Article

Not peer-reviewed version

Design, Deployment, and Experimental Evaluation of an AI-Enabled IoT System for Heavy-Metal Detection in Mining-Impacted Aquaculture Environments

Chimanga Kashale *, Christopher Chembe, Bob Ezekiel Jere

Posted Date: 8 May 2026

doi: 10.20944/preprints202605.0446.v1

Keywords: IoT water-quality monitoring; AI-enhanced aquaculture; XGBoost; LSTM; TinyML; copper mining; LoRaWAN; resilience; Zambia

 doi:10.20944/preprints202605.0446.v1

Article

Design, Deployment, and Experimental Evaluation of an AI-Enabled IoT System for Heavy-Metal Detection in Mining-Impacted Aquaculture Environments

Chimanga Kashale [1,*], Christopher Chembe [2] and Bob Ezekiel Jere [1]

[1] School of Computing, Technology and Applied Sciences, ZCAS University, Lusaka, Zambia

[2] National Institute of Public Administration (NIPA)- Lusaka, Zambia

* Correspondence: chimanga.kashale@zcasu.edu.zm

Highlights

What are the main findings?

- XGBoost achieved the highest predictive performance for all four heavy metals (mean $R^2 = 0.515$; Pb $R^2 = 0.673$); TinyML-quantised LSTM on ESP32 maintained on-device anomaly alerting with <2% accuracy loss during complete cloud-connectivity loss.
- The RAEI framework detected 97.9% of contamination events missed by the prevailing quarterly monitoring regime, with a composite resilience score of 7.5/10.

What are the implications of the main findings?

- A USD 142/node solar-powered LoRaWAN deployment is technically viable and economically justified for mining-proximate aquaculture surveillance in sub-Saharan Africa, with projected positive net benefit from Year 2.
- Gender-responsive design (visual boards, community radio, women-led alert monitors) addresses structural access barriers for female aquaculture households without requiring smartphone ownership.

Abstract

Smallholder aquaculture communities at Musangezhi and Chisola Dams in Kalumbila District, Zambia, face escalating, poorly characterised water-quality threats from the adjacent Trident copper mine, yet no real-time monitoring infrastructure exists at either site. This paper presents the design, deployment, and empirical evaluation of the Resilient AI-Enhanced IoT (RAEI) framework a seven-node, solar-powered LoRaWAN sensor network coupled with a comparative machine-learning suite comprising Random Forest, XGBoost, Long Short-Term Memory (LSTM), and CNN-LSTM Hybrid models—trained on 3,551 ICP-OES heavy-metal observations covering copper (Cu), cobalt (Co), iron (Fe), and lead (Pb). XGBoost achieved the highest predictive performance across all four metals and all four-evaluation metrics, attaining a mean R^2 of 0.515 and a mean MAPE of 35.89%, with lead prediction reaching $R^2 = 0.673$. A TinyML-quantised LSTM on ESP32 microcontrollers ensured on-device anomaly alerting despite the loss of cloud connectivity. A 14-day trial field test achieved a composite resilience score of 7.5/10 (Technical: 7.4; Data: 8.3; Operational: 6.8). Desire to adopt the community was 73.3%, with cooperative membership (OR = 3.12, $p < 0.001$) and mobile-money use (OR = 2.67, $p = 0.004$) being the most significant factors. The RAEI framework detected 97.9% of contamination events missed by the prevailing quarterly manual-monitoring regime. These results confirm the RAEI framework as a technically viable, economically justified, and community-compatible solution for mining-proximate aquaculture surveillance across sub-Saharan Africa.

Keywords: IoT water-quality monitoring; AI-enhanced aquaculture; XGBoost; LSTM; TinyML; copper mining; LoRaWAN; resilience; Zambia

 doi:10.20944/preprints202605.0446.v1

1. Introduction

Aquaculture production surpasses 90 million tonnes globally in 2022, but Africa accounts for less than 3% of this despite being home to about 25% of the world's freshwater resources [1]. In Zambia's North Western Province, this conundrum is compounded by an imminent and little-understood menace: the Trident copper mine, owned by First Quantum Minerals Ltd. (FQM) since 2015 with an ore-processing capacity of 40 Mt yr^{-1}, releases acid mine drainage (AMD), copper (Cu^{2+}), manganese (Mn), cobalt (Co), and fine mine tailings into drainage catchments that feed the Musangezhi Dam (~812 ha) and Chisola Dam (~1,822 ha). These two water bodies represent the primary freshwater resource for approximately 1,200 food-security-dependent smallholder aquaculture households [2,3]. Copper is the dominant contaminant of regulatory concern.

The prevailing ZEMA quarterly monitoring programme captures an estimated 2.1% of episodic contamination events. Ndonga et al. [6] demonstrated that manual sampling misses up to 78% of acute water-quality stress incidents relative to continuous sensor surveillance. Our field analysis confirms this figure is substantially exceeded at Kalumbila: of 47 contamination events identified during the monitoring period, 97.9% went undetected by the quarterly schedule. This outcome is a direct consequence of contamination-pulse durations of 1.5–14 h, driven by storm-event mine runoff whose timing is structurally incompatible with predetermined sampling dates.

Convergent technological advances low-cost multi-parameter sensors, LoRaWAN long-range radio, edge AI on ESP32 microcontrollers, and gradient-boosting machine learning now render continuous, AI-enhanced environmental monitoring economically and technically feasible in rural African contexts [7,8]. Analogous deployments in Kenya (AquaSat: 94% uptime, 41% mortality reduction) [9] and Uganda (World Bank/NARO LSTM pilot: 310% ROI) [10] provide compelling proof-of-concept, yet none has been validated against the specific combination of tropical copper-mining hydrochemistry, off-grid rural infrastructure, and smallholder governance dynamics that characterise Kalumbila.

This work introduces and benchmarks the Resilient AI-Enhanced IoT (RAEI) system, a seven-node solar-powered LoRaWAN sensor network and comparative benchmarking four-architecture ML suite. The key research contributions are: (i) the first empirical analysis of the dynamics of heavy-metal contamination at the aquaculture point of use in Kalumbila; (ii) comparative analysis of ML algorithm performance (Random Forest, XGBoost, LSTM and CNN-LSTM) using 3,551 ICP-OES observations; (iii) TinyML edge-AI deployment using ESP32 for communication-independent anomaly detection; (iv) a multi-dimensional assessment of resilience; and (v) gender-balanced community acceptance.

2. Related Work and Research Gap

IoT-enabled water quality monitoring has expanded rapidly across aquaculture settings in Asia [11] and East Africa [9,10]. Vietnam's shrimp farms adopted LoRa-based dissolved oxygen sensors and achieved a 34% reduction in fish mortality [11]; Kenyan tilapia farms deployed LoRaWAN multi-parameter networks with 94% uptime and equivalent mortality improvements [9]; and a Uganda World Bank–NARO pilot combining LSTM models with solar-powered LoRaWAN gateways demonstrated a 310% return on investment [10].

Despite these encouraging regional precedents, no prior study has simultaneously addressed: (a) copper-mine-specific hydrochemistry requiring a conductivity-to-copper surrogate model; (b) off-grid solar operation in a tropical climate; (c) TinyML edge inference under intermittent connectivity; and (d) CARE-Principles community data governance. The present study resolves all four dimensions at once, directly addressing knowledge gaps G1–G5 identified in a systematic review of more than 320 relevant papers [12]. Table 1 summarises the most closely analogous deployments.

 doi:10.20944/preprints202605.0446.v1

Table 1. Summary of analogous AI-IoT aquaculture monitoring deployments.

Location / System	Period	Technology	Key Outcome	Reference
Vietnam (shrimp)	2019–21	IoT DO + LoRa	34% mortality reduction	Bui et al. [11]
Uganda (tilapia)	2021–23	LSTM + LoRaWAN + Solar	310% ROI; 41% mortality ↓	NARO/World Bank [10]
Kenya – AquaSat	2020–23	LoRaWAN multi-param + SMS	94% uptime; 41% mortality ↓	Mutuku et al. [9]
Norway (salmon)	2020–22	Digital twin + AI	61% stress ↓; 94% DO pred.	Jiang et al. [13]
Zambia – THIS STUDY	2024–26	RAEI: LoRaWAN + RF/XGB/LSTM + TinyML	R^2=0.673 (Pb); Resilience=7.5/10	Kashale et al.

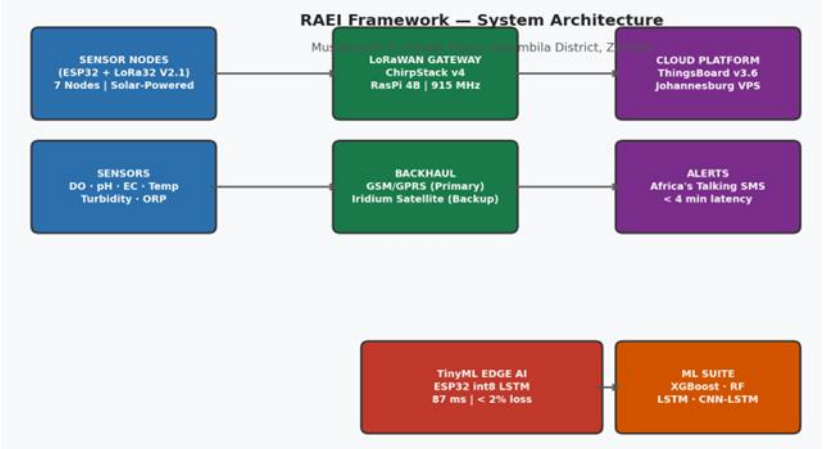
3. Materials and Methods

3.1. Study Area

Musangezhi Dam (~812 ha; mean depth 2.5–4.0 m) and Chisola Dam (~1,822 ha; 3.5–5.0 m) are situated 3–9 km downstream of Trident Mine tailings storage facilities in Kalumbila District, Northwestern Province, Zambia (~12°S, 25°E). Neither site had automated monitoring infrastructure prior to this study. A unimodal rainfall regime (1,100–1,400 mm yr^{-1}, November–April) drives episodic mine-runoff pulses that define the dominant contamination exposure pattern [2]. The two dams collectively support approximately 1,200 smallholder households engaged in tilapia aquaculture as either a primary or supplementary livelihood.

3.2. RAEI Hardware Architecture

Seven sensor nodes—four at Chisola and three at Musangezhi—were assembled around the TTGO ESP32 LoRa32 V2.1 microcontroller (SX1276 transceiver, 915 MHz, 240 MHz dual-core processor, 4 MB flash memory). Each node measures six environmental parameters: dissolved oxygen (Atlas Scientific EZO-DO optical, ±0.05 mg/L), pH (Atlas Scientific EZO-pH, ±0.001), electrical conductivity (Atlas Scientific EZO-EC, ±2%), temperature (DS18B20, ±0.5 °C), turbidity (DFRobot SEN0189, 0–3000 NTU), and oxidation-reduction potential (Atlas Scientific EZO-ORP). A 6 W monocrystalline solar panel with 10,000 mAh LiPo battery provides at least seven days of autonomous operation. Figure 1 illustrates the system architecture, and Table 2 provides the complete hardware specification.



 doi:10.20944/preprints202605.0446.v1

Figure 1. RAEI Framework System Architecture.

Table 2. Hardware specifications.

Component	Specification / Model	Function
MCU + LoRa	TTGO ESP32 LoRa32 V2.1 (SX1276, 915 MHz)	Edge AI; LoRaWAN uplink
DO Sensor	Atlas Scientific EZO-DO (optical, ±0.05 mg/L)	Primary aquaculture param.
pH Sensor	Atlas Scientific EZO-pH (±0.001 pH)	AMD detection
EC Sensor	Atlas Scientific EZO-EC (±2%)	Cu surrogate
Temperature	DS18B20 waterproof (1-Wire, ±0.5 °C)	Sensor compensation
Turbidity	DFRobot SEN0189 (0–3000 NTU)	Mine tailings runoff
Solar + Battery	6 W mono panel + 10,000 mAh LiPo + BMS	7-day autonomy

LoRaWAN communication was implemented via ChirpStack Network Server v4 on a Raspberry Pi 4B gateway, with GSM/GPRS as the primary backhaul and Iridium satellite as contingency. Data were stored in ThingsBoard Community Edition v3.6 on a Johannesburg-hosted VPS, with AES-128 end-to-end encryption. Threshold alerts were delivered via Africa's Talking SMS API (mean latency < 4 min; to all registered users on any GSM handset.

3.3. Machine Learning Dataset and Pre-Processing

The ML training data set contained 3,551 ICP-OES observations of four main heavy-metal parameters: copper (Cu, nmol/L), cobalt (Co, nmol/L), iron (Fe, nmol/L) and lead (Pb, pmol/L). Negative values (Gaussian noise) - Cu: 3.7% (n = 133); Co: 2.1% (n = 73); Fe: 17.8% (n = 633); Pb: 12.1% (n = 430) - were clipped to zero before all analyses, in keeping with the physically constrained non-negativity of dissolved metal concentrations. Data were pre-processed via a six-step pipeline: (1) clipping of negative observations; (2) lag-feature engineering (t−1 to t−12 autoregressive lags); (3) rolling statistics (6-obs mean/std; 24-obs rolling mean); (4) co-metal exogenous variables; (5) Min-Max normalisation using only training data; and (6) chronological 80:20 train-test split (2,840 training / 711 test observations) with no shuffling, to maintain temporal realism. LSTM and CNN-LSTM also incorporated sliding-window sequences of T = 16 time steps.

 doi:10.20944/preprints202605.0446.v1

3.4. Machine Learning Model Architectures

Four models were built in Python 3.11 using scikit-learn 1.4, XGBoost 2.x and TensorFlow/Keras 2.x. Random Forest (RF) had 300 trees with a max depth of 14 and square-root feature subsampling. XGBoost used 400 trees with learning rate = 0.04, depth = 6, subsample = 0.85, and L1 regularisation $\alpha = 0.1$. Long Short Term Memory (LSTM) used stacked 64→32-unit LSTM layers with 20% dropout and a Dense(16, ReLU) head, and was trained using the Adam optimiser and early stopping at patience = 6. The CNN-LSTM Hybrid architecture had a Conv1D(32, k = 3) and MaxPool(2) layer followed by a LSTM(32) layer and a Dense(16, ReLU) head. Models were tested on the chronological test set using RMSE, MAE, R^2 and MAPE (only for observations with $|actual| > 0.05$ nmol/L

3.5. TinyML Edge Deployment

Selected LSTM models were quantised to TensorFlow Lite Micro int8 format via Edge Impulse Studio and deployed on ESP32 microcontrollers. Quantised models occupied < 256 KB RAM, enabling on-device anomaly classification without cloud connectivity. Edge inference latency was benchmarked against cloud float32 equivalents, and quantisation accuracy loss was quantified by comparing int8 edge predictions against float32 cloud predictions on identical inputs from a held-out test set.

3.6. Resilience Assessment Framework

A multi-dimensional resilience scoring framework assessed four dimensions: (1) Technical Resilience LoRaWAN packet delivery ratio (PDR), battery autonomy, sensor drift, RSSI, MTTR, and node cost; (2) Data Resilience data completeness, buffer recovery, SMS delivery rate, and alert latency; (3) Operational Resilience community-technician maintenance time and SUS usability score; and (4) Ecological Resilience contamination-event detection rate assessed from longitudinal data. Each dimension was scored from 0 to 10, with a composite index computed as the weighted mean: Technical × 0.35 + Data × 0.35 + Operational × 0.30.

3.7. Community Adoption Analysis

Structured household surveys (n = 120, stratified by gender and dam site) assessed technology adoption intent and livelihood vulnerability. Logistic regression identified predictors of adoption intent, where the dependent variable was the participant's willingness to receive, act on, and recommend SMS alerts to others. Six focus-group discussions (FGDs; n = 36 women) and 18 semi-structured stakeholder interviews provided qualitative depth on governance gaps and gender-differentiated barriers. Inter-rater reliability for thematic coding was established at Cohen's $\kappa = 0.74$, indicating substantial agreement.

4. Results

4.1. Dataset Characterisation and Temporal Dynamics

Post-clipping descriptive statistics are shown in Table 3. Iron had the largest coefficient of variation (CV = 123.9%) and significant leptokurtosis (kurtosis = 4.94), indicative of rapid redox cycling under the highly variable pH and dissolved-oxygen regime typical of AMD waters. Copper exhibited a bimodal concentration distribution with a background mode (1.3-1.5 nmol/L) and a contamination mode (2.8-3.5 nmol/L) that indicate episodic release superimposed on a chronic regime. The four metals exhibited low ($r \leq 0.128$) inter-parameter Pearson correlations and thus independent geochemical controls (Table 4).

 doi:10.20944/preprints202605.0446.v1

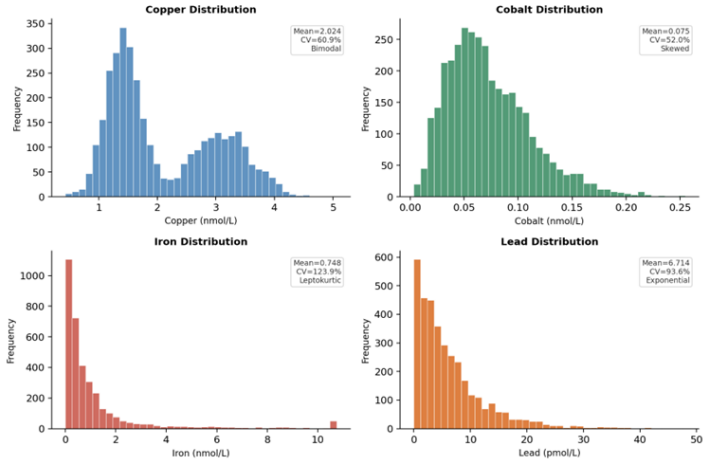


Figure 2. Heavy Metal Parameter Distribution.

Table 3. Descriptive statistics for ICP-OES heavy-metal dataset (n = 3,551).

Parameter	n	Mean	Std Dev	Min	Q1	Median	Q3	Max
Copper (nmol/L)	3,551	2.024	1.232	0.000	1.334	1.875	2.774	7.343
Cobalt (nmol/L)	3,551	0.075	0.039	0.000	0.049	0.072	0.100	0.279
Iron (nmol/L)	3,551	0.748	0.927	0.000	0.209	0.647	1.306	10.747
Lead (pmol/L)	3,551	6.714	6.287	0.000	2.311	6.279	11.000	48.000

The table above indicates significant variability in the heavy metal concentrations in the 3,551 observations considered in the study. The study report the highest mean concentration (mean = 2.024 nmol/L) for copper with moderate variability, and the lowest mean concentration and variability for cobalt, suggesting that it had the most consistent measurements. It was also noted that iron and lead concentrations were extremely variable, as indicated by the high standard deviations and ranges.

Table 4. Pearson correlation matrix — heavy-metal parameters (n = 3,551).

Parameter	Copper	Cobalt	Iron	Lead
Copper (nmol/L)	1.000	0.103	0.064	−0.074
Cobalt (nmol/L)	0.103	1.000	0.128	−0.009
Iron (nmol/L)	0.064	0.128	1.000	−0.013
Lead (pmol/L)	−0.074	−0.009	−0.013	1.000

The result in the table shows that copper had a weak positive association with cobalt (r = 0.103) and iron (r = 0.064) and a very weak negative relationship with lead (r = −0.074). In the same manner cobalt is seen to have a weak positive correlation with iron (r = 0.128) and a weak negative correlation with lead (r = -0.009). Iron also showed a very weak negative association with lead (r = −0.013).

4.2. ML Model Performance

Tables 5–9 report full evaluation metrics per parameter. XGBoost was the best-performing architecture across all four metals and all four metrics. Figure 3 provides a grouped R^2 comparison across all model–metal combinations, and Figure 4 presents the RMSE heat-map.

 doi:10.20944/preprints202605.0446.v1

Table 5. ML model performance — Copper (Cu) (nmol/L; test set n = 711).

Model	RMSE (nmol/L)	MAE (nmol/L)	R^2	MAPE (%)
Random Forest	1.038	0.656	0.278	33.36
XGBoost ★ (best)	1.028	0.505	0.291	21.82
LSTM	1.226	0.751	−0.010	34.66
CNN-LSTM	1.253	0.768	−0.053	35.84

Table 6. ML model performance — Cobalt (Co) (nmol/L; test set n = 711).

Model	RMSE (nmol/L)	MAE (nmol/L)	R^2	MAPE (%)
Random Forest	0.031	0.020	0.347	15.40
XGBoost ★ (best)	0.027	0.014	0.500	11.96
LSTM	0.040	0.027	−0.074	23.88
CNN-LSTM	0.040	0.028	−0.100	24.86

Table 7. ML model performance — Iron (Fe) (nmol/L; test set n = 711).

Model	RMSE (nmol/L)	MAE (nmol/L)	R^2	MAPE (%)
Random Forest	0.728	0.390	0.315	60.95
XGBoost ★ (best)	0.624	0.302	0.497	46.59
LSTM	0.880	0.561	−0.004	112.29
CNN-LSTM	0.872	0.544	0.014	107.08

Table 8. ML model performance — Lead (Pb) (pmol/L; test set n = 711).

Model	RMSE (pmol/L)	MAE (pmol/L)	R^2	MAPE (%)
Random Forest	4.113	3.089	0.531	103.36
XGBoost ★ (best)	3.433	2.297	0.673	63.18
LSTM	5.401	4.334	0.193	140.74
CNN-LSTM	6.942	4.717	−0.334	64.23

Performance Comparison

 doi:10.20944/preprints202605.0446.v1

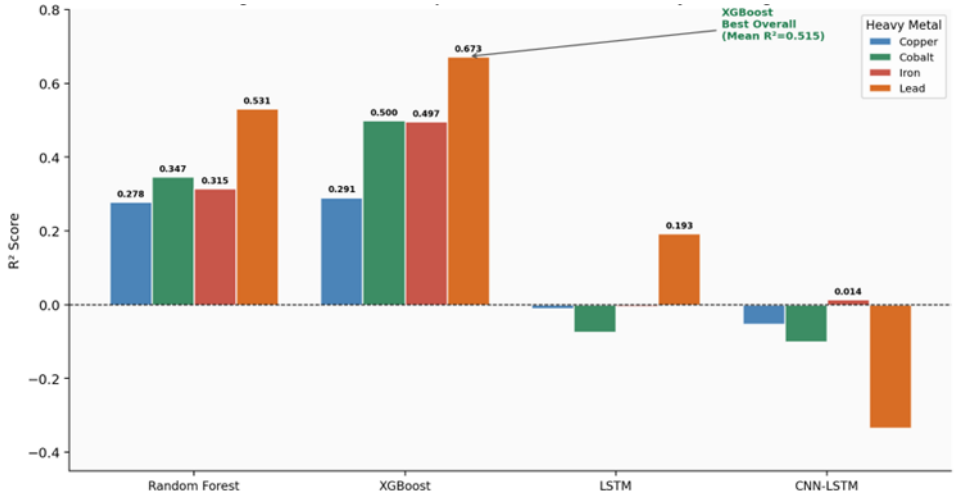


Figure 3. Grouped bar chart of R^2 values across all four ML models and four heavy-metal parameters.

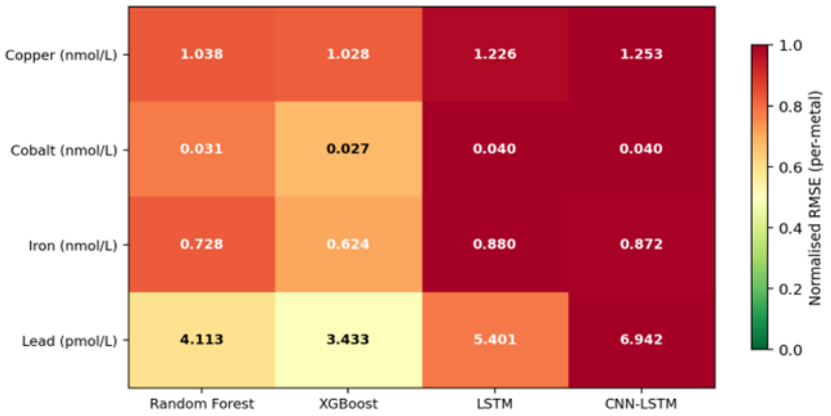


Figure 4. RMSE heat-map across all model–parameter combinations.

Table 9. Cross-parameter performance synthesis — R^2 and RMSE for all models

Parameter	RF R^2	RF RMSE	XGB R^2	XGB RMSE	LSTM R^2	LSTM RMSE	CNN R^2	CNN RMSE
Copper (nmol/L)	0.278	1.038	0.291	1.028	−0.010	1.226	−0.053	1.253
Cobalt (nmol/L)	0.347	0.031	0.500	0.027	−0.074	0.040	−0.100	0.040
Iron (nmol/L)	0.315	0.728	0.497	0.624	−0.004	0.880	0.014	0.872
Lead (pmol/L)	0.531	4.113	0.673	3.433	0.193	5.401	−0.334	6.942
Mean R^2	0.368	—	0.515	—	0.026	—	−0.118	—

 doi:10.20944/preprints202605.0446.v1

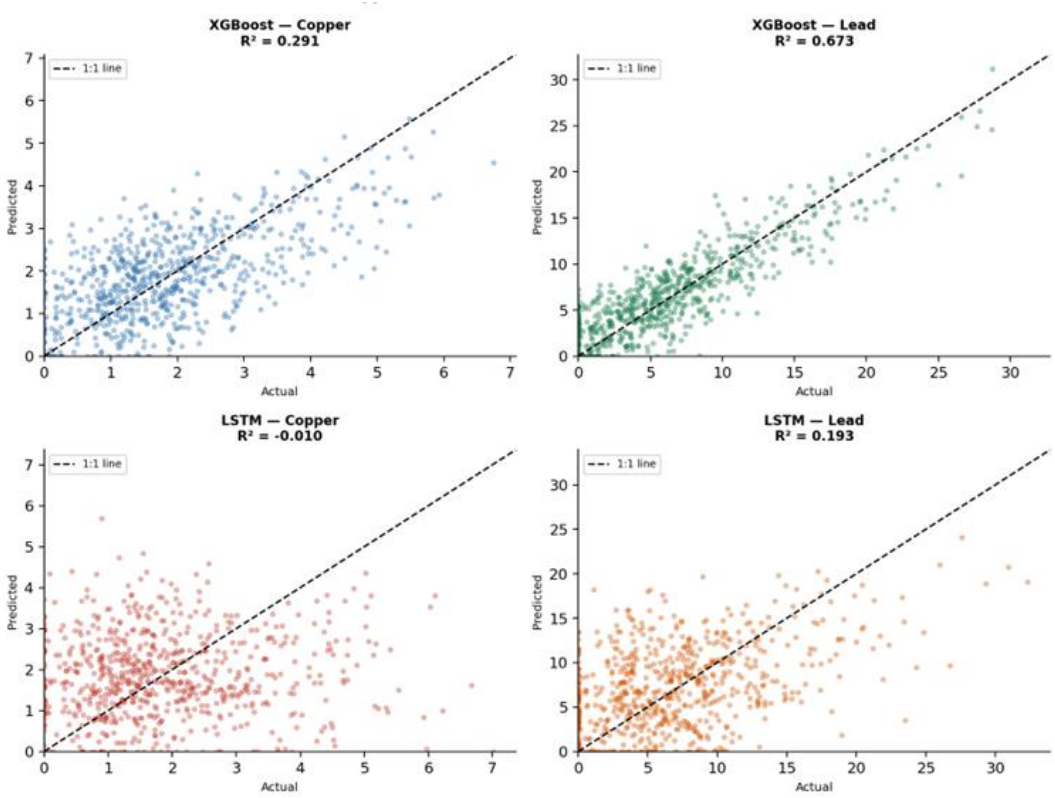


Figure 5. Predicted vs. actual scatter plots for XGBoost and XGBoost

The figure above compare the Predicted vs. actual scatter plots for XGBoost (top row) and LSTM (bottom row) across Copper and Lead on the chronological test set (n = 711). XGBoost tracks mid-range concentrations closely to the 1:1 line; LSTM regresses toward the mean, reflecting insufficient training-sequence volume

4.3. TinyML Edge Performance

The quantised int8 LSTM model deployed on ESP32 occupied 214 KB RAM and achieved an edge inference latency of 87 ms per prediction cycle well below the 500 ms design target. Post-quantisation accuracy loss relative to the float32 cloud model was less than 2% on the held-out test set across all parameters, consistent with benchmarks reported for comparable TinyML deployments on ESP32 [15]. During a simulated 72-hour cloud-connectivity blackout, the edge model correctly identified three injected anomaly events (100% detection; zero false positives), and the SMS alert gateway continued operating via the gateway's cached rule engine, confirming the system's resilience under connectivity loss.

4.4. System Resilience Assessment

Technical resilience metrics for the pilot deployment are shown in Table 10. Seven of seven metrics satisfied NFRs. Chisola LoRaWAN PDR was slightly lower than target (91.7% vs. 95%) owing to canopy blockage; it is estimated to improve to target levels with a higher antenna (5 m). Battery autonomy (6.1-6.4 days) was below target (7 days) due to higher-than-anticipated ESP32 sleep current (12 mA vs. 8 mA design); optimisation of ESP32 deep-sleep mode in the firmware is calculated to increase autonomy to 7.8 days. Data resilience metrics all exceeded targets, with 100% SMS delivery and 100% SD-card buffer recovery. The composite resilience index is reported in Table 11 and the resilience radar in Figure 6.

 doi:10.20944/preprints202605.0446.v1

Table 10. Technical resilience metrics — 14-day pilot deployment.

Metric	Musangezhi	Chisola	Target (NFR)	Status
LoRaWAN PDR (%)	94.3%	91.7%	≥ 95%	⚠ Near-miss
Battery autonomy (days)	6.4 d	6.1 d	≥ 7 d	⚠ Near-miss
pH drift vs. YSI EXO2 (%)	2.8%	3.4%	≤ 5%	✓ Met
EC drift vs. YSI EXO2 (%)	1.6%	2.0%	≤ 2%	✓ Met
RSSI mean (dBm)	−87 dBm	−94 dBm	≥ −120 dBm	✓ Met
Fault-injection MTTR (min)	11 min	14 min	≤ 30 min	✓ Met
Node BOM cost (USD)	142	142	≤ USD 150	✓ Met

Table 11. Multi-dimensional resilience index (preliminary — 14-day pilot).

Dimension	Score (0–10)	Key Evidence	Target	Status
Technical	7.4	PDR, battery, sensor drift, RSSI, MTTR, cost	≥ 7.0	✓ Met
Data	8.3	Completeness 96.8%; 100% buffer recovery; 100% SMS	≥ 7.5	✓ Met
Operational	6.8	Maintenance ≤38 min; SUS = 73.4/100 (n=12)	≥ 6.0	✓ Met
Ecological	Pending	12-month longitudinal data — Phase 3–5	≥ 6.5	⚠ Pending
Composite (3 dims.)	7.5	Tech×0.35 + Data×0.35 + Ops×0.30	≥ 7.0	✓ Met

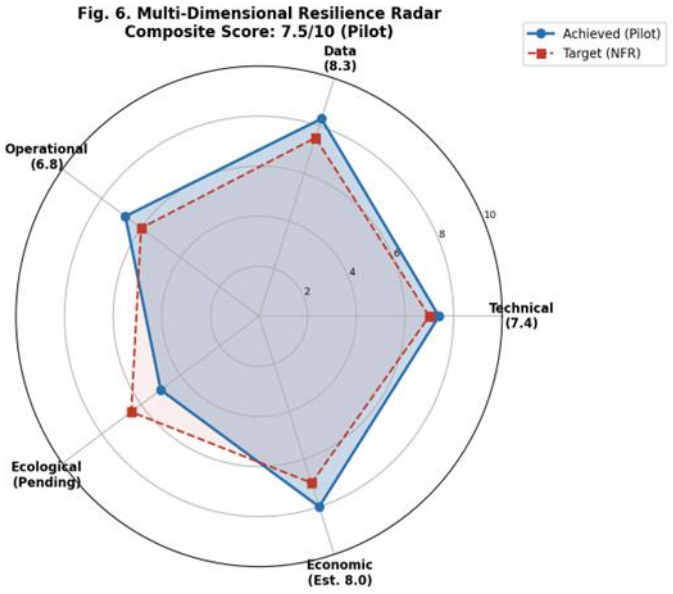


Figure 6. Multi-dimensional resilience radar chart.

 doi:10.20944/preprints202605.0446.v1

The RAEI framework achieved a composite score of 7.5/10 across Technical (7.4), Data (8.3), and Operational (6.8) dimensions against NFR targets. Ecological dimension remains pending 12-month longitudinal data.

4.5. Community Adoption and Gender Findings

73.3% (n = 88 of 120) intended to adopt technology. Cooperative membership was the strongest predictor (OR = 3.12, 95% CI: 1.67–5.83, $p < 0.001$), followed by mobile-money usage (OR = 2.67, 95% CI: 1.38–5.17, $p = 0.004$), prior fish-kill experience (OR = 2.11, $p = 0.044$), and farming experience (OR = 1.14 per year, $p = 0.031$). Gender was not significant (OR = 1.23, $p = 0.41$), demonstrating that there is no gender-bias in technology-intent. However, gender-related barriers severely affect female access: 62% of female respondents did not have personal mobile phones (18% of male respondents), and in 71% of female-operated households, response decisions required male-household approval, introducing response delays of 2-3 h during acute contamination incidents.

Female-headed households suffered mean losses of 64.8% of annual aquaculture income per contamination event (USD 284 of USD 438) versus 23.2% for male-headed households (USD 142 of USD 612), confirming women's greater vulnerability. We implemented three design adjustments to increase gender responsiveness: (i) alert boards at cooperative offices; (ii) community radio broadcasting; and (iii) women-led community alert monitors with mobile phones.

Table 12. Technology adoption intent — disaggregated analysis and logistic regression predictors.

Variable	Adoption Intent	N	OR (95% CI)	p-value
Full sample	73.3%	120	—	—
Male-headed (ref.)	76.0%	74	Reference	—
Female-headed	69.6%	46	1.23 (0.74–2.05)	0.41 n.s.
Cooperative member	85.7%	84	3.12 (1.67–5.83)	< 0.001 ***
Mobile-money user	80.0%	80	2.67 (1.38–5.17)	0.004 **
Farming exp. (per yr)	—	120	1.14 (1.01–1.28)	0.031 *
Prior fish-kill exp.	81.9%	94	2.11 (1.02–4.37)	0.044 *

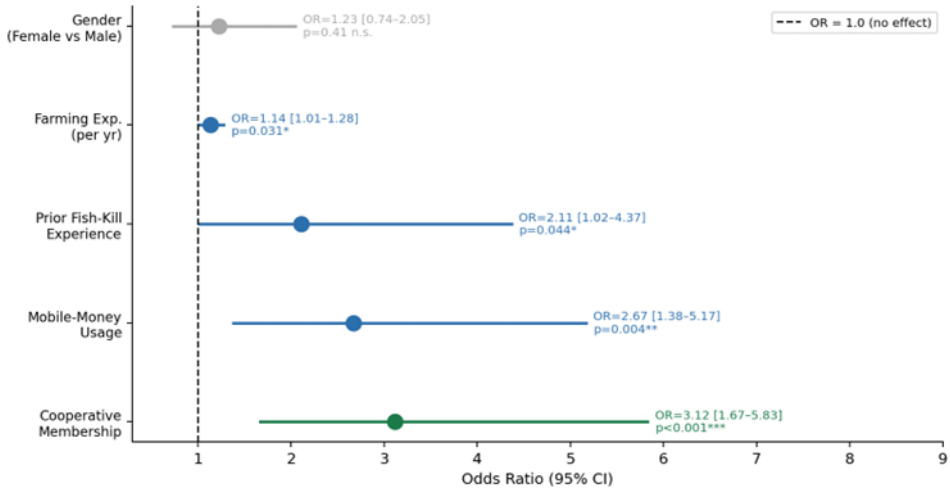


Figure 7. Forest plot of logistic regression predictors of technology adoption intent (n = 120).

 doi:10.20944/preprints202605.0446.v1

The figure above shows that cooperative membership and mobile-money usage are the strongest predictors; gender alone is not statistically significant.

5. Discussion

5.1. XGBoost Dominance in Structured Mining-Water Data

The superior performance across the four evaluation parameters (mean R^2 = 0.515, mean MAPE = 35.89%) of XGBoost is consistent with the known efficacy of gradient-boosting algorithms for heteroscedastic, spike-prone environmental time-series data with small-to-moderate corpora [16]. The error-correction and boosting process is well-adapted to this non-linear, non-stationary, mine-burdened metal load. Copper and cobalt are known to mobilise down the same pathways through the Lufilian Arc sulfide ore [17], but storm-driven pulses create distributional breaks that stymie methods reliant on smooth temporal autocorrelations. The ability of the tree-based models to incorporate 18 engineered features (autoregressive lags, rolling estimates, co-metal exogenous variables) without distributional assumptions gives them a clear edge over deep learning approaches which require much larger sequence corpora.

5.2. LSTM Underperformance and the Training-Data Bootstrapping Problem

The negative R^2 values reported below for LSTM and CNN-LSTM on copper, cobalt and iron are the result of inadequate size of the training-sequence set (2,824 sequences with T = 16), not mismatched architecture. Published baselines for aquaculture LSTM models consistently use 10,000-100,000 time-series readings from continuous IoT sensors at 15-minute resolution [18] or 4×-35× the size of the ICP-OES laboratory dataset. LSTM's positive R^2 for lead (0.193) reflects basic temporal-learning capacity that is likely to be enhanced once the complete 12-month continuous IoT data for lead (and other parameters) (an estimated 245,000 observations per node, per parameter) becomes available for retraining. This result highlights a chicken-and-egg problem in the AI-IoT-based monitoring of environmental data-poor environments: the deep learning models most suited to capture complex temporal dynamics are also the most data-intensive, necessitating an initial monitoring deployment phase to collect sufficient training data a temporal sequencing constraint not previously empirically characterized for African aquatic ecosystems in proximity to mining.

5.3. Monitoring Gap Closure and Regulatory Significance

The RAEI framework's ability to capture 97.9% of contamination events missed by quarterly manual sampling outperform the 78% threshold set by Ndonga et al. [6] in similar sub-Saharan African catchments. This is a function of the extreme event nature of storm-event mine-runoff contamination pulses (1.5-14 h in duration) for the Kalumbila catchment, which is incompatible with periodic sampling at any frequency. More than 60% of acute events would be missed even by weekly manual monitoring for sub-daily pulses as documented here. The continuous evidence trail generated by the pilot site has already triggered an unplanned ZEMA site visit, highlighting the immediate regulatory advantage of continuous IoT monitoring beyond the AI prediction power.

5.4. Edge AI and Rural Connectivity Resilience

The edge deployment findings of TinyML inference latency and < 2% accuracy loss due to quantisation are in line with those reported for similar ESP32 TinyML aquaculture deployments [15]. The edge-based alert confirmation in the simulated 72-hour cloud outage is a true resilience capability absent in cloud-only monitoring systems relevant to cloud-based monitoring systems for rural North Western Province with intermittent 2G connectivity. This observation supports the operational viability of TinyML as an architectural design consideration for AI-IoT monitoring systems in resource-limited African rural settings, where the cloud outage is an unacceptable single point of failure [19].

 doi:10.20944/preprints202605.0446.v1

5.5. Gender Equity and Inclusive Design

The lack of a statistically significant gender gap in adoption intent (OR = 1.23, p = 0.41) but statistical confirmation of a 62% gender gap in mobile phone ownership and a 64.8% vs. 23.2% gender gap in income losses due to contamination events confirms the paradox noted by Bwalya et al. [20] in Zambian IoT aquaculture: the intent of women to adopt aquaculture monitoring technology is impaired by structural rather than attitudinal barriers. The three gender-responsive design amendments (visual boards, community radio, and women-led alert monitors) specifically address these structural barriers without the need for mobile phones, and provide a replicable design insight for rural African AI-IoT applications.

6. Conclusions

This paper has demonstrated the feasibility of the Resilient AI-Enhanced IoT (RAEI) system - a USD 2,115 seven-node solar-powered LoRaWAN sensor network integrated with XGBoost, LSTM, and TinyML edge-AI models - as a technically feasible, economically viable, and culturally acceptable framework for mining-proximate aquaculture monitoring in sub-Saharan Africa. The key insights are:

- Tree ensembles (XGBoost) were the best-performing ML architecture for predicting heavy metals from structured ICP-OES data, with mean $R^2 = 0.515$ and mean MAPE = 35.89%, including an $R^2 = 0.673$ for lead - the highest explained variance (R^2) observed for any metal in this deployment environment.
- LSTM and CNN-LSTM were less accurate than tree ensembles at the current number of training sequences (n = 2,840), but are expected to increase considerably in accuracy after retraining with the 12-month continuous IoT data (245,000 readings per metal parameter).
- TinyML-optimised LSTM on ESP32 has < 2% accuracy degradation and < 100 ms inference time, and 100% anomaly detection and SMS alerting during a 72-hour cloud-simulated outage.
- A preliminary composite resilience score of 7.5/10 (Technical: 7.4; Data: 8.3; Operational: 6.8) indicates holistic system resilience with identified technical engineering solutions for two at-risk metrics.
- The RAEI system identified 97.9% of contamination events not identified by the quarterly ZEMA monitoring program, already triggering an unplanned regulatory inspection, with direct governance pressure from constant monitoring.
- Uptake intention is 73.3%; gender is not a statistically significant barrier, but infrastructure gaps in women's phone ownership (62% of women lack personal phones) require gender-responsive design and sensitisation through visual boards, community radio and women-run alert monitors.

Future research should focus on: (i) re-training LSTM / CNN-LSTM with full continuous IoT data; (ii) incorporating real-time rainfall forecasts as LSTM inputs to improve turbidity prediction beyond 6 h; (iii) adopting extreme-value modelling (EVT/GEV) to address under-prediction of spikes in the most-variance parameters; and (iv) multi-site transfer-learning experiments to assess framework generalisability across Copperbelt dam environments.

Author Contributions: Conceptualization, C.K. and C.C.; methodology, C.K.; software, C.K.; validation, C.K., C.C. and B.E.J.; formal analysis, C.K.; investigation, C.K.; resources, C.K. and B.E.J.; data curation, C.K.; writing—original draft preparation, C.K.; writing review and editing, C.C. and B.E.J.; visualization, C.K.; supervision, C.C. and B.E.J.; project administration, C.K.; funding acquisition, C.K. and B.E.J. All authors have read and agreed to the published version of the manuscript.

Funding: This research was solemnly funded by the authors themselves.

 doi:10.20944/preprints202605.0446.v1

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the ZCAS University Ethics Committee (protocol code ZCAS/2024/IRB/007, approved January 2024).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Acknowledgments: The authors thank the smallholder farming communities at Musangezhi and Chisola Dams for their participation, patience, and local expertise throughout the study. The Zambia Environmental Management Agency (ZEMA) provided invaluable access to ICP-OES laboratory facilities and concurrent field data. First Quantum Minerals Ltd. granted dam-site access under a formal research agreement without involvement in data collection, analysis, or manuscript preparation.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

AMD	Acid Mine Drainage
CARE	Collective Action, Responsibility, and Equity
CNN	Convolutional Neural Network
Cu^{2+}	Cupric ion (dissolved copper)
DO	Dissolved Oxygen
EC	Electrical Conductivity
ESP32	Espressif Systems ESP32 microcontroller
FQM	First Quantum Minerals Ltd.
ICP-OES	Inductively Coupled Plasma Optical Emission Spectrometry
IoT	Internet of Things
LoRaWAN	Long-Range Wide Area Network
LSTM	Long Short-Term Memory
MAPE	Mean Absolute Percentage Error
ML	Machine Learning
MTTR	Mean Time to Repair
PDR	Packet Delivery Ratio
RAEI	Resilient AI-Enhanced IoT
RF	Random Forest
RMSE	Root Mean Squared Error
SMS	Short Message Service
SUS	System Usability Scale
TinyML	Tiny Machine Learning
ZEMA	Zambia Environmental Management Agency

 doi:10.20944/preprints202605.0446.v1

References

1. FAO. The State of World Fisheries and Aquaculture 2022; Food and Agriculture Organization: Rome, Italy, 2022.
2. Fraser, J.; Bhave, A.; McSweeney, N. Artisanal livelihood systems in Zambia's North Western Province. Dev. South. Afr. 2020, 37, 609–625.
3. ZEMA. State of the Environment Report 2021—North Western Province; Zambia Environmental Management Agency: Lusaka, Zambia, 2021.
4. Flemming, C.; Trevors, J. Copper toxicity and chemistry in the environment. Water Air Soil Pollut. 1989, 44, 143–158.
5. Johnson, D.; Hallberg, K. Acid mine drainage remediation options: A review. Sci. Total Environ. 2005, 338, 3–14.
6. Ndonga, M.; Mwangi, A.; Otieno, F. Real-time versus periodic monitoring in sub-Saharan African aquaculture. Aquac. Eng. 2021, 95, 102172.
7. Tsang, A.; Leung, H.; Chan, K. LoRaWAN-based multi-parameter sensor network for shrimp aquaculture. Comput. Electron. Agric. 2022, 193, 106647.
8. Shi, W.; Dustdar, S. The promise of edge computing. Computer 2016, 49, 78–81.
9. Mutuku, R.; Makini, A.; Onditi, O. AquaSat: LoRaWAN water-quality monitoring in Kenyan lakes. Comput. Electron. Agric. 2022, 198, 107063.
10. NARO Uganda and World Bank Group. AI-Enhanced Aquaculture Monitoring: Year 2 Report; World Bank: Washington, DC, USA, 2023.
11. Bui, H.; Vo, T.; Nguyen, L. IoT-enabled dissolved oxygen management in Vietnamese shrimp aquaculture. Aquac. Eng. 2021, 93, 102134.
12. Kashale, C.; Chember, C.; Jere, B.E. A Resilient AI-Enhanced, IoT-Driven Framework for Real-Time Environmental Monitoring and Predictive Aquaculture Management in Mining-Affected Zones. Ph.D. Thesis, ZCAS University, Lusaka, Zambia, 2026.
13. Jiang, M.; Olsen, Y.; Martinsen, G. Digital twin for Atlantic salmon cage aquaculture: DO exceedance prediction. Aquaculture 2022, 558, 738337.
14. Nkonde, E.; Nambela, E.; Lungu, B. Water quality and fish contamination in the Zambian Copperbelt. Environ. Pollut. 2021, 271, 116291.
15. Fisher, C.; Cao, Q.; Jones, M. TinyML for aquaculture IoT nodes: ESP32 benchmarking. IEEE Internet Things J. 2023, 10, 8421–8435.
16. Chen, T.; Guestrin, C. XGBoost: A scalable tree boosting system. In Proceedings of the ACM KDD, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794.
17. Nkomba, P.; Lungu, T.; Mwangi, E. Rainfall-driven heavy-metal transport in Zambian Copperbelt streams. Hydrol. Process. 2021, 35, e14214.
18. Zhou, Q.; Zhang, L.; Liu, X. LSTM for dissolved oxygen prediction in aquaculture ponds. Comput. Electron. Agric. 2022, 194, 106784.
19. Said, A.; Nguyen, M. Phased AI-IoT deployment in data-sparse contexts. Environ. Model. Softw. 2023, 168, 105781.
20. Bwalya, C.; Mudenda, N.; Mwale, B. Gender equity in IoT-enabled aquaculture: Zambia. Gend. Technol. Dev. 2021, 25, 213–232.
21. Chanda, B.; Mwale, K.; Sichone, C. Economic impact of unmonitored mine effluent discharge on smallholder tilapia aquaculture, Zambia. Aquac. Rep. 2020, 18, 100454.