Infrastructure Anomaly Detection and Failure Prediction

- An Imbalanced Data Problem

A/Prof Yang Wang

Nov 2022

2016
- Water Professional of the year winner

- ANNY excellence in analytics

2017
- iAwards merit

2018
- EUREKA PRIZE
- AWA High Commendation

2019
- iAwards merit
- Smart Cities Awards

2020
- iAwards winner
- ITS award finalist

Data Driven Solutions for Infrastructure and Asset Management

Help asset and infrastructure owners identify and predict asset behavior, reduce future risk, optimize their performance through data analytics and machine learning.

Road and bridge: structural health monitoring

Water and sewer: failure prediction

Power facilities: Proactive maintenance

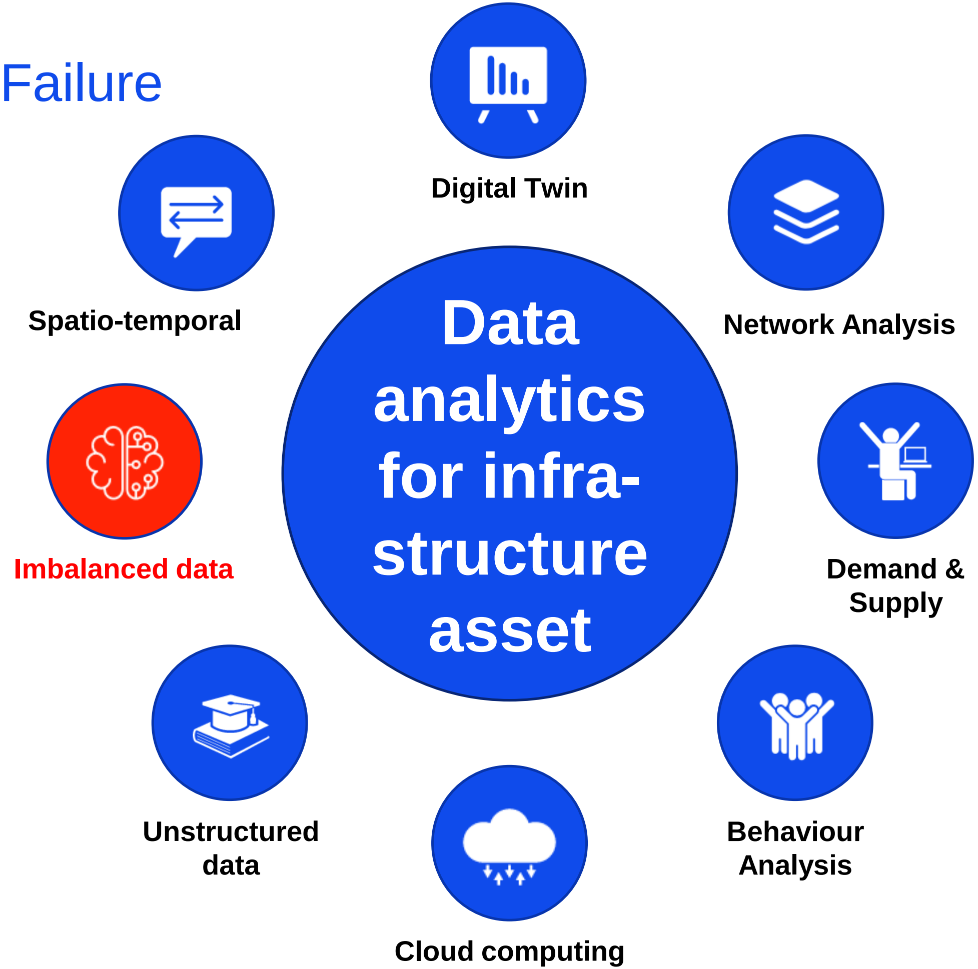
Telecommunication and broadband: demand forecasting

Train and rail: performance optimization

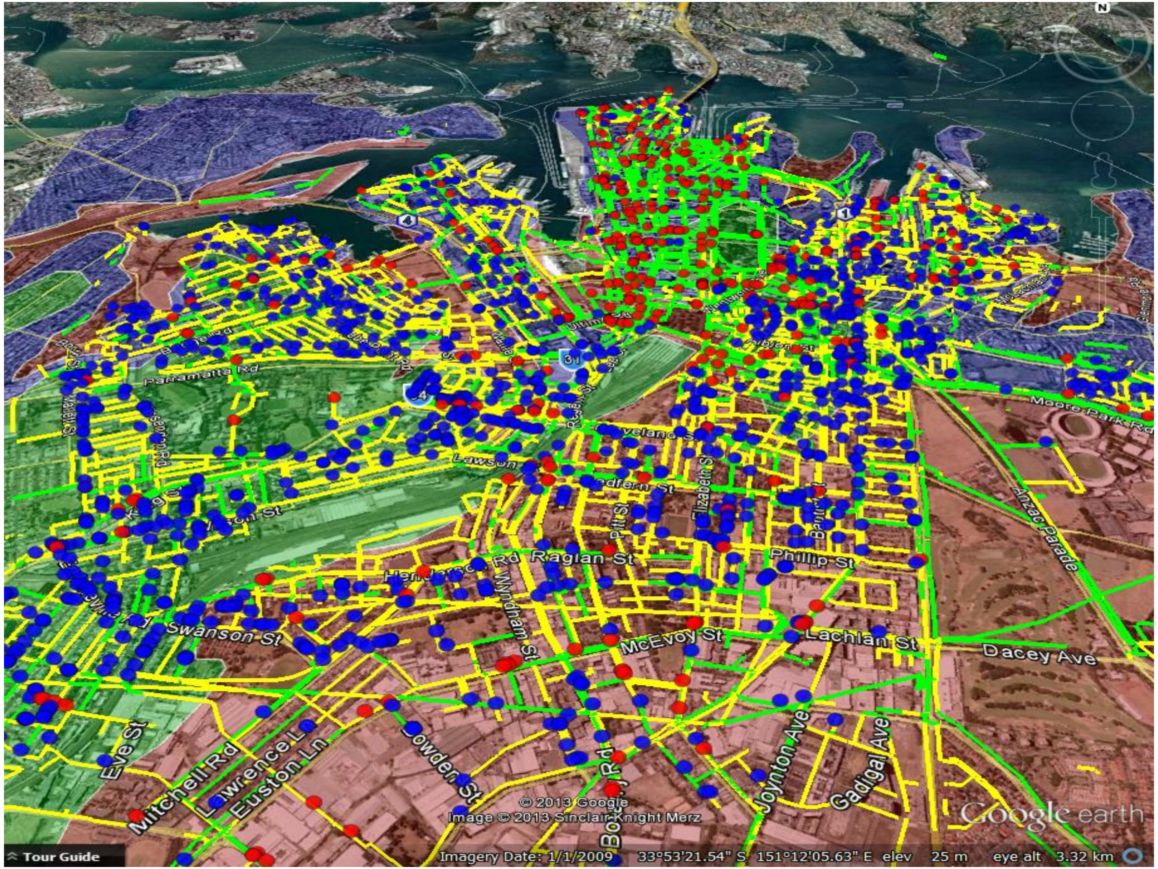
Finance: Cashflow prediction

Infrastructure Anomaly & Failure

- Imbalanced data problem

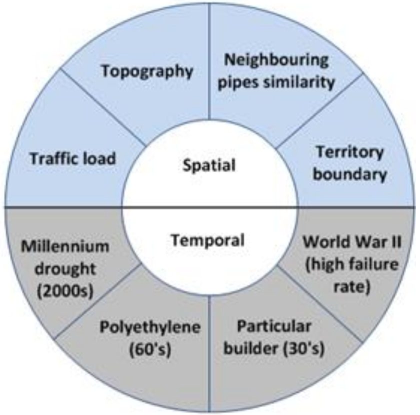


Water Main Failure Prediction



Water Main Failure Prediction

- Pipe failure prediction based on learning from historical failure records and attributes related to failure.

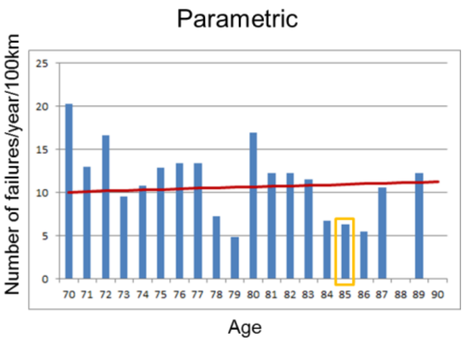


- Multiple factors (20+)
- Diverse failure patterns
- Sparse failure data (1%)
- Incomplete dataset
- Long term prediction with confidence estimation

Innovative Solutions

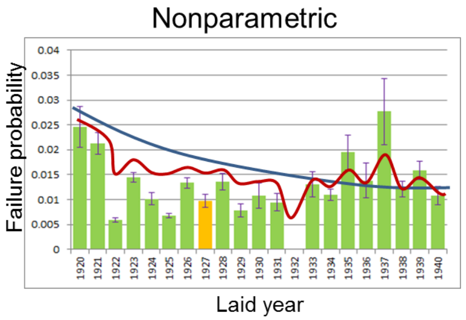
Novelty

- Our novel data-driven failure prediction solution is based on non-parametric learning models and non-homogenous stochastic process model



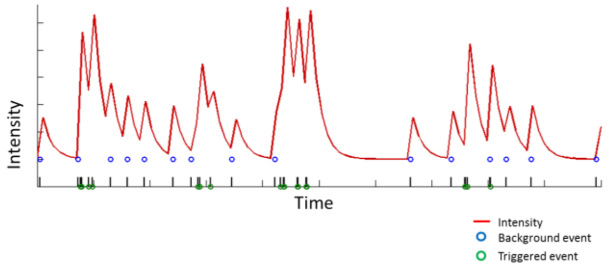
Non-parametric learning

- Avoid strong assumptions on model structure
- Complexity growing with the data observations
- Consider all the available factors
- Model spatial connectivity



Non-homogenous stochastic point process-based statistical model (interaction point process)

- A generic and adaptable tool for modelling series of events (e.g./ pipe failures).
- Trigger intensity is determined by previous failures.



***Intensity**: **Expected number of events** (pipe failures) at time points.

Hierachical Beta Process

- Top level: network to pipe groups

  Group failure patterns

  $q_k \sim \mathrm{Beta}(c_0 q_0, c_0(1-q_0))$, where $k = 1, \ldots, K$

- Middle level: pipe group to individual pipes

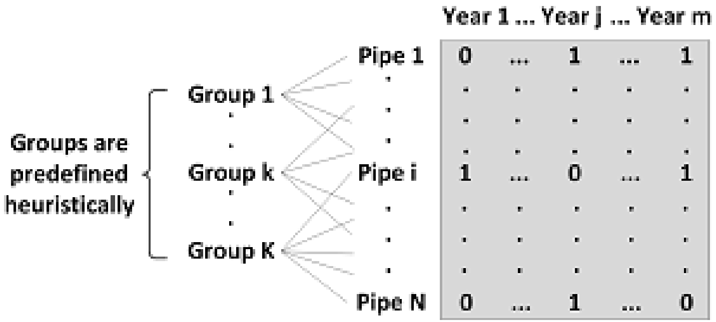
  Pipe failure probabilities

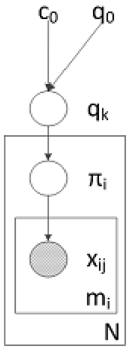
  $\pi_{k,i} \sim \mathrm{Beta}(c_k q_k, c_k(1-q_k))$, where $i = 1, \ldots, n_k$

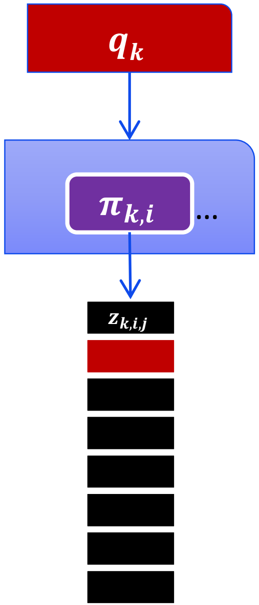
- Bottom level: Individual pipe to failure observations

  Failure observations over years

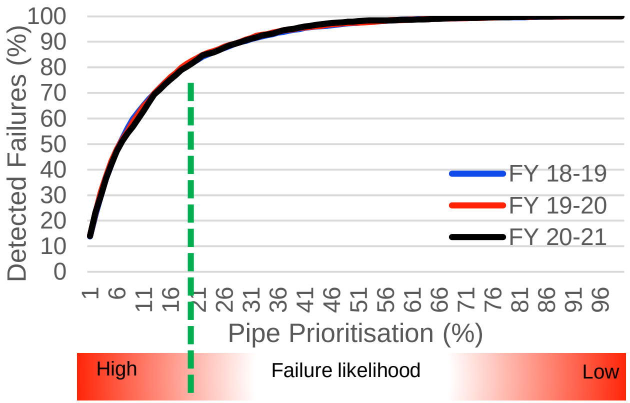
  $z_{k,i,j} \sim \mathrm{Ber}(\pi_{k,i})$, where $j = 1, \ldots, m_{k,i}$

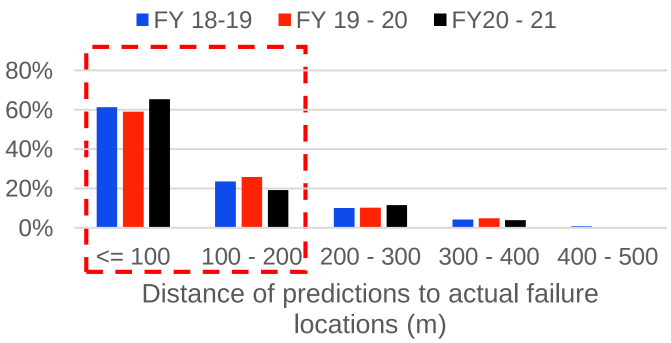


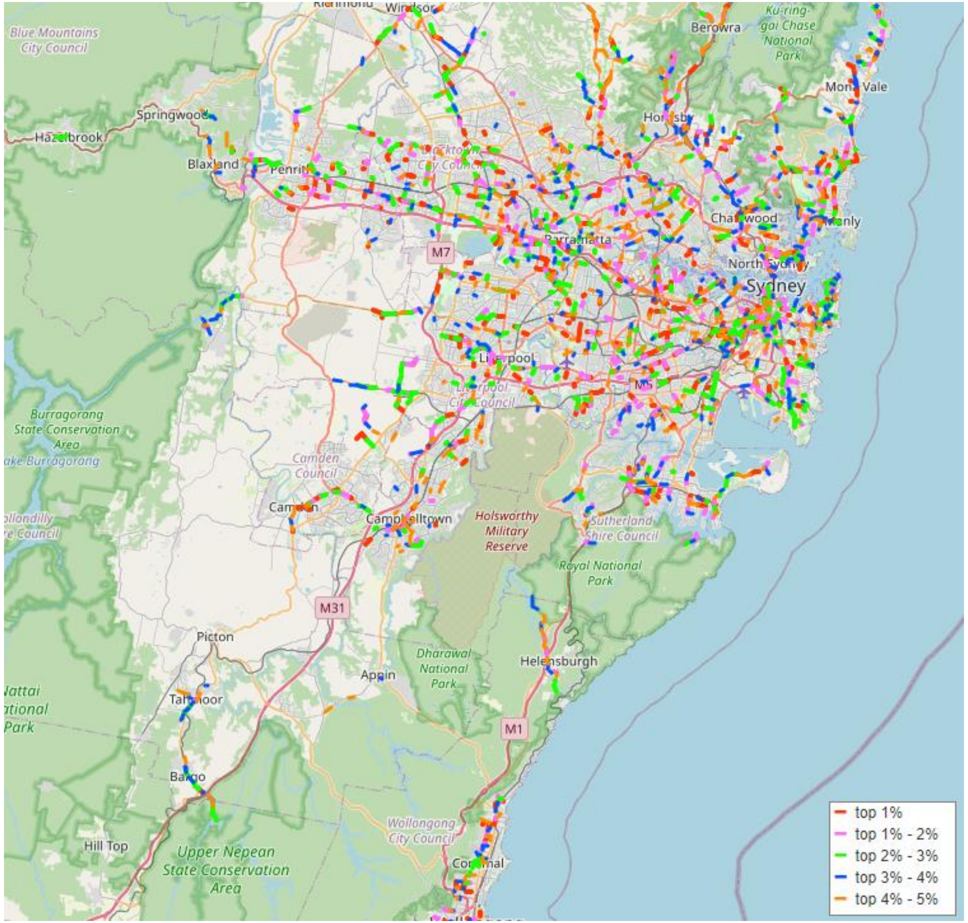




Water Main Failure Prediction

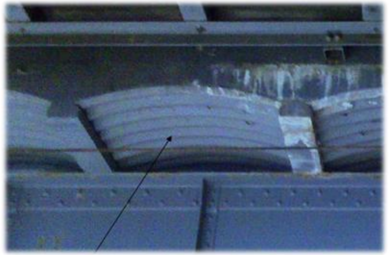






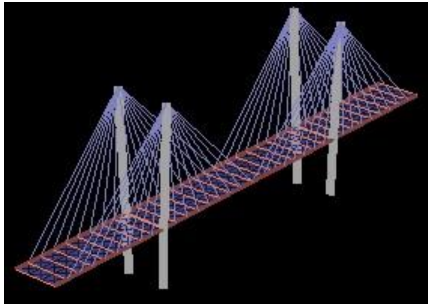
Structural Health Monitoring

- Problems:
  - Aging structure: monitor structural integrity of 800 steel and concrete jack arch supports (joints) under bridge deck
  - Current practice: visual inspection every 2 years and difficult access
- Needs:
  - Damage detection: time-based → condition-based monitoring
  - Monitor every one of 800 joints: efficient and reliable data management and analysis techniques

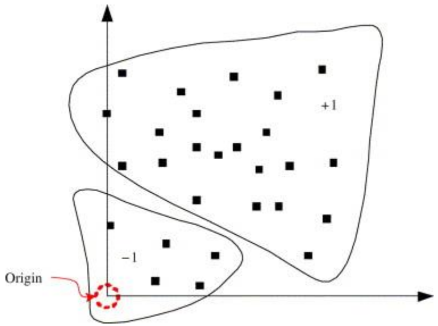


Damage Identification Techniques

- Model-driven vs data-driven approach
  - Numerical model may not be available or accurate, may not cater well to variation in environmental and operational conditions.
  - Data-driven approach establishes a model from data, using machine learning techniques.

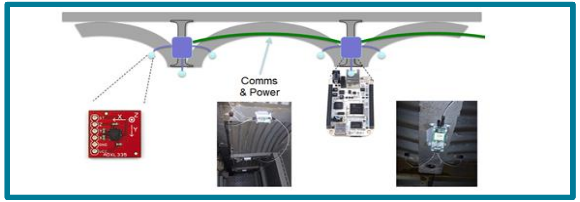


- Our approach using machine learning
  - Data corresponding to damage are often not available: a benchmark model is built using only healthy data.
  - New data not conforming with trained model are considered as damage.

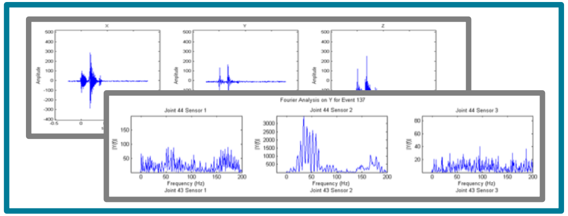


Machine Learning Flowchart

Excitation



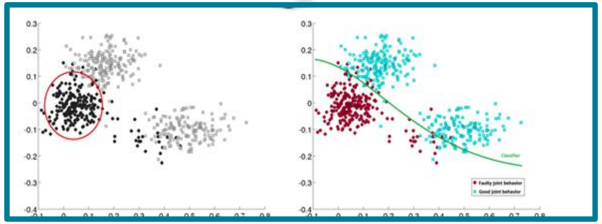
Data Acquisition



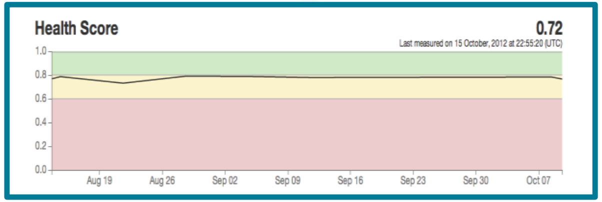
Feature Extraction

Train

Test



Machine Learning Analysis

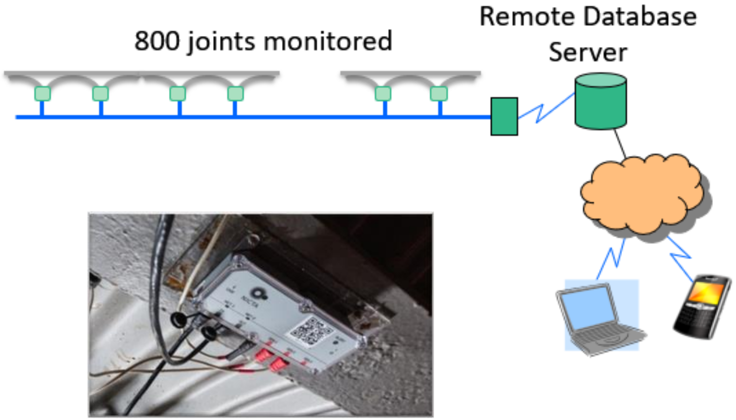


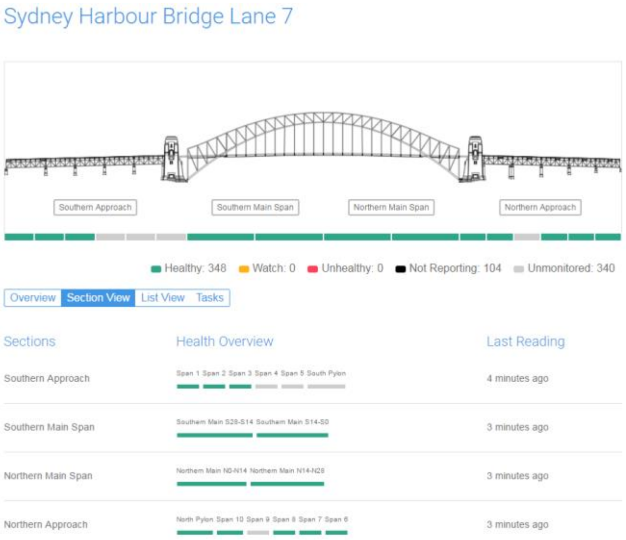
Structural Health Score

Decision Making

Structural health monitoring of SHB

- System: data acquisition system + data analytics + user dashboard
- Real-time and condition-based monitoring
- Data-driven machine learning technique for damage detection





Track Inspection Process

1. MTP vehicles inspecting the track

- Whole network inspection every 2 weeks
- Captures images via easement, track and rail cameras

Inspection Data

- Content:
  - Inspection images
  - Image GPS positions
- Data for this project:
  - 3 inspections:
  - 7 sessions for city circle
  - Each inspection has about **2M** images/3.5TB

2. Inspectors reviewing images back in the office

- One inspector reviewing 30-40 km/day on average
- Identify points of interest (POIs) for on-ground inspection and repair

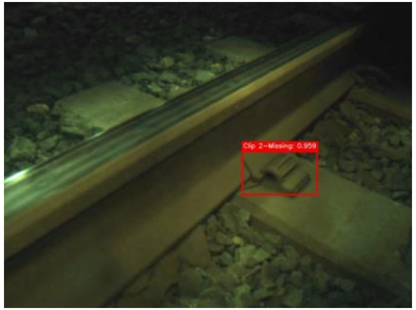
POI Data

- Content:
  - POI table
  - POI images
- Data for this project:
  - Total POIs: 243,031
  - Defects (POIs with external ID): 3,270
  - POI images: 64,403
  - Defect images: 13,562
  - Data size: about 10GB

3. On-ground inspection and repair

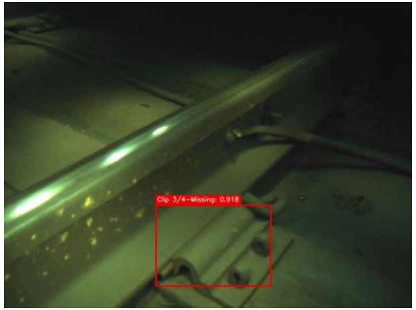
- On-ground technicians performs further inspection and repair

Different Types of Clip Defects



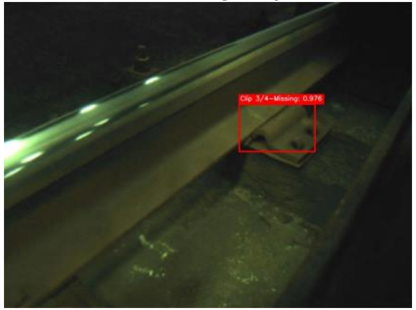
Missing clip

Missing clip

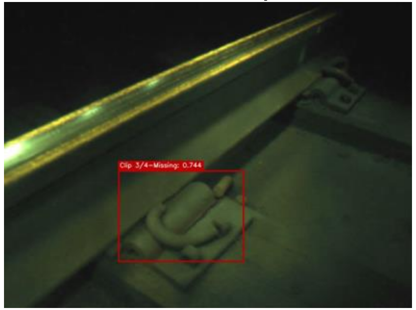


Loose clip

Missing clip

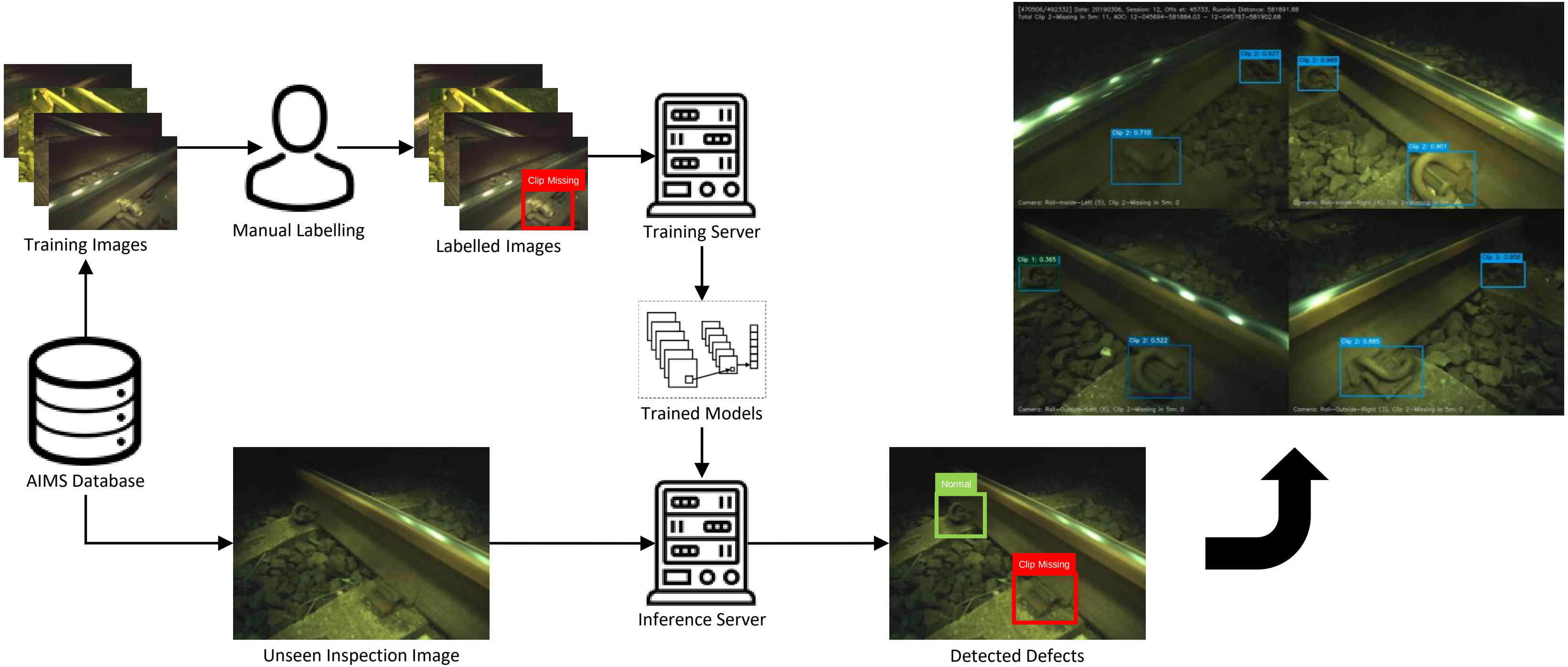


Missing clip



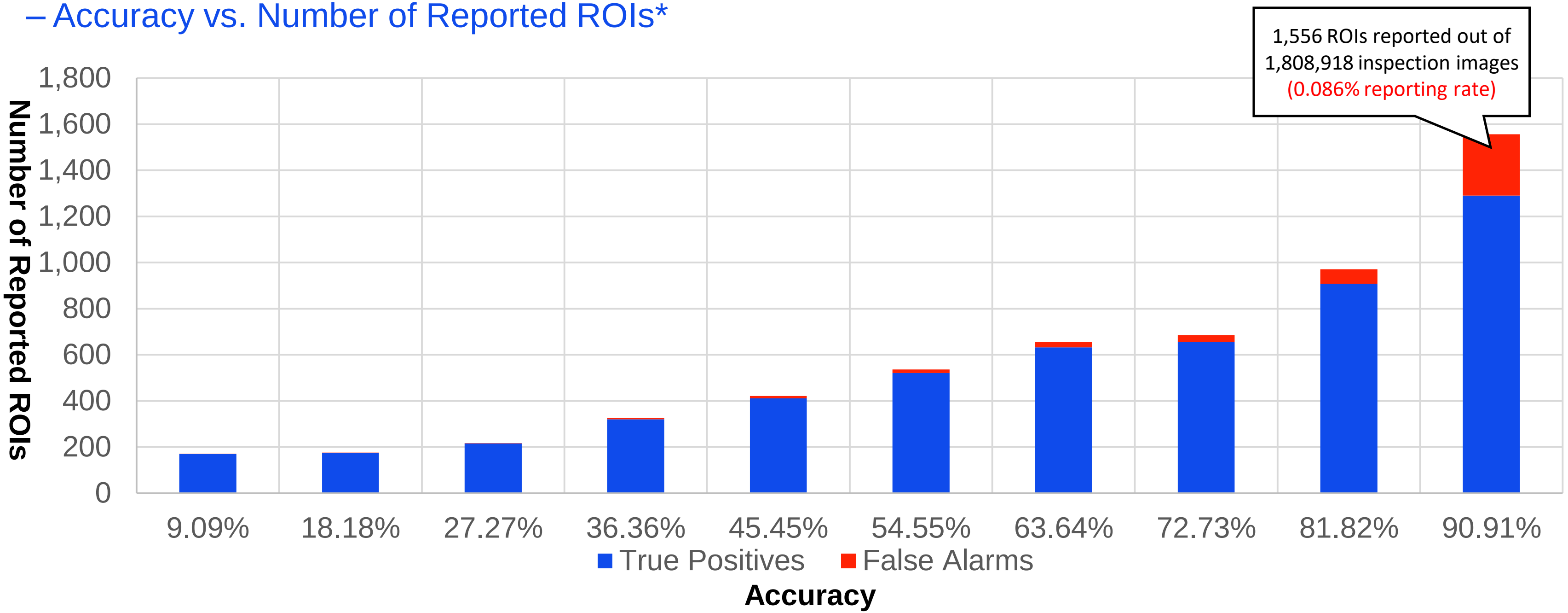
Broken clip

Machine Learning System for Defect Detection



Detecting Missing Clips on 03/10/2018 Inspection

– Accuracy vs. Number of Reported ROIs*



* On average, 3 to 4 ROIs correspond to 1 unique clip as the clip might be captured multiple times on inspection images.

Data Driven Solutions


Road & bridge

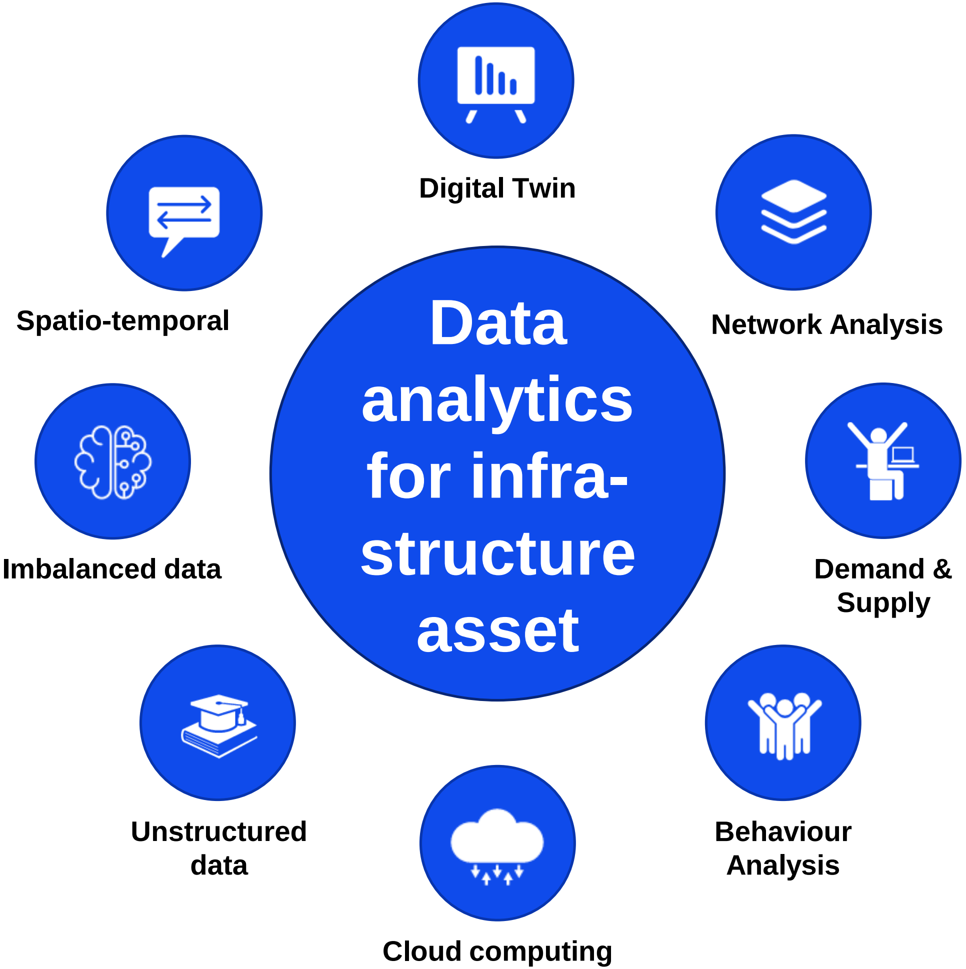

Water & Sewer


Power facilities


Engines/Transformers


Manufacture/Hardware


Finance



Thank you

Data-driven Approaches