KNS Spring Meeting
ICC Jeju, May 6-8, 2026

사단법인 한국원자력학회
KOREA NUCLEAR SOCIETY

Intelligent Data Tiering Design for Optimizing Data Availability and System Load in Nuclear Activity Detection Platforms

Sueyeon Lee, Gayeon Ha, Yonhong Jeong*

Center for Nuclear Nonproliferation Strategy & Technology, Korea Institute of Nuclear Nonproliferation and Control (KINAC)
**Corresponding author: jyh1404@kinac.re.k*

Introduction

Backgrounds

- Commercial satellite imagery + multilingual OSINT → exponential data growth in nuclear monitoring
- Current systems: uniform storage — all data treated equally regardless of value

Result : access latency ↑, resource waste ↑, false positive rates ↑

Objectives

- Design a dynamic 3-tier framework that classifies data by analytical urgency and verified value in real time
- Minimize system load + maximize detection accuracy through intelligent tiering

Case Study

Case A: CTI Systems & Data Lifecycle Management

- CTI faces the same challenge: massive heterogeneous data + real-time threat identification
- Sun et al. (2023): staged pipelines with distinct latency needs — collection (ms) → validation (batch) → archival (long-term) [1]
- Wagner et al. (2019): orgs processing 10,000+/day use differentiated hot/cold pathways [2]

→ Direct mapping to nuclear monitoring: urgent anomalies (Tier 1) → proliferation validation (Tier 2) → long-term analysis (Tier 3)

Case B: Adaptive Intelligent Tiering (AIT)

- Pang et al. (2023): DL access prediction + RL placement optimization → up to 85% gain vs. LRU/LFU [3]
- Key: tier placement by contextual value, not file age or frequency

→ Nuclear application: months-old thermal imagery instantly becomes high-priority when suspicious procurement is detected — static rules cannot capture this

	LRU / LFU (Traditional)	AIT / Proposed (RL-based)
Strategy	Static rules: file age or access count	Context-aware predictive placement via RL
Performance	Baseline	Up to 85% improvement
Adaptability	No real-time adaptation; requires manual tuning	Continuously learns; adapts to workload changes
Value Assessment	Access frequency only	Multi-factor: temporal, confidence, access, PIR

Table 1. AIT vs. Traditional Policies (Pang et al., 2023)

Methodology

Proposed 3-Tier Architecture

- **Tier 1 — Hot (7-day):** NVMe SSD, sub-ms latency.

Real-time: thermal anomalies, OSINT triggers, steam discharge → GPU-accelerated LLM inference

- **Tier 2 — Warm (30-90 day):** SSD/HDD hybrid, ~seconds.

Pattern matching, geospatial contextualization, expert review → routed by Bayesian confidence S(d)

- **Tier 3 — Cold (indefinite):** HDD arrays, ~minutes.

Facility profiles, proliferation networks → trend analysis & AI training

- **Data distribution:** Hot 5-10% | Warm transitional | Cold 85-90%

Attribute	Tier 1 (Hot)	Tier 2 (Warm)	Tier 3 (Cold)
Retention	7 days	30–90 days	Indefinite
Storage	NVMe SSD	SSD/HDD Hybrid	HDD Array
Latency	Sub-millisecond	Second-scale	Minute-scale
Score (S)	Pre-scoring (input)	$0.30 \leq S < 0.75$	$S \geq 0.75$ → Tier 3; $S < 0.30$ → Purge
Data Volume	5–10%	Transitional	85–90%
Key Content	Satellite anomalies, OSINT triggers, thermal imagery	Pattern matching, geospatial context, expert review	Facility profiles, proliferation networks

Table 2. Three-Tier Data Management Architecture

Migration Logic & Scoring Function

```
S(d) = w₁·f_temporal(d) + w₂·f_confidence(d) + w₃·f_access(d) + w₄·f_PIR(d)
```

Weight Optimization (NEW):

- $W = \{w_1, w_2, w_3, w_4\}$ optimized via **Deep Q-Network (DQN)**
- State: current tier occupancy + processing latency
- Reward: $R = \alpha \cdot \text{throughput} + \beta \cdot \text{PIR\_detection} - \gamma \cdot \text{cost}$
- Incentivizes: max throughput & PIR accuracy, min operational cost

Decision Thresholds (IC-Aligned):

- **τ = 0.75** → Promote to Tier 3 (median of IC High Confidence 0.70–0.80)
- **ρ = 0.30** → Purge (lower bound of IC Low Confidence)
- [0.30, 0.75) → remains in Tier 2 for continued expert evaluation

Each component normalized to [0, 1]:

Component	Variable	Calculation	Data Source	Remarks
Temporal Decay	$f_{temporal}(d)$	$e^{-\lambda(t_{now}-t_{created})}$	System metadata	Exponential decay
Confidence Score	$f_{confidence}(d)$	$\frac{1}{n}\sum_{i=1}^{n} P_i(v)$	Validation Layer	Bayesian probability
Access Frequency	$f_{access}(d)$	$\frac{\log(1+\text{freq})}{\log(1+\text{freq}_{max})}$	Access log	Log-normalization
PIR Correlation	$f_{PIR}(d)$	$\frac{V_d \cdot V_{PIR}}{\lVert V_d\rVert \lVert V_{PIR}\rVert}$	OSINT / Mission vectors	Cosine similarity

Table 3. Scoring Function Components

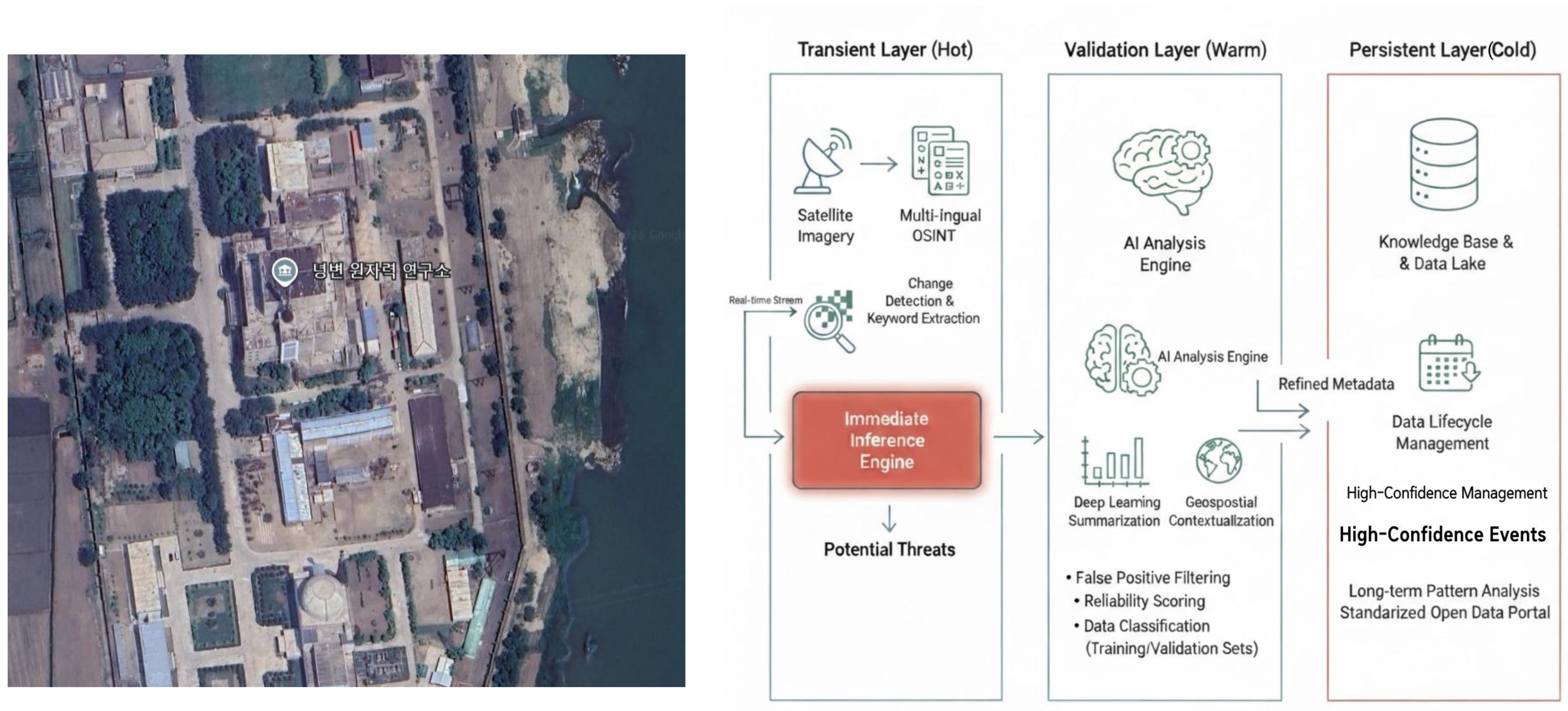

Fig 1. (a) Satellite imagery of a nuclear facility and (b) proposed 3-tier data tiering architecture

Expected Results

Based on empirical benchmarks from analogous CTI and AIT systems:

- Processing latency: **60–85% reduction** via Tier 1 GPU-accelerated inference
- Storage costs: **40–50% reduction** (only 5–10% in Hot Storage)
- Throughput: **100× increase** (500 → 50,000 items/day) via automated Validation Layer
- False positives: **70–80% reduction** through multi-stage Bayesian validation
- Linear scalability — Cold tier (85–90%) scales at minimal cost
- RL-based migration adapts to evolving threat landscapes without manual tuning
- IC-aligned thresholds (τ=0.75, ρ=0.30) ensure principled data routing

Conclusions

- Proposed a dynamic 3-tier framework that addresses information overload in nuclear activity detection through context-aware, adaptive data routing
- RL-optimized multi-factor scoring with IC-aligned thresholds (τ=0.75, ρ=0.30) enables principled tiering — unlike static HSM rules, data value changes are captured dynamically
- Validated Tier 3 intelligence serves as high-quality AI training sets; Tier 1 real-time streams enable proactive threat detection
- Framework demonstrates linear scalability — Cold tier (85-90%) scales at minimal cost with automatic RL adaptation
- Future work: pilot deployment, real-world threshold refinement, federated learning for multi-agency sharing, Explainable AI (XAI)

References

[1] N. Sun et al., Cyber Threat Intelligence Mining for Proactive Cybersecurity Defense, IEEE Comm. Surveys & Tutorials, Vol.25, No.3, 2023.
[2] T. D. Wagner et al., Cyber threat intelligence sharing: Survey and research directions, Computers & Security, Vol.87, 2019.
[3] L. Pang et al., Adaptive intelligent tiering for modern storage systems, Performance Evaluation, Vol.160, 2023.