ; on-device redaction for camera feeds; see TinyML SHM exemplar [19].

• Wildfire smoke/flame: ≥ 15 FPS on embedded GPU; alerting tuned for low false-negatives; edge-cloud co-training under drift.

Flood monitoring and rapid inundation assessment benefit from distributed hydro-IoT nodes that compute edge-level exceedances and rise-rate alerts, then uplink compact events via LPWAN. LoRa-based deployments demonstrate low-power, long-range feasibility for real-world flood monitoring, offering modular sensor interfacing and sustained operation on constrained energy budgets [13]. For map-level vulnerability, near-real-time SAR-based flood mapping (e.g., Sentinel-1) complements in-situ signals; deep segmentation models trained on public datasets have produced rapid inundation products that condition network-level exposure

and fragility analyses [14]. Integrating these streams—local thresholds at the edge and regional maps at the near-edge/cloud—yields robust nowcasts for transport-asset risk and routing decisions under uncertainty [3], [14].

Structural health monitoring increasingly uses edge computing to pre-process vibration and strain signals, extract features, and run compact classifiers that flag anomalies or condition changes, thereby reducing uplink bandwidth and enabling continuous monitoring. Reviews document the migration of deep models from cloud to edge for tasks including crack detection, modal identification, and damage localization, with compression and quantization enabling deployment on constrained processors [11].

Wildfire early warning camera networks augmented with edge-deployed smoke and flame detectors can flag incipient events faster than human scanning when visibility permits. Embedded implementations report real-time operation on field-rated system-on-modules after backbone replacement with MobileNet-class encoders and quantized detectors, achieving approximately 26 frames per second on Xavier NX in controlled tests [12]. These systems illustrate how edge-AI can deliver low-latency alerts when backhaul is congested by concurrent emergencies.

Table 3 summarizes representative deployments and their edge-AI roles.

Domain	Representative Study	Edge-AI Function	Platform/Connectivity	Indicative Outcome
Flood monitoring	[13] Ragnoli et al., 2020	On-sensor thresholds; event compression	Battery nodes + LoRaWAN	Sustained low-power operation
Rapid flood mapping	[14] Katiyar et al., 2021	Deep segmentation (SAR) at near-edge/cloud	Compute node + open data	Near-real-time inundation maps
Wildfire detection	[12] Zheng et al., 2023	Quantized smoke/flame detection	Jetson-class embedded SOC	≈26 FPS on embedded hardware

VIII. Evaluation Metrics and Benchmarks

Evaluation & Reproducibility Scaffold

Report a minimal device-to-decision card: latency budget (sensor→decision, ms); energy per inference (mJ) and 24 h duty-cycle sustainability; uplink payload (bytes/event, events/day); model size (KiB/MB) and peak RAM (KiB); task metrics (AUROC/F1) and decision-centric loss (missed-hazard reduction at fixed alert budget); packet-loss tolerance at p%; and thermal stability (W @ °C).

Evaluation must account for end-to-end latency from sensor to decision, energy per inference or duty-cycle cost for battery sustainability, communication footprint, robustness to drift, and task accuracy with decision-relevant false-negative trade-offs. For vulnerability assessment, evaluation should additionally report risk-relevant metrics such as reduction in missed hazardous states at a given alert burden, alignment with fragility curves, and value-of-information for routing or maintenance decisions. Where possible, metrics should be stratified by hazard intensities and asset typologies relevant to the target network, consistent with the AI RMF's emphasis on context-appropriate risk measurement [9].

IX. Open Challenges and Research Agenda

Limitations

Field deployments face concept drift, sensor spoofing/poisoning risks during federated adaptation, and sparse multi-sensor datasets with energy/latency annotations. We therefore mark claims where evidence is thin and encourage open benchmarks with device-level energy traces.

Non-stationarity and domain shift pose persistent challenges as climate drivers and asset conditions evolve. Reliable online adaptation and uncertainty quantification at the edge remain active research areas, especially under strict compute and energy budgets. Turning edge-curated signals into interpretable vulnerability narratives requires hybrid models and interpretable surrogates for operator trust. Public, labeled, multi-sensor datasets with energy and latency annotations and risk-calibrated scoring remain sparse for climate-hazard tasks; community benchmarks should incorporate device-level measurements and decision-centric losses. Federated TinyML and unattended devices are vulnerable to model poisoning, sensor spoofing, and supply-chain risks; incorporating SP 800-213A capabilities and AI-RMF-conformant governance is essential to avoid brittle or unsafe behavior in critical settings [9], [10]. Finally, operational integration into digital-twin-enabled workflows and asset-management systems requires tested interfaces, versioned models, and assurances around testing, evaluation, verification, and validation (TEVV).

X. Conclusion

The intensification of climate hazards demands vulnerability assessment that keeps pace with events and operates under degraded conditions. Edge-AI integrated IoT systems—grounded in open data models and messaging standards, equipped with on-device detection and adaptive learning, and governed by rigorous security and risk-management frameworks—offer a pragmatic path from raw signals to actionable, risk-calibrated indicators. Evidence from flood monitoring, structural health monitoring, and wildfire detection demonstrates that such systems can achieve real-time or near-real-time performance while containing bandwidth and preserving data stewardship [3], [11]– [13]. Future work should prioritize non-stationary learning, risk-aware evaluation, and trustworthy integration into digital twins and decision processes.

REFERENCES

[1] IPCC, Climate Change 2023: Synthesis Report. Intergovernmental Panel on Climate Change, 2023.
[2] A. R. Crimmins, C. W. Avery, D. R. Easterling, et al., Fifth National Climate Assessment. U.S. Global Change Research Program, 2023. doi: 10.7930/NCA5.2023.
[3] E. E. Koks, M. Thober, S. W. Bierkens, et al., "A global multi-hazard risk analysis of road and railway infrastructure assets," Nature Communications, 10, 2677, 2019. doi: 10.1038/s41467-019-10442-3.
[4] Y. Chen and X. Ran, "Deep Learning With Edge Computing: A Review," Proceedings of the IEEE, 107(8), 1655–1674, 2019.
[5] V. Tsoukas, A. Gkogkidis, E. Boumpa, and A. Kakarountas, "A Review on the Emerging Technology of TinyML," ACM Computing Surveys, 56(10), 2024. doi: 10.1145/3661820.
[6] M. Ficco, F. Palmieri, A. Castiglione, et al., "Federated learning for IoT devices: Enhancing TinyML with transfer learning," Information Fusion, 99, 102189, 2024. doi: 10.1016/j.inffus.2023.102189.
[7] OGC, SensorThings API Part 1: Sensing, OGC 15-078r6, Open Geospatial Consortium, 2016.
[8] OASIS, MQTT Version 5.0, OASIS Standard, 2019.
[9] NIST, Artificial Intelligence Risk Management Framework (AI RMF 1.0), NIST AI 100-1, 2023. doi: 10.6028/NIST.AI.100-1.
[10] NIST, IoT Device Cybersecurity Guidance for the Federal Government: IoT Device Cybersecurity Requirement Catalog, SP 800-213A, 2021. doi: 10.6028/NIST.SP.800-213A.
[11] Y. Lin, "Edge Computing for Structural Health Monitoring: A Review," Sensors, 22, 2022. doi: 10.3390/s22218231.
[12] H. Zheng, J. Duan, Y. Dong, and Y. Liu, "Real-time fire detection algorithms running on small embedded devices based on MobileNetV3 and YOLOv4," Fire Ecology, 19, 31, 2023. doi: 10.1186/s42408-023-00189-0.
[13] M. Ragnoli, G. Barile, A. Leoni, G. Ferri, and V. Stornelli, "An Autonomous Low-Power LoRa-Based Flood-Monitoring System," Journal of Low Power Electronics and Applications, 10(2), 15, 2020. doi: 10.3390/jlpea10020015.
[14] V. Katiyar, J. S. S. Jensen, N. Kussul, et al., "Near-Real-Time Flood Mapping Using Off-the-Shelf Synthetic Aperture Radar Data and Deep Learning," Remote Sensing, 13(12), 2334, 2021. doi: 10.3390/rs13122334.
[15], [16] X. Zheng, Y. Zhao, D. Wu, et al., "A Comprehensive Survey on Spatio-Temporal Graph Neural Networks," arXiv preprint arXiv:2312.04806, 2024.
[16] G. Jin, Y. Liang, Y. Fang, Z. Shao, J. Huang, J. Zhang, and Y. Zheng, "Spatio-Temporal Graph Neural Networks for Predictive Learning in Urban Computing: A Survey," IEEE Transactions on Knowledge and Data Engineering, vol. 36, no. 10, pp. 5388–5408, 2024. doi: 10.1109/TKDE.2023.3333824.
[17] M. I. Zakaria, W. A. Jabbar, and N. Sulaiman, "Development of a Smart Sensing Unit for LoRaWAN-Based IoT Flood Monitoring and Warning System in Catchment Areas," Internet of Things and Cyber-Physical Systems, vol. 3, pp. 249–261, 2023. doi: 10.1016/j.iotcps.2023.04.005.
[18] P. Roosipuu, I. Annus, A. Kuusik, N. Kändler, M. M. Alam, et al., "Monitoring and Control of Smart Urban Drainage Systems Using NB-IoT Cellular Sensor Networks," Water Science & Technology, vol. 88, no. 2, pp. 339–354, 2023.
[19] C. Huang, H. Wei, Y. Zhang, and S. Pan, "Tiny-Machine-Learning-Based Supply Canal Surface Crack Detection on Microcontroller Units," Sensors, vol. 24, no. 13, p. 4124, 2024. doi: 10.3390/s24134124.
[20] C. Mydlarz, P. S. V. Challagonda, B. Steers, J. Rucker, T. Brain, B. Branco, et al., "FloodNet: Low-Cost Ultrasonic Sensors for Real-Time Measurement of Hyperlocal, Street-Level Floods in New York City," Water Resources Research, vol. 60, no. 5, e2023WR036806, 2024. doi: 10.1029/2023WR036806.