AI TO PROTECT CRITICAL INFRASTRUCTURE

JACKIE, PATRICK, PAIGE, KAHU,
ABIA AND ANNIE

INTRODUCTION

AI and IoT-Driven City-Scale Sensing for Roadside Infrastructure Maintenance

MOTIVATION

Reactive maintenance (citizen reports, manual audits) is slow, costly, and incomplete - many issues go unreported.

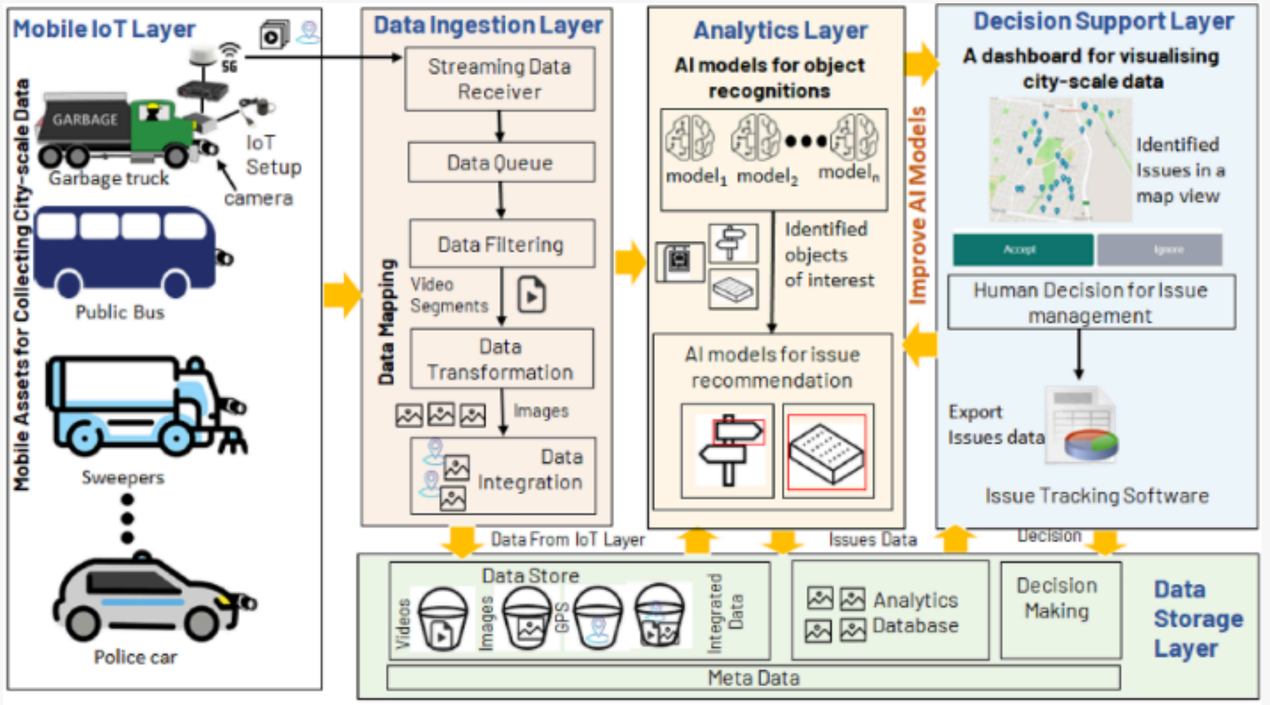
Cities need proactive, scalable sensing as networks expand and budgets tighten.

Opportunity: Repurpose municipal fleets & AI/IoT/5G to automate real-time detection of roadside “points of maintenance” (PoMs) and improve safety and service delivery.

SOLUTION AND KEY FINDINGS

Framework: AIoT-CitySense (5 layers: Mobile IoT, Ingestion, Storage, Analytics, Decision Support).

Deployment: “Mobile IoT-RoadBot” on 11 waste trucks in Melbourne with stereo cameras, GNSS, edge compute, 5G, data to AWS AI services. 1 FPS frame sampling; map dashboard for triage.

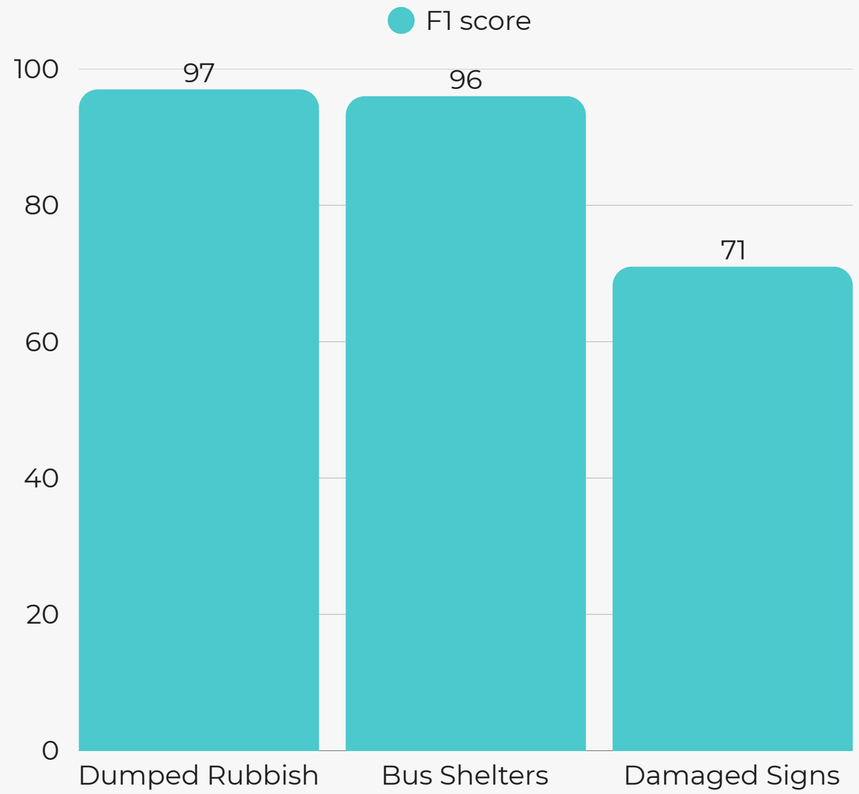


ML pipeline:

1. **Stage 1:** Pre-trained detector filters relevant objects (e.g., road signs, bus shelters) and supports optional privacy filtering.
2. **Stage 2:** Custom models per use case:
   a. Damaged signs: object detection with boxes, incremental learning and class imbalance (~10% positives).
   b. Dumped rubbish: binary classifier & post-filtering to remove bins/trees.
   c. Bus shelters: custom model (Rekognition Custom Labels).

Results	Damaged signs	Dumped rubbish	Bus shelter
F1 Score	71	97	96
	Solution only	Solution & reports	Reports only
Detection	85%	98%	2%

RESULTS



Results:

85% more dumped-rubbish PoMs than resident reports; 94% of overlaps found 3–4 days earlier.

Ops: ~5k images/truck/day; ~A$200/day cloud cost fleet-wide.

IMPLICATIONS / ETHICAL CONSIDERATIONS

Equity and governance:

1. Potential to improve service equity across suburbs if routes cover the whole municipality and prioritization is transparent.

Operational impact:

1. Proactive, city-scale sensing using existing fleets; faster remediation, reduced reliance on citizen labor, better coverage in low-participation areas.
2. Human-in-the-loop workflow (accept/ignore) enables continuous model improvement and accountable actioning.

Privacy and transparency:

1. Cameras capture people/plates/properties, requires robust, audited anonymisation (face/plate blurring), data minimization/retention limits, clear purpose limitation, and public communication.
2. Cloud/vendor dependence (Rekognition) raises cost control, portability, and resilience considerations.

LIMITATIONS / FUTURE WORK

Measurement and modeling:

1. Report standard detection metrics (mAP@IoU), condition-specific breakdowns (day/night/rain), latency, and calibration conduct independent field audits for ground truth.
2. Cascade risk: misses in Stage 1 cap recall consider end-to-end multi-task detectors (e.g., YOLO/Detectron2) and add segmentation for fine-grained damage.
3. Address class imbalance (focal loss, class weighting, targeted sampling, hard-negative mining), adopt active learning with human-in-the-loop.

Robustness and scale:

1. 10–20% data loss (coverage/ weather) push lightweight prefilters/ inference to edge to reduce bandwidth/cost and mitigate outages, add drift monitoring.
2. Generalisation from a single council/ fleet/camera vantage, apply domain adaptation and standardise interfaces for other fleets.

Scope expansion:

1. Extend to potholes, cracks, lane marking quality, and other assets. Integrate prioritisation models and work-order systems.

CYBER THREATS TO CRITICAL ENERGY INFRASTRUCTURE

MOTIVATION

- **Sophisticated Cyberattacks:** Traditional defenses are insufficient
- **High Stakes:** Economic and social stability are at risk.

SOLUTION AND KEY FINDINGS

Using AI to shift from reactive to proactive cybersecurity.

The paper proposed 3 solutions to detect anomalies or suspicious patterns at the **beginning** of an attack.

AI SOLUTIONS

Deep Learning

To detect ransomware in energy control systems

AI SOLUTIONS

Deep Learning

To detect ransomware in energy control systems

Machine Learning

To detect anomalies in operational hardware

AI SOLUTIONS

Deep Learning

To detect ransomware in energy control systems

Machine Learning

To detect anomalies in operational hardware

Predictive Analytics

To detect system failures or attacks in the power distribution network

RESULTS

SIGNATURE-BASED SYSTEMS

74.9% of threats accurately detected

12% false positive rate

AI SYSTEMS

94.7% of threats accurately detected

4% false positive rate

70% faster incident response

IMPLICATIONS AND ETHICAL CONSIDERATIONS

Security operators can intervene earlier in the threat cycle

Reduced alert fatigue

Faster security responses

LIMITATIONS AND FUTURE WORK

1. Adversarial attacks
2. Explainability of AI systems
3. Automated mitigations

PHYSICAL SECURITY PROTECTION OF CRITICAL INFRASTRUCTURE

MOTIVATION

Current surveillance systems can result in human-related errors, high costs and limited scalability.

SOLUTION AND KEY FINDINGS

Smart video-surveillance systems using distributed CPS

Uses edge devices to perform local video analytics and cloud servers to aggregate the data from these nodes.

Core innovation: Systems ability to reconfigure itself by switching modes based on detected threats.

RESULTS

System performance and efficiency

Performs in real-time at **32-98 fps**.

76% reduction is resource consumption.

Human detection

Achieves an F1-score of **0.9646 -** nearly matching the best model, but **36x smaller** in size

Facial recognition

Achieves **99.5%** on a standard data set

Object tracking

Performs **effectively** in multi-camera setups

IMPLICATIONS AND ETHICAL CONSIDERATIONS

Risk of function creep

Lack of privacy and anonymity

Demographic bias

LIMITATIONS AND FUTURE WORK

1

Environmental robustness

2

Scalability of re-identification

3

Vulnerability to adversarial attacks

SENSOR FAULT DETECTION IN WASTEWATER TREATMENT PLANTS.

MOTIVATION

Traditional methods struggle to identify complex, time-based sensor failures, which can cause environmental harm and energy waste.

DATA & SOLUTION

Data Collection

- Data from 12 sensors over 1 year ~5.3 million points
- Includes:
  - Chemical measurements: ammonia, oxygen, nitrates
  - Operational measurements: blower frequencies, temperatures

Data Preprocessing

- Missing values filled with the last known reading
- Engineered features from ammonia (e.g., mean, variance)
- Normalised for training

LSTM Architecture

type of neural network that processes sequences of data, remembering important patterns over time while ignoring irrelevant information, making it ideal for time-series

- Multi-layered neural network with 4 LSTM layers (60 units each)
- Processes 60-minute data windows
- Learns temporal patterns for normal vs. faulty behaviour
- Gates decides what to remember, forget, or pass forward

RESULTS

*The LSTM method achieved **96.5%** overall accuracy, with an **11%** improvement in precision compared to traditional approaches, which scored 93% accuracy and 65.9–70.7% precision.*

Mehood / Metric	Variance Analysis	PCA-SVM	LSTM
Accuracy (%)	93.3	93	96.5
Recall (%)	80.9	94.1	93.9
Precision (%)	70.7	65.9	81.7
F1-score (%)	75.4	77.5	87.4

**ARIMA could not distinguish between normal and faulty conditions because it only considers short-term memory.*

LIMITATIONS

One sensor fault type focus - only collective faults in ammonia sensors, not other sensor failures or fault categories

Geographical specificity - being a treatment plant in Valdobbiadene (where Prosecco wine is produced), experience unusual patterns of organic mass during the harvest season.

Single Facility Validation - only used data and tested in the facility that caters to a Population Equivalent (PE) load of 10,000.

AND FUTURE WORK

- **Multi site validation** (<1,000 PE to > 100,000 PE)
- **Cross-regional testing** in facilities without seasonal agricultural loading patterns
- **Multi-sensor fault detection** - expand beyond ammonia sensors (e.g. dissolved oxygen, pH)

IMPLICATIONS / ETHICAL CONSIDERATIONS

Over-reliance on automation
reduction in human oversight and backup monitoring

Algorithmic Transparency
Black Box/Limited interpretability

Workforce Displacement
“goal is a system with minimal human supervision”

CLIMATE CHANGE & PROTECTING CRITICAL INFRASTRUCTURE

MOTIVATION

- Effects of Climate Change getting worse and posing threat to Critical National Infrastructure
- 2021 Texas Snowstorms: 4.5 million homes without power
- 2023 Cyclone Gabrielle: critical road & transport infrastructure destroyed, ~$1b spent on recovery
- Events have severe consequences including loss of life

Solution #1:

AI MODELS TO PREDICT IMPACT

Train models to produce predictive simulations of the impact of climate change on nearby infrastructure

Delteres: Project based out of the Netherlands to simulate effects of sea level rise and its impact on infrastructure

Inputs:

- 🏘️ Nearby Infrastructure Information
- 🌏 Climate Data

Outputs:

- 🌊 Predictions of how infrastructure (eg, roads, drainage systems, etc) will be impact by sea level rise
- 🧱 Protection recommendations

Goal:

- 🏰 Help build resilient infrastructure and prevent outages / damage caused by sea level rise

Solution #2:

AI POWERED EARLY WARNING SYSTEMS

Train models to identify precursor indicators to climate events for early alerting and response effrots.

FloodNet: Deep learning model to identify when flooding is likely to occur and provide early alerting for timely responses.

Same techniques can be used to build models for alerting on other events (eg, wildfires)

Inputs:

- 🏞️ Images of bodies of water
- 💦 Data metrics (eg, water level & flow rate)

Outputs:

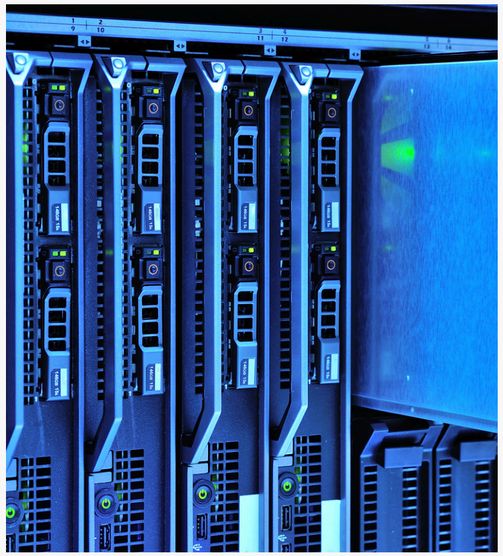
- 🎲 Likelihood of flooding
- ⚠️ Severity of possible flooding

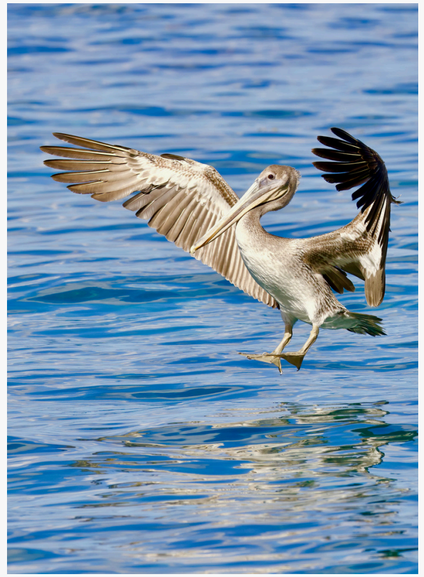
Goal:

- 📣 Provide early warnings to help cities prepare & initiate timely responses

LIMITATIONS

- Availability of sufficient volumes of high-quality region-specific data
  - Models need high quality, up to date information
- Limited results from real world testing
  - How do results compare to existing human capabilities?





ETHICAL CONSIDERATIONS

- Performance & Explainability
  - Requires low rate of false negatives
  - Humans need to understand how AI got to its outputs
- Climate Considerations
  - Energy use & carbon footprint of AI models
  - Contributing to underlying problem

DISCUSSION

Questions

- Given that AI models like deep learning and LSTMs are not always transparent, what ethical and safety concerns arise when they are used to make critical decisions in energy infrastructure? What would make you comfortable (or uncomfortable) with these "black box" algorithms managing the power grid?

- How can we ensure accountability and maintain public trust when an AI system, rather than a human, makes a decision that leads to a widespread power outage or security breach?

- The papers suggest AI can reduce "alert fatigue" for human operators. But what is the ideal division of labour between AI and humans in critical infrastructure? Where should the AI's role end and the human's begin?

- What are the areas where you can see AI having the most impact or being the most trusted to protect critical infrastructure, and what level of human oversight would be needed to implement these systems in the real world?

- Critical infrastructure is often invisible to the public until something goes wrong. How much transparency should there be about the role AI is playing in protecting these systems? Would too much detail risk security, or would it help build public trust?

- While designed for security, could the anonymised data from this system be used for other beneficial smart city applications? For instance, optimising pedestrian flow, managing emergency evacuations, or planning public services? How could this be done without compromising the privacy of people mentioned earlier?