National Edge AI Hub

Artificial Intelligence Theme of Edge AI Hub

Energy and Resource Efficient Artificial Intelligence

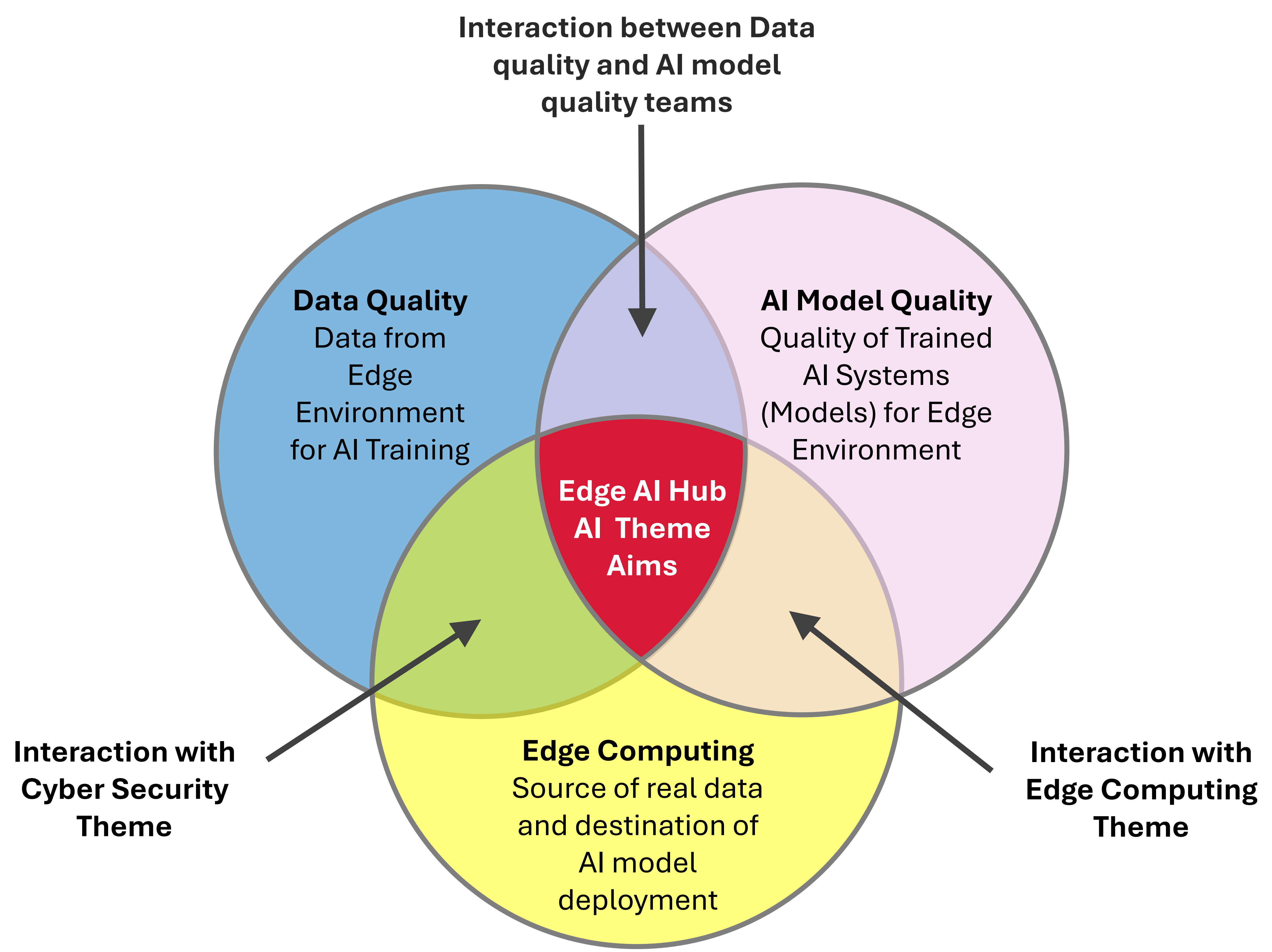
Dr Varun Ojha

Edge AI Hub, Newcastle University

@

APRIL Hub, The University of Edinburgh

22 November 2024


Interaction between Data quality and AI model quality teams
Data Quality
Data from Edge Environment for AI Training
AI Model Quality
Quality of Trained AI Systems (Models) for Edge Environment
Edge AI Hub AI Theme Aims
Edge Computing
Source of real data and destination of AI model deployment
Interaction with Cyber Security Theme
Interaction with Edge Computing Theme

National Edge AI Hub

AI Theme Challenges / Research Aims

- **Monitoring of Data/Model Quality**

  to monitor how cyber-disturbances impact age of data, AI algorithms learning quality and the overall application resilience?

- **Recovery of Data/Model Quality**

  to recover data and AI model quality that are impacted by cyber-disturbances and ensure suitability for AI model deployment on devices at Tiers 1, 2 of EC architectures ?

- **Assurance of Continuity of Data/Model Quality**

  to assure AI algorithms continually adapt to EC environments where unknown cyber-disturbances that were not presented in the original training dataset?

Potential Research Problems

- Monitoring
  - **RP1.** Investigate, characterise, and develop ontologies of data challenges and AI model challenges for edge computing environment.
  - **RP2.** Data and model quality assurance to data quality challenges such as faults, missing data, hardware failure, sensor degradation; diverse data source; sensor/data heterogeneity.
- Recovery
  - **RP3.** Investigate and develop data and model quality certification/robustness to various challenges such as data distribution shift, impurities, adversarial attacks, hardware resources limitations, etc.
  - **RP4.** Investigate the model quality certification/robustness to cyber disturbances, cyber-attacks, on federated/distributed EC environment.
- Assurance
  - **RP5.** Identify quality issues with AI model implementation on edge and offer mitigation strategies to resolve the challenges for ensuring model quality continuity.

Our Smart City Testbench

Newcastle University's Urban Observatory Sensors

>£8 million pounds (Capital investment)

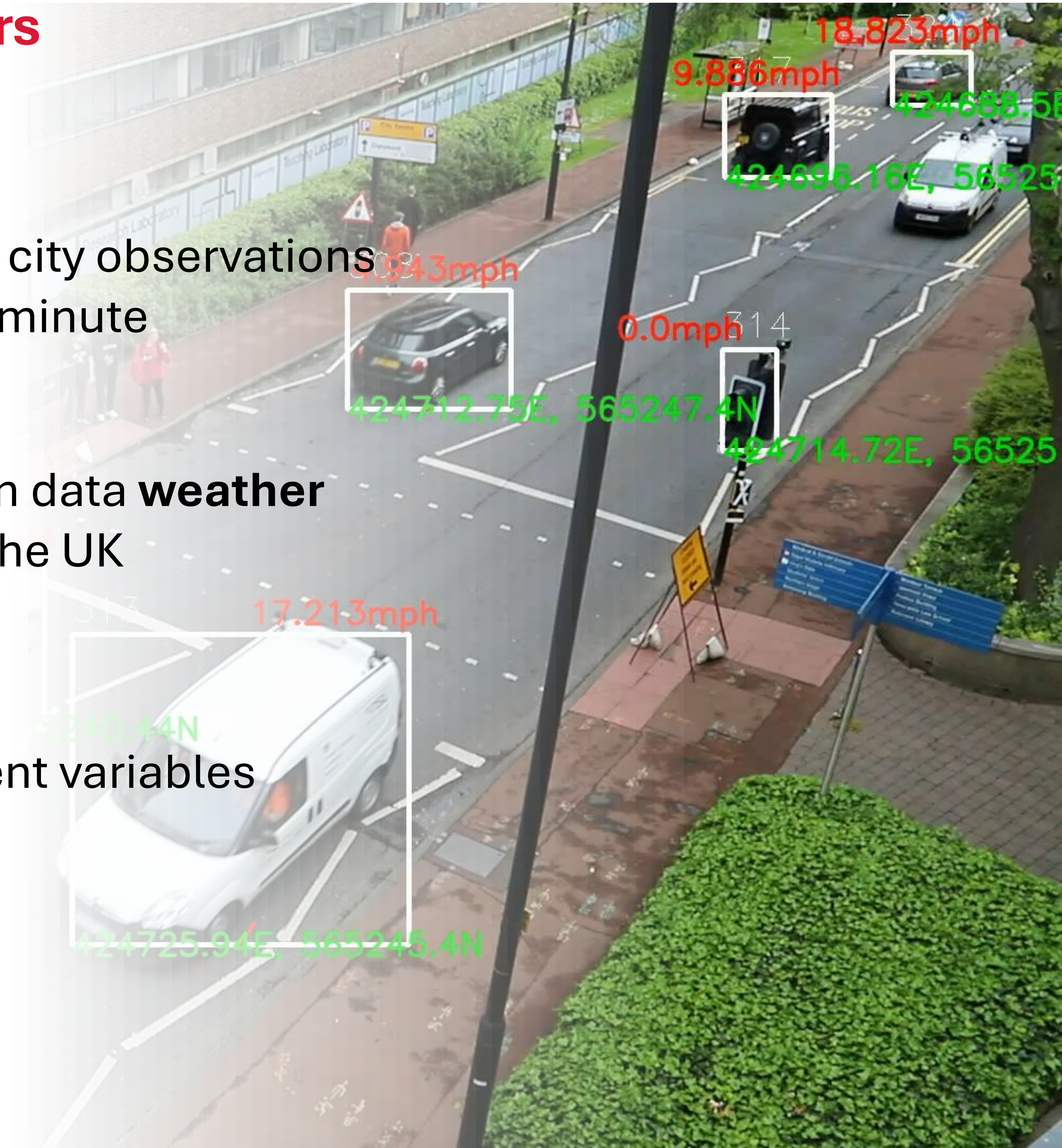
10 billion city observations 10,000 a minute

Billions of smart building observations

Only open data **weather radar** in the UK

CCTV: **500** views, **500m+** images, 24 real-time feeds

65 different variables



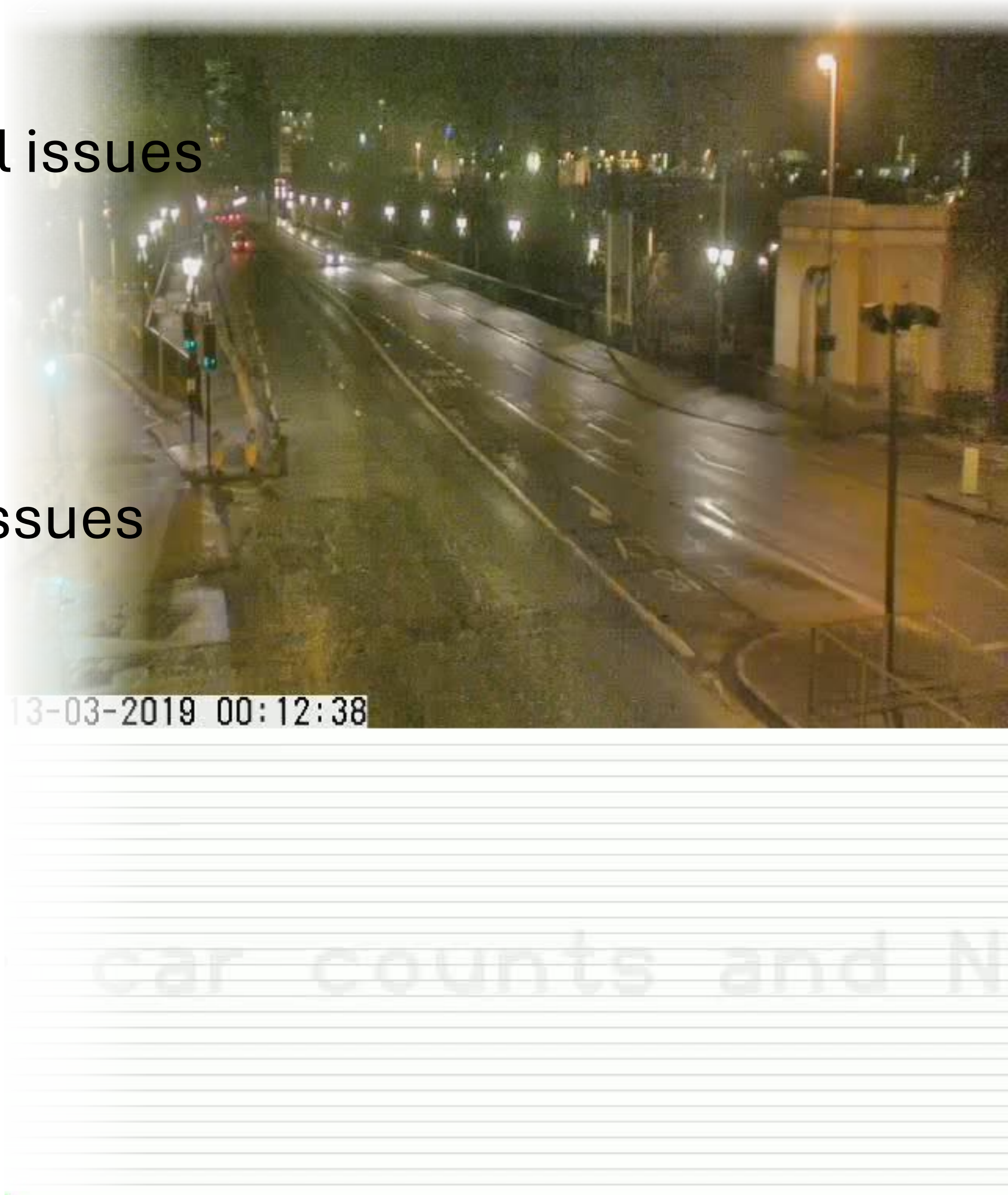
Source: Phil et al (Newcastle)

National Edge AI Hub

National Edge AI Hub

Our Experience with Data Quality Challenges

- **Data quality**
  - degradation of sensors over time
  - anomalous values, random spikes, or environmental issues
  - data out of range, out distribution, uncertainty
- **Data stream issues**
  - data retrieval - source API failure
  - source API failure, network failure, network overload
  - system throughput - queues building up, hardware issues
- **Cyber security**
  - adversarial attacks
  - denial of services, spoofing
- **Failure**
  - hardware failure at sensor



Source: Phil et al (Newcastle)

National Edge AI Hub

Data Certification (Safe ML) – Example Solution

Trusted dataset for AI model training

Our Solutions:

- **D-ACE** – a framework for certifying training datasets using various characteristics
- **SafeML** – a framework for safety monitoring of ML models at run time

We will extend these to Edge AI

- D-ACE for certifying datasets in federated Edge AI architecture
- Safety of Federated Learning algorithms in EdgeAI architecture

Expectation
AI model training data

Reality
data in reality for testing AI model

Source: Thakker et al (Hull)

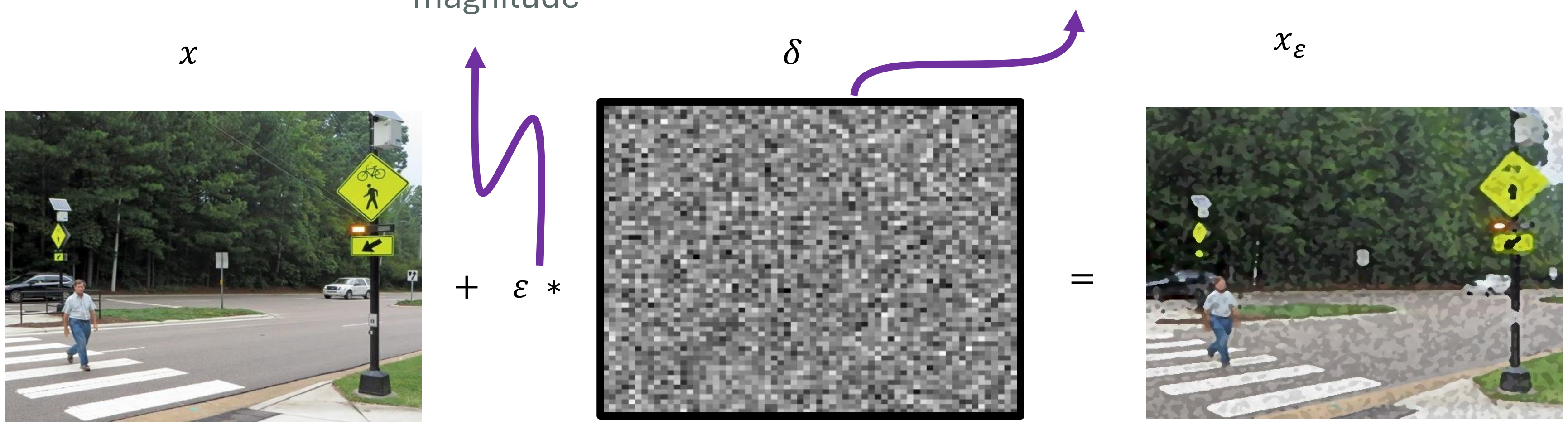
Model Certification – Example Solution

Models adversarial Attacks Mitigation in Autonomous Vehicle and Vehicle to Everting Transportation Communication Scenario

Calculated using Deep Neural Networks (DNNs) weights (white-box attack)

Perturbation magnitude

x δ x_{ε}



$+ \ \varepsilon \ * \qquad = $

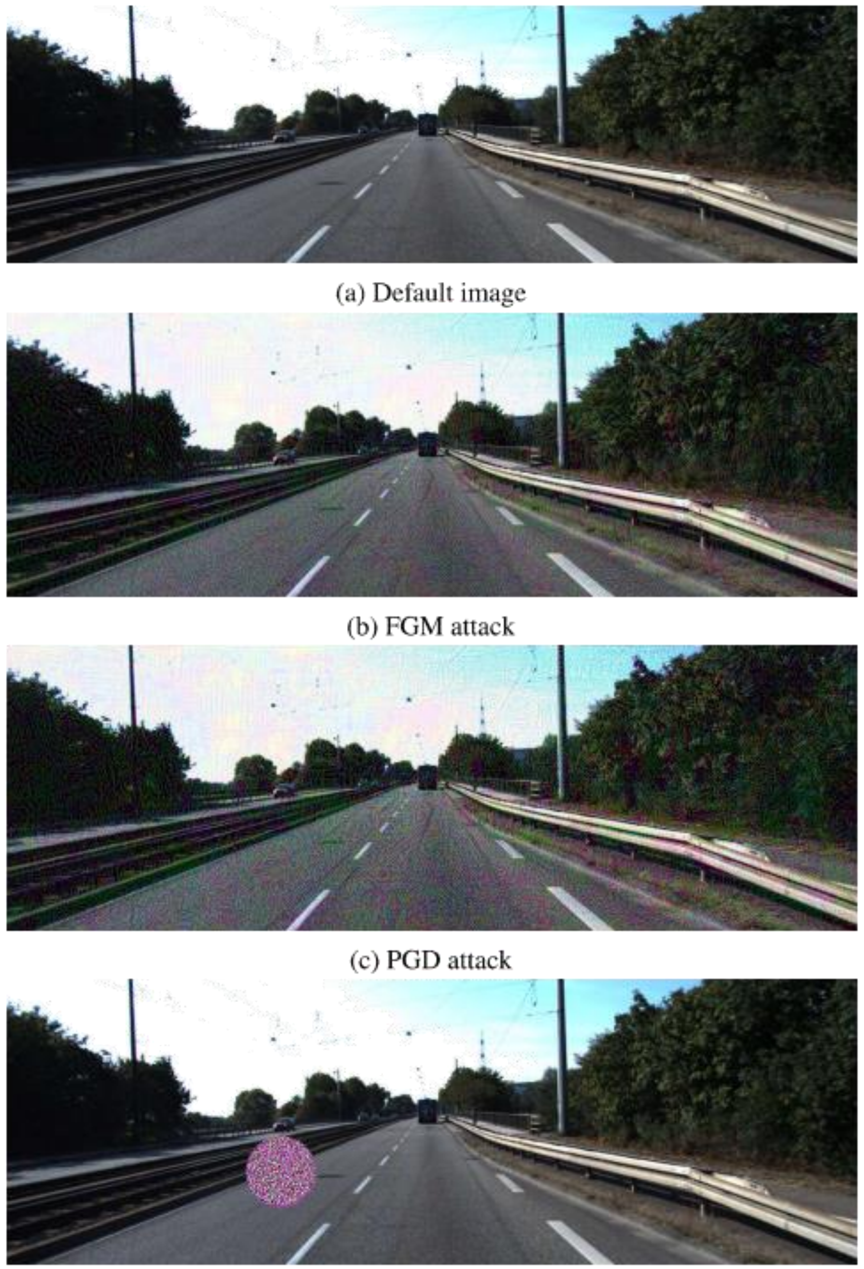
Input example Predicted as ‘***Pedestrian***’

Adversarial perturbation

Adversarial example Predicted as ‘***Speed limit change***’

One of objectives of the AI Model Quality analysis is to subject AI model to the ‘**worst case conditions**’ (such as adversarial cyber/attacks) and evaluate the ***ability for a model to remain invariant*** under such settings.

Source: Ojha et al (Newcastle)



(a) Default image

(b) FGM attack

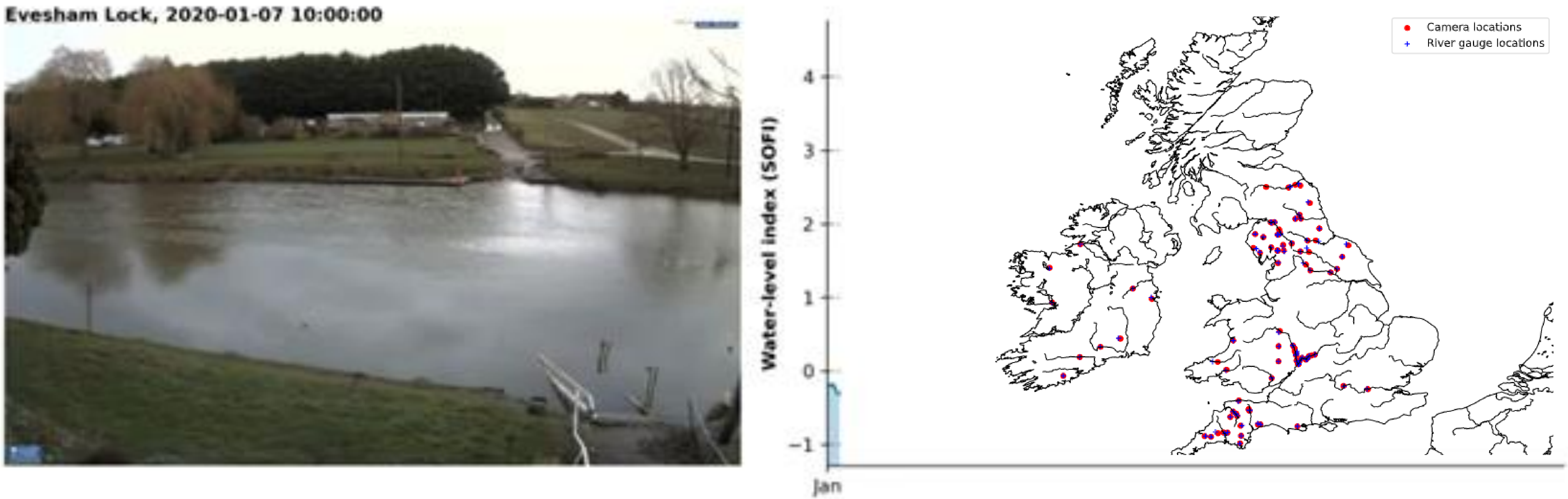
(c) PGD attack

(d) AP attack

Edge AI for Flood Tracking and Monitoring

Fusion of Environmental Agency Data Edge Data (CCTV Cameras) across UK & Ireland

Our research help automat tracking and monitoring of flood saturation

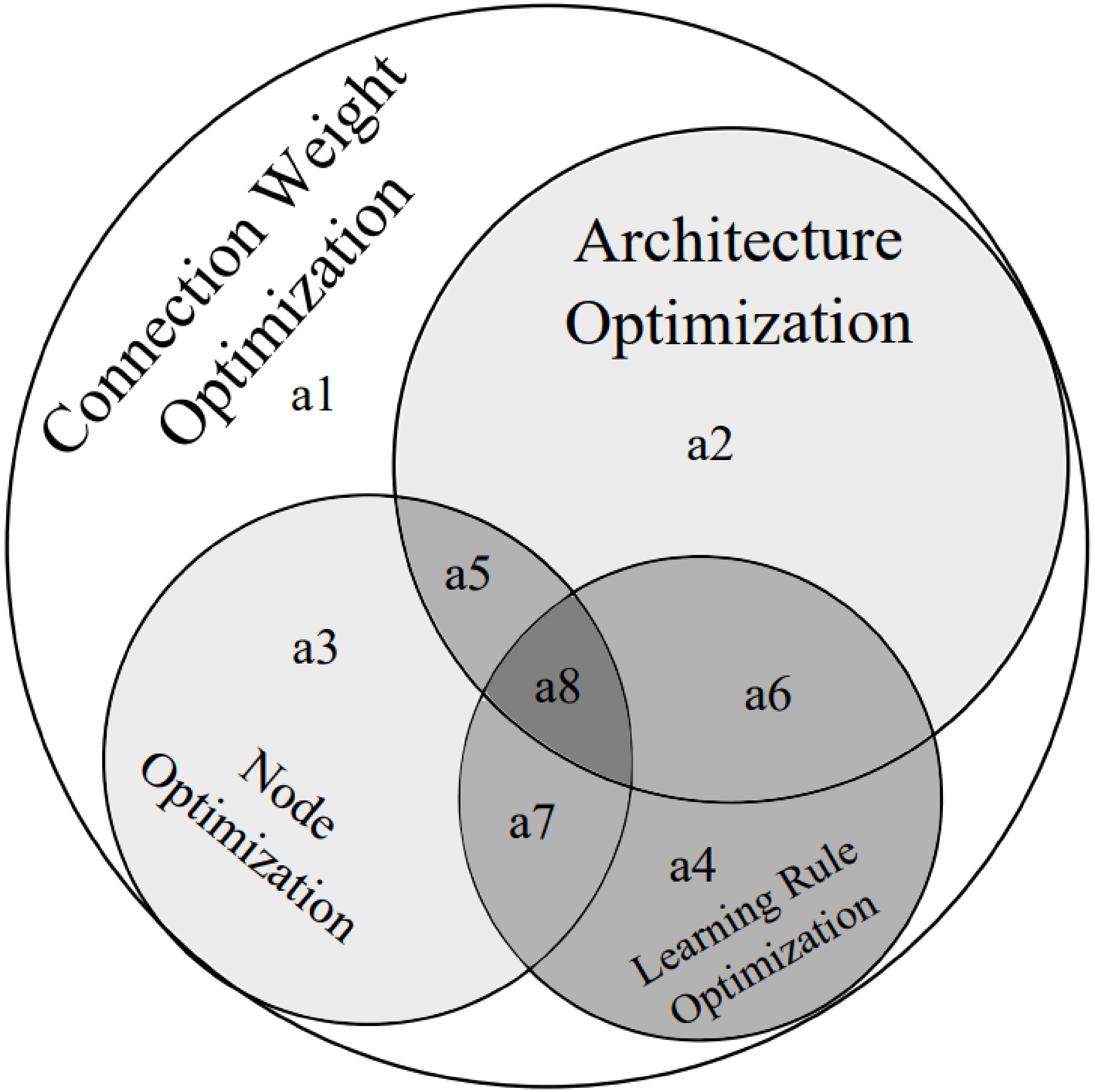


We achieve 94% accuracy in correctly predicting real flood events The River Avon and River Severn.

Source: Ojha et al (Newcastle)

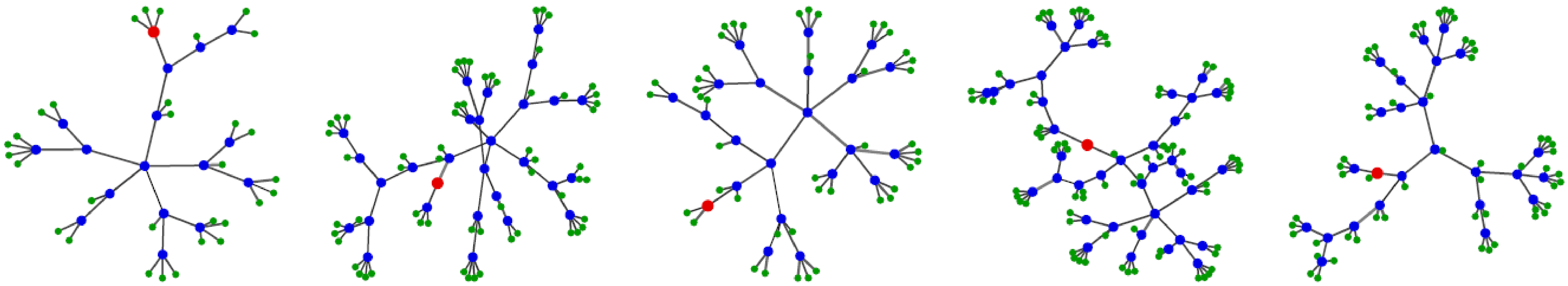
Energy Efficient AI Models

(Ojha et al.)



Backpropagation neural tree: Performance on regression

Regression results



(a) baseball (.85, 48) (b) dee (.89, 89) (c) diabetese (.63, 67) (d) friedman (.95, 116) (e) mpg6 (.9, 82)

Algorithm	Bas	Dee	Dia	Frd	Mpg	Avg Acc	Avg Weights
BNeuralT	0.665	0.837	0.492	0.776	0.867	**0.727**	152
MLP	0.721	0.829	0.49	0.943	0.874	0.772	**1041**

Ojha and Nicosia (2022), *Neural Networks*

Backpropagation neural tree: Performance on regression

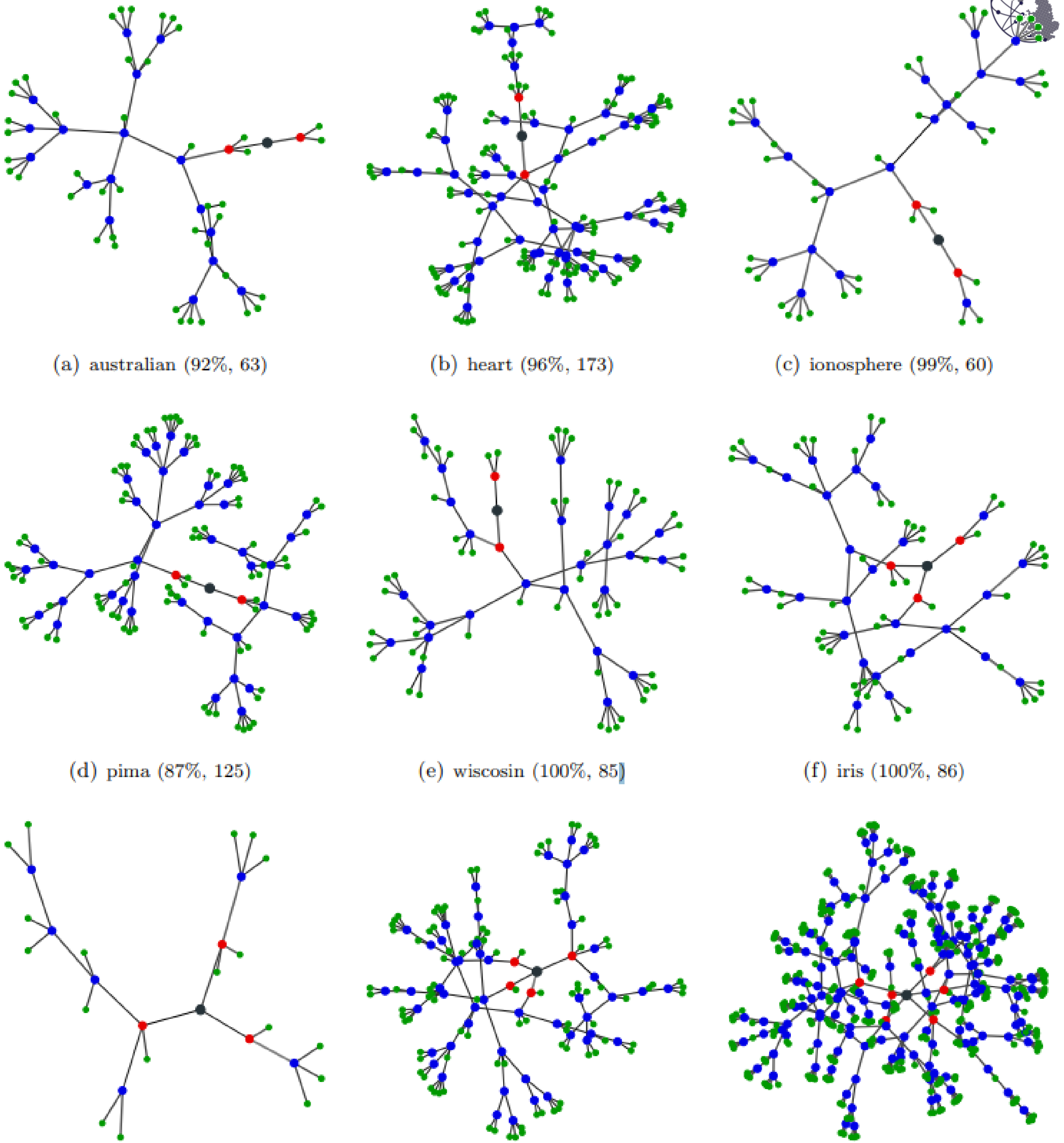
Regression results

- Neural Trees use only 14.6% of MLP parameters
- Accuracy differs only 5.8% lower than the best MLP result

Backpropagation neural tree: Performance on Classification

Classification results.

Data	BNeuralT	MLP
Aus	0.895	0.876
Hrt	0.897	0.833
Ion	0.952	0.882
Pma	0.822	0.774
Wis	0.986	0.984
Irs	0.992	0.972
Win	0.991	0.991
Vhl	0.75	0.826
Gls	0.732	0.635
Avg. Accuracy	**0.891**	**0.863**
Avg. Weights	**261**	1969



(a) australian (92%, 63) (b) heart (96%, 173) (c) ionosphere (99%, 60)

(d) pima (87%, 125) (e) wiscosin (100%, 85) (f) iris (100%, 86)

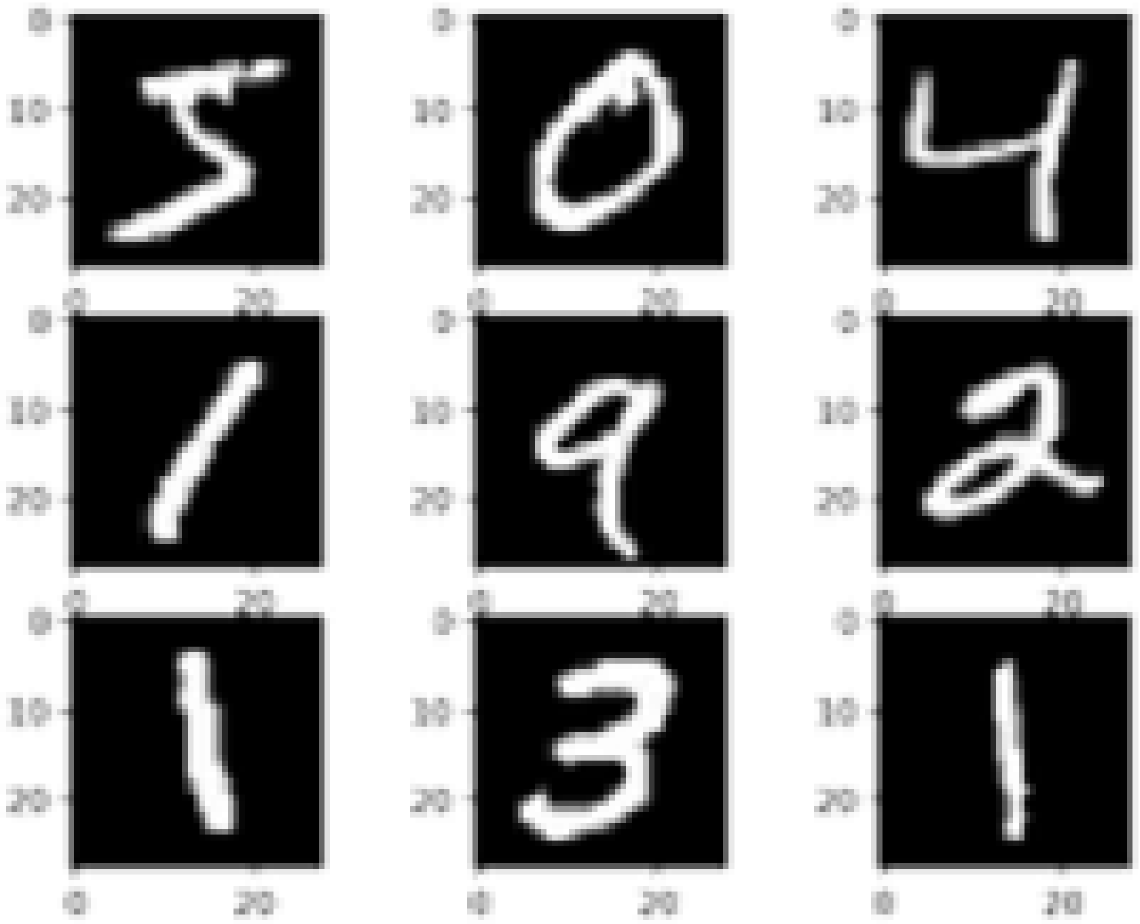
Ojha et al (2022), Neural Networks

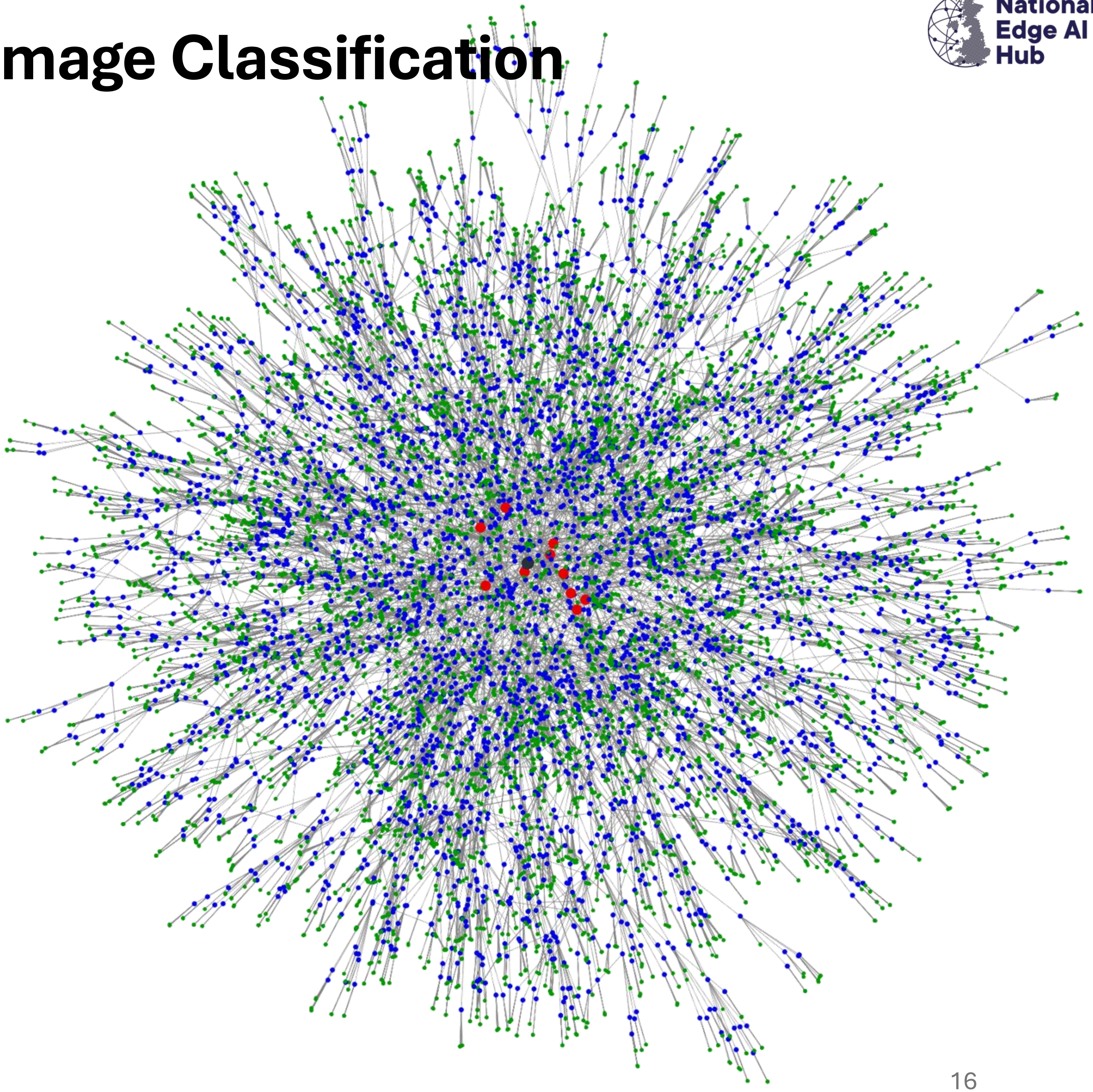
Backpropagation neural tree: Performance on classification

Classification results

- Neural Trees used **only 13.25% parameters** of MLP
- Accuracy is **2.65% better than the best MLP** result

Backpropagation neural tree: Image Classification

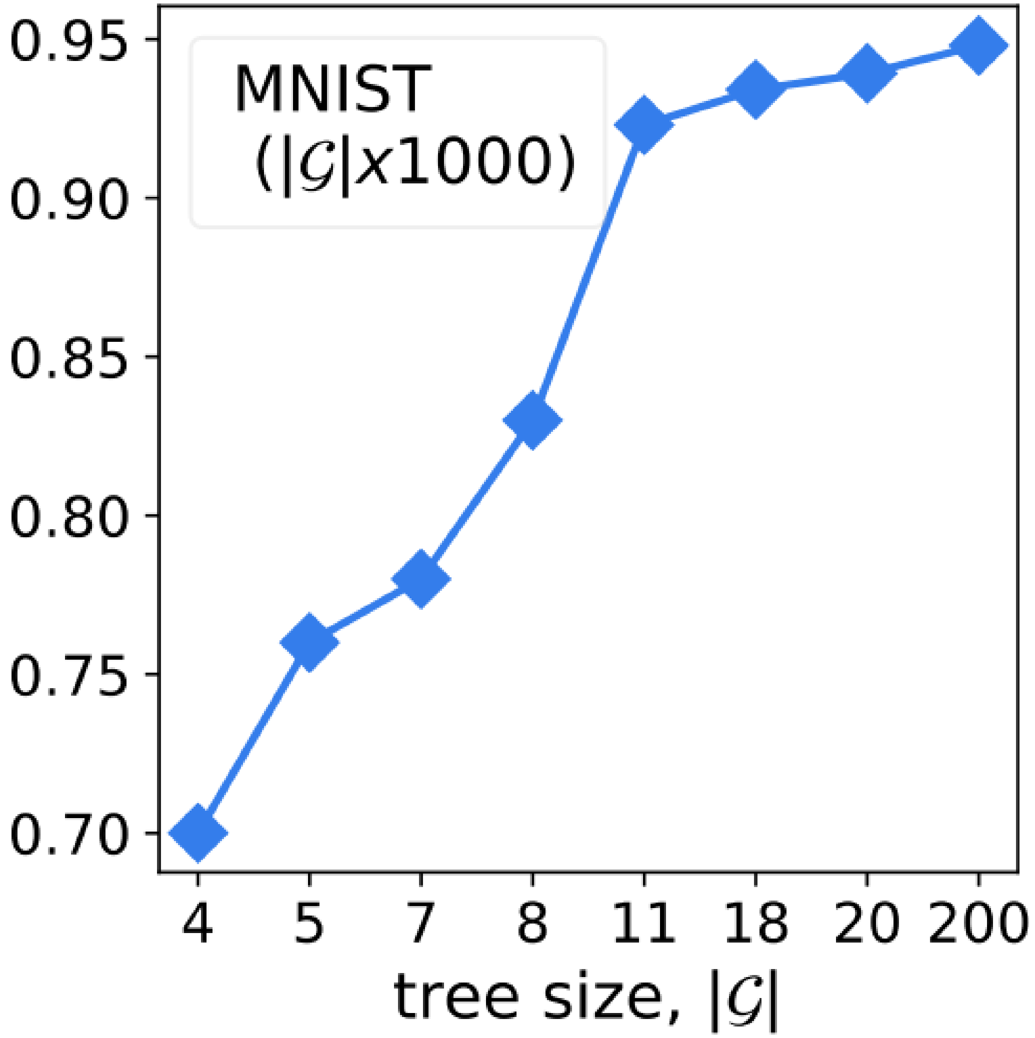




Ojha and Nicosia (2022), *Neural Networks*

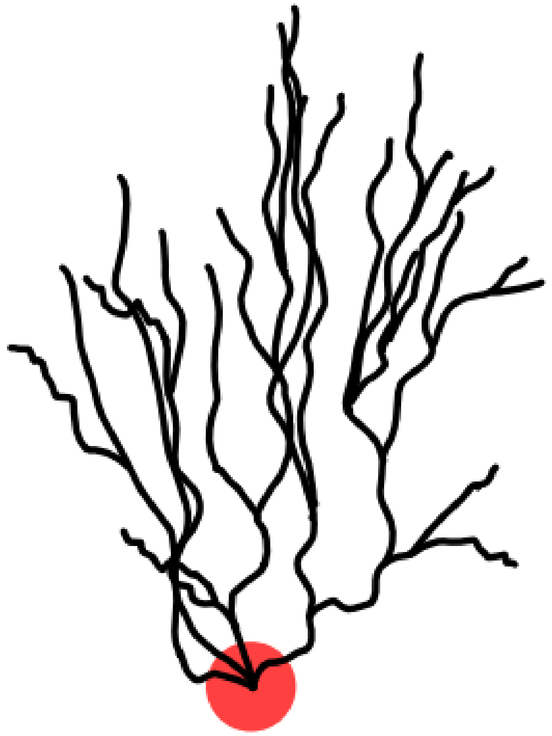
Model size vs accuracy

	Algorithms	Error(%)
BNeuralTs	BNeuralT-10K (pixels)	7.74
	BNeuralT-18K (pixels)	6.58
	BNeuralT-20K (pixels)	6.08
	BNeuralT-200K† (pixels)	**5.19**
Classification Trees	GUIDE (pixels, oblique split)	26.21
	OC1 (pixels, oblique split)	25.66
	GUIDE (pixels)	21.48
	CART-R (pixels)	11.97
	CART-P (pixels)	11.95
	C5.0 (pixels)	11.69
	TAO (pixels)	11.48
	TAO (pixels, oblique split)	5.26

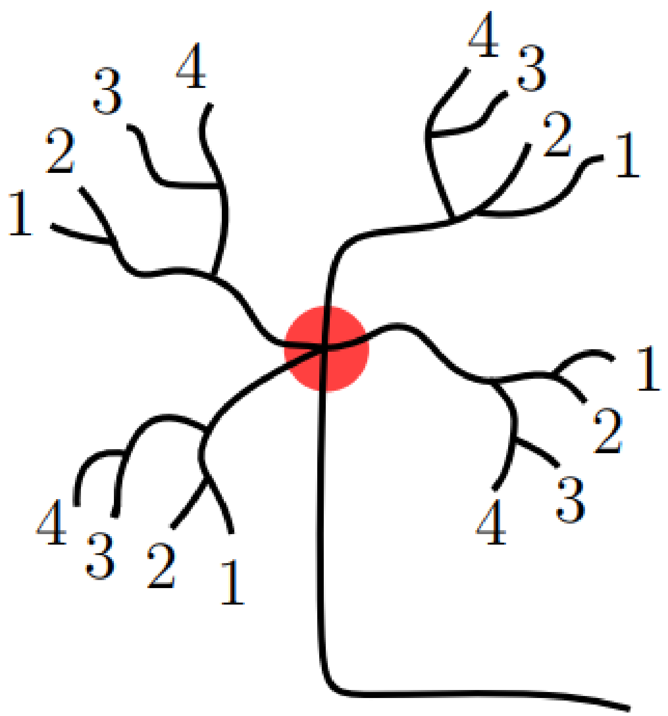


Single neuron

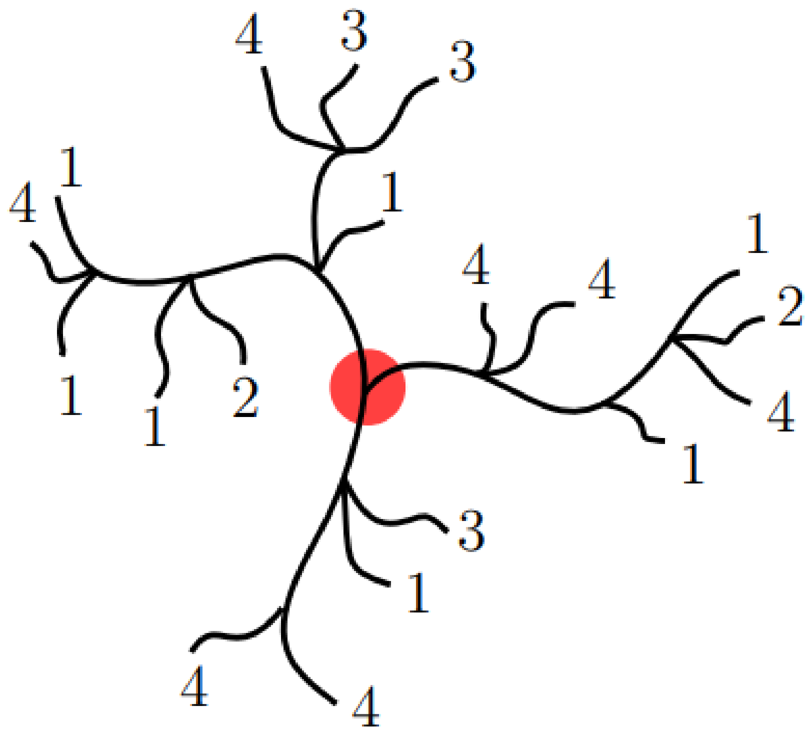
Biological Inspiration for Backpropagation Neural Trees



Travis et al. (2005)

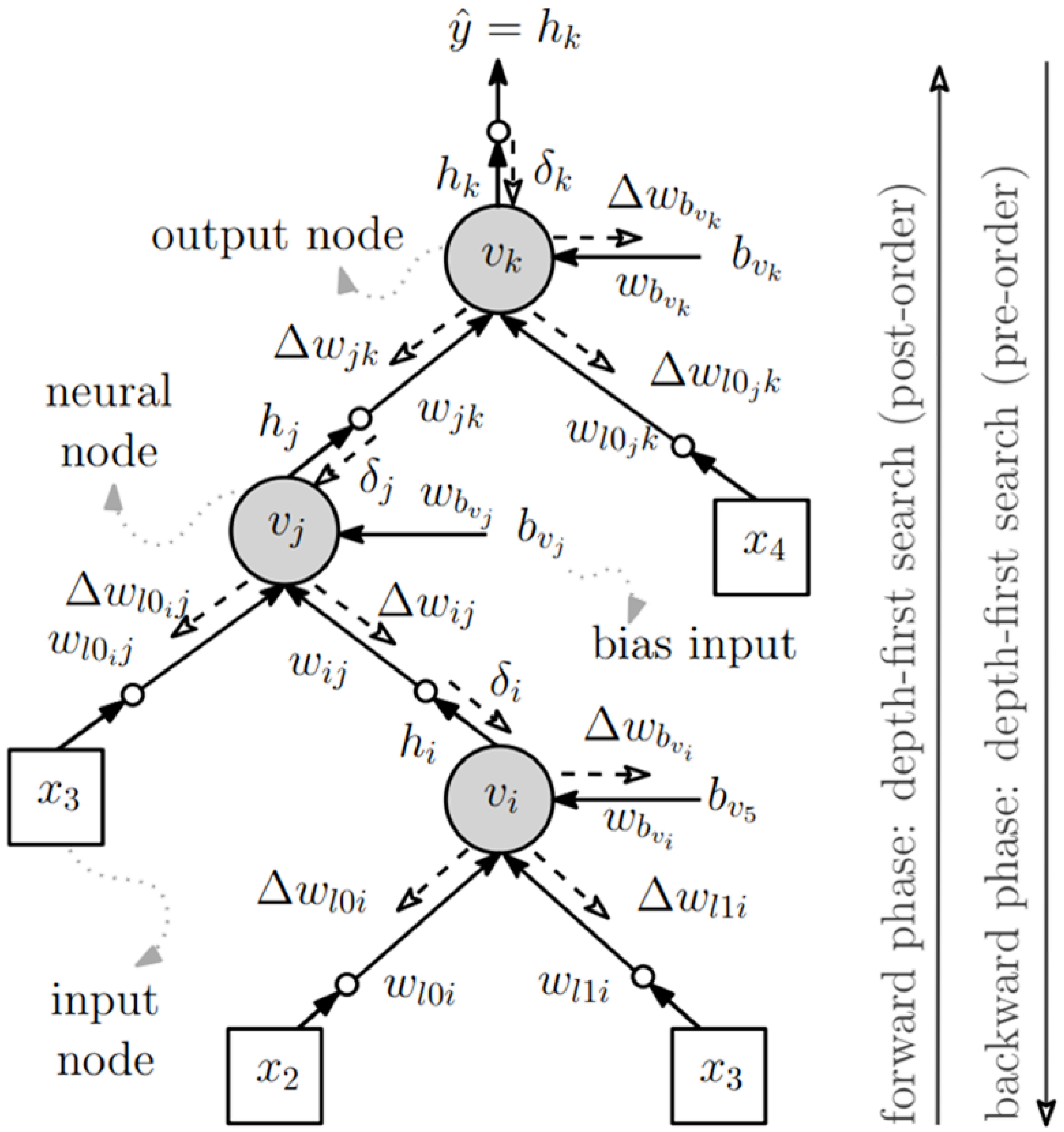


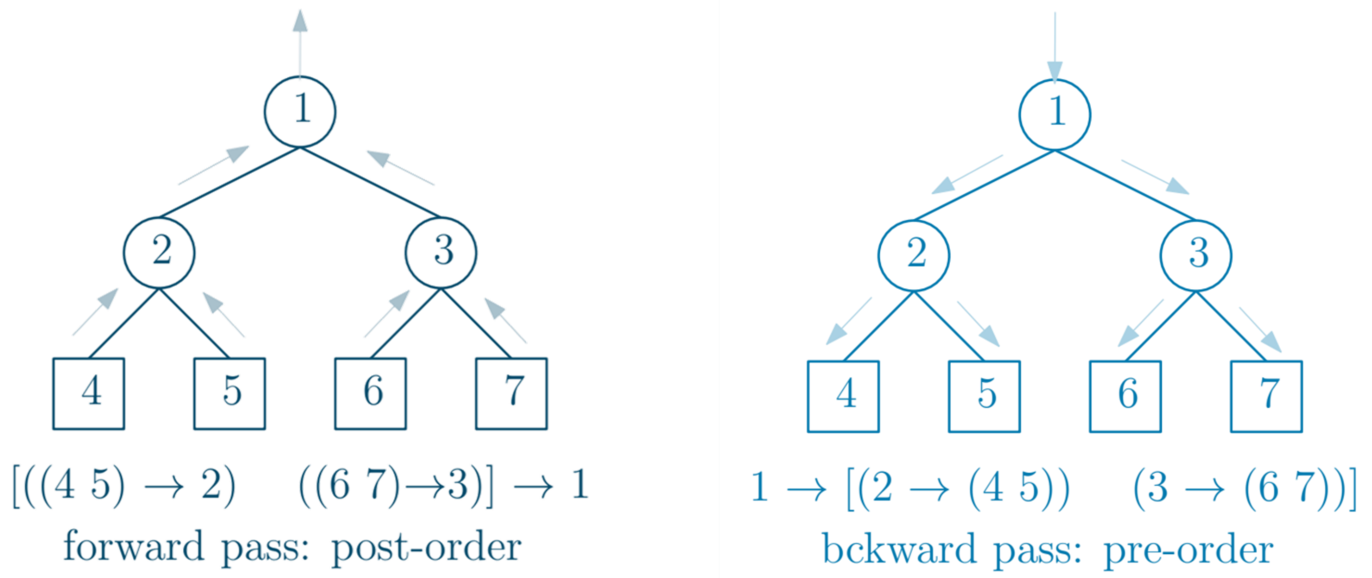
Jones and Kording (2021)

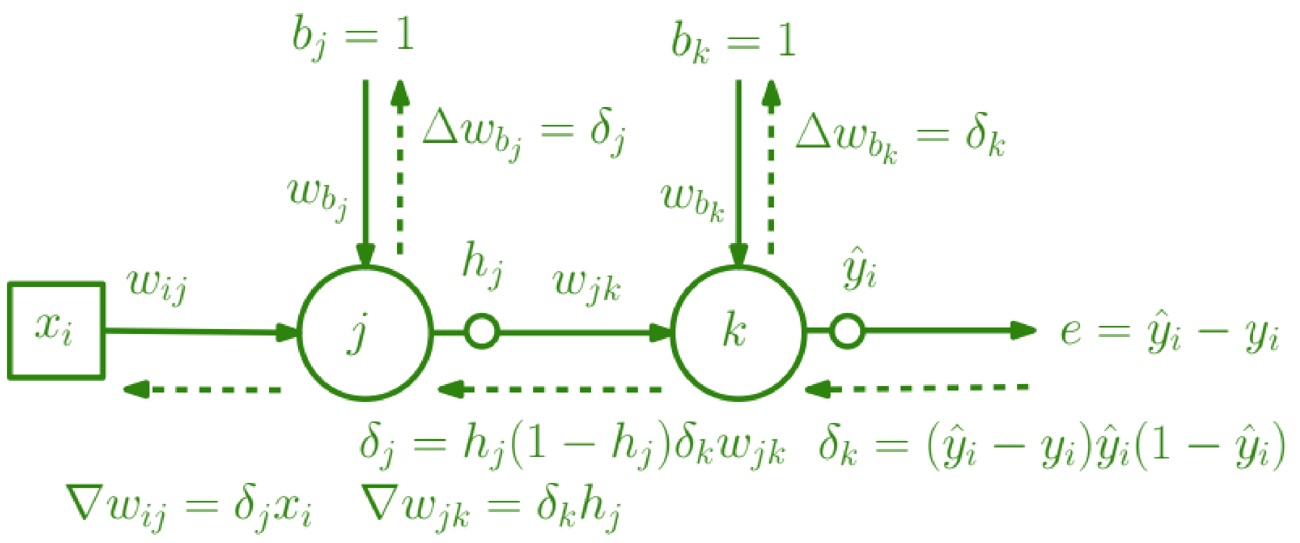


Ojha and Nicosia (2022)

Backpropagation neural tree

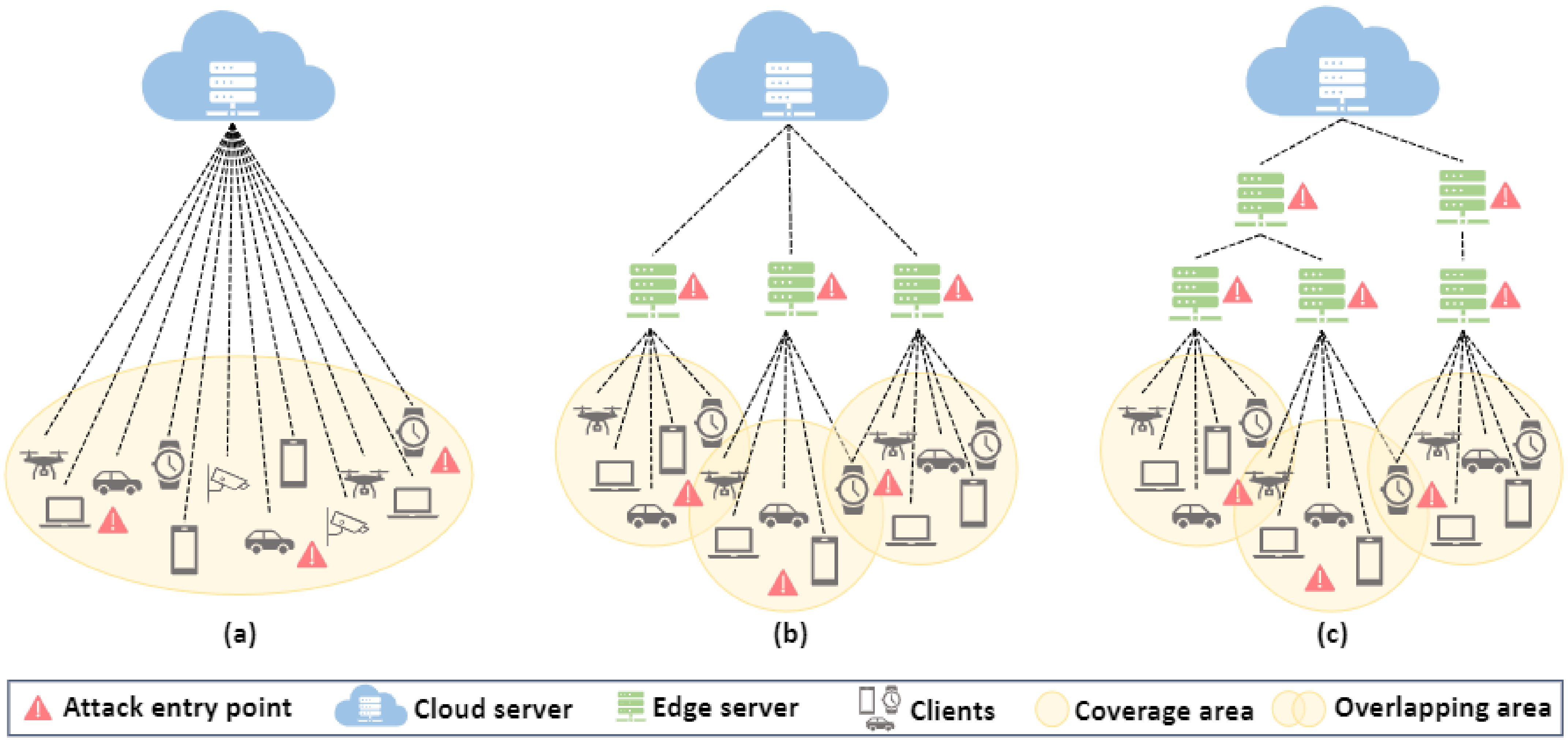






Federated leaning

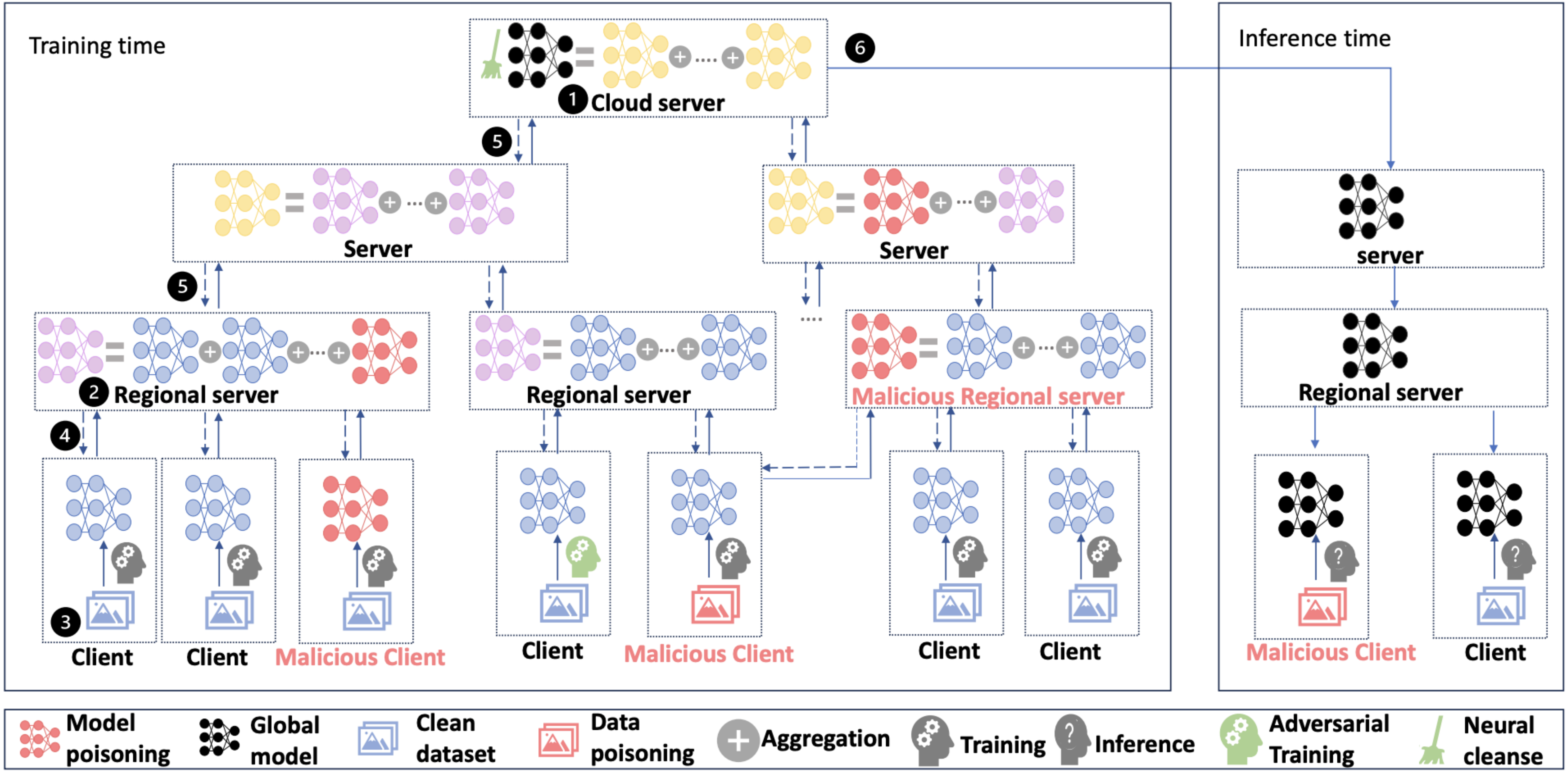
Leaning on user's edge devices rather than on cloud



FL network architectures: (a) 2-level FL; (b) 3-level HFL; (c) 4-level HFL

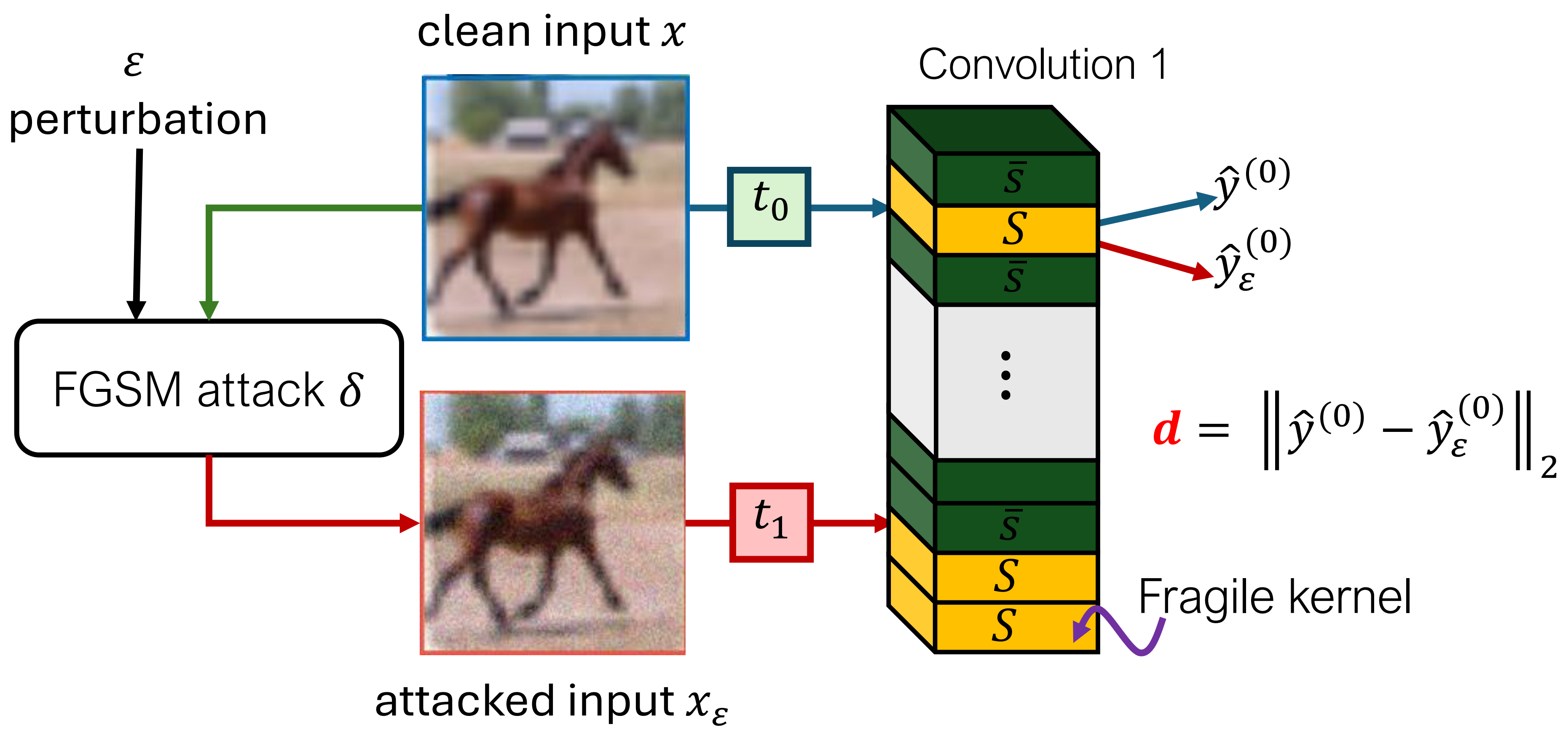
Alqattana et al. ICANN, (2024)

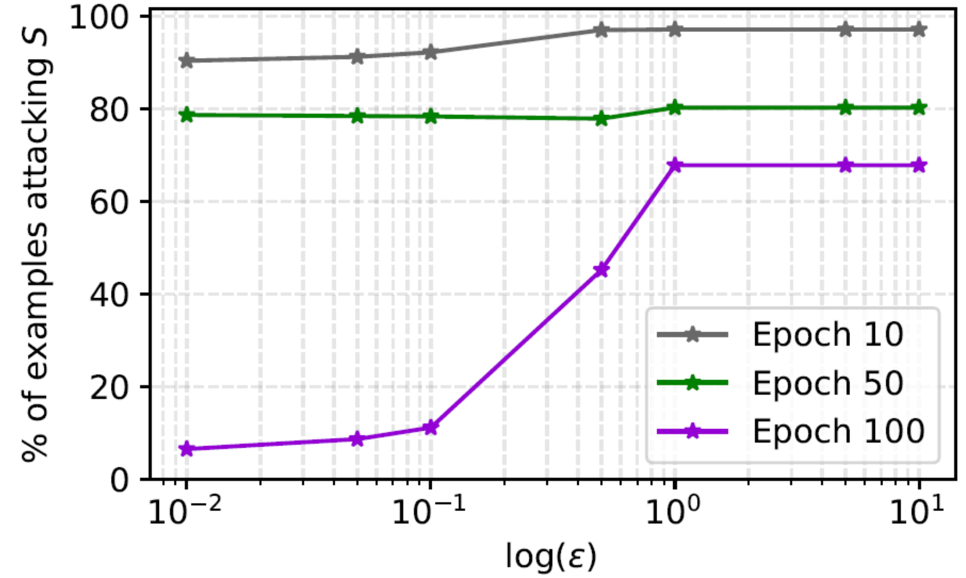
Attack on AI models in distributed systems



Attacks in individual (single)models

We measure the **average magnitude difference** $\boldsymbol{d}$ **at the output** of the first convolutional layer, between fragile and non-fragile neurons, on both clean and adversarial inputs.



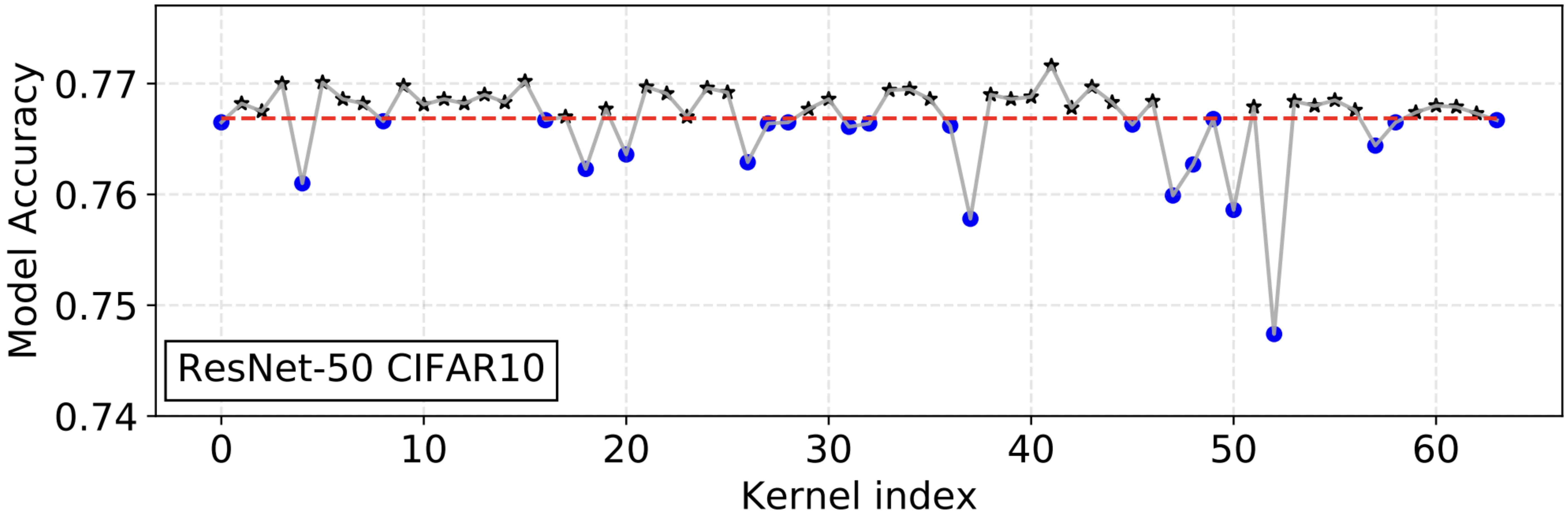


if avg. distance of fragile kernels S
greater than
avg. distance of non-fragile kernels $\bar{S}$
then
x_ε attacks fragile kernels

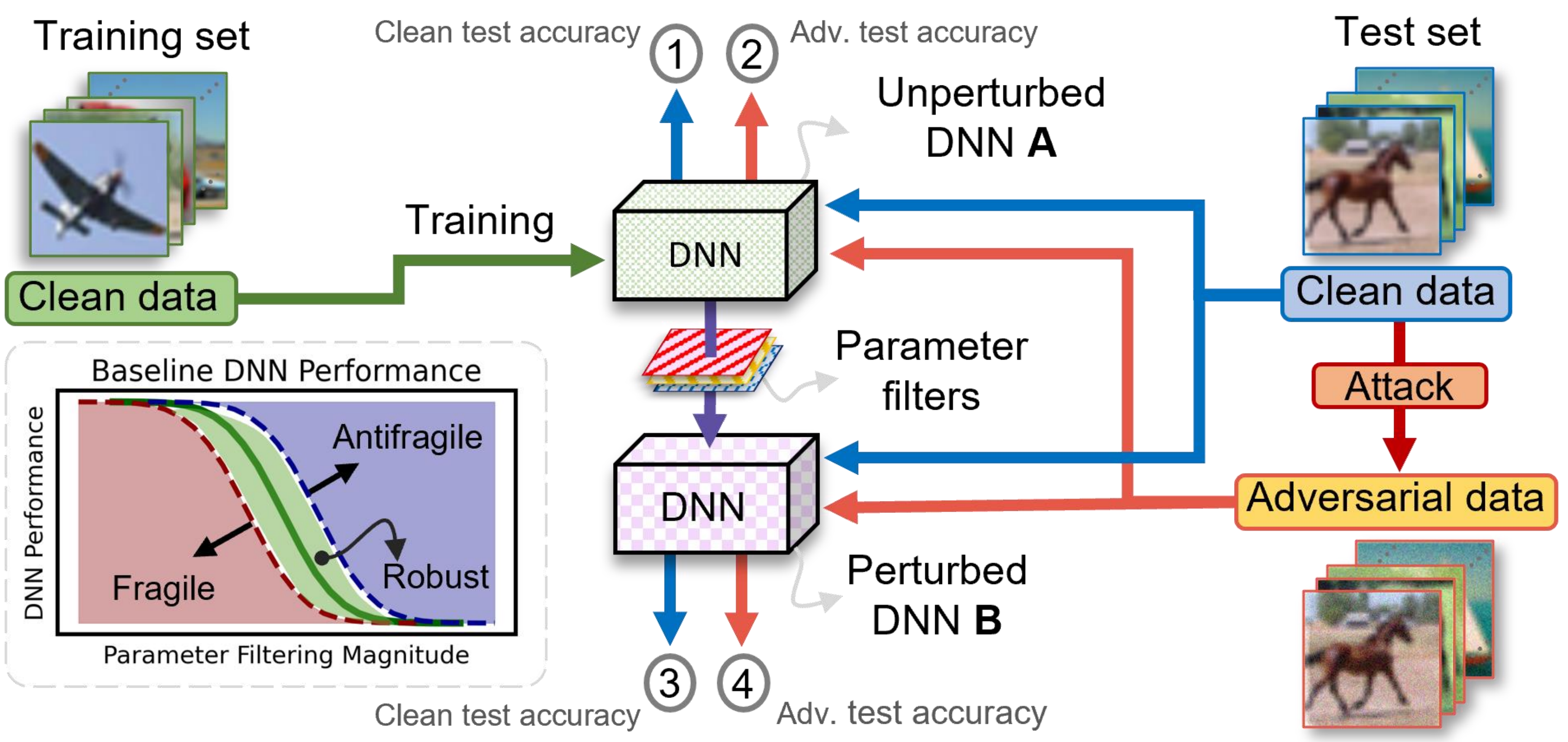
Ojha et al. ICANN, (2021)

Fragility of individual neural kernels

Fragile kernels shown in blue below mean performance line in red and null kernels S 0 are shown in black star above mean line in red



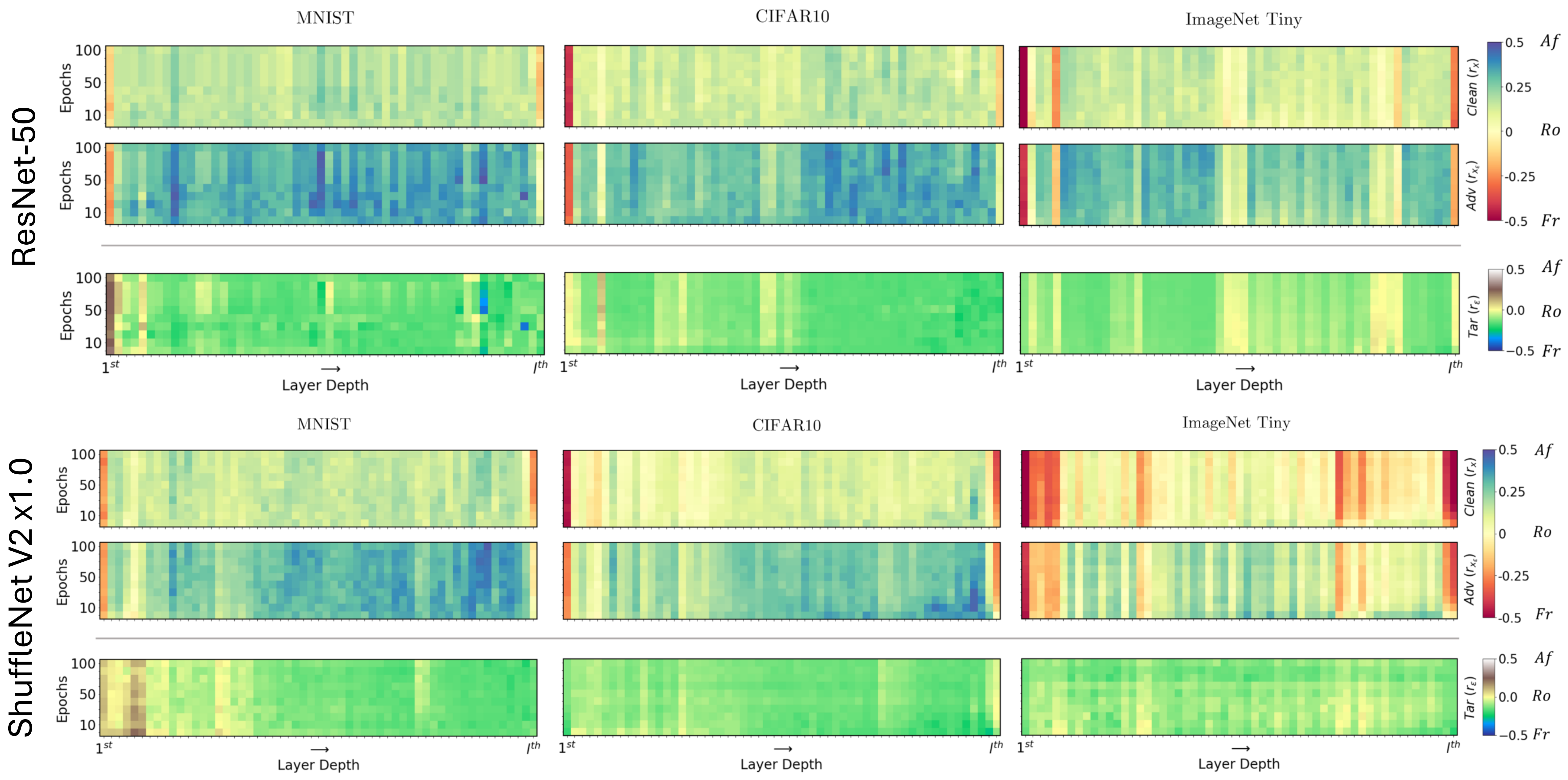
Fragility, robustness and antifragility



- a new method of parameter filtering (**synaptic filtering**)
- **synaptic filtering of all layers and parameters** of a DNN architecture.
- **compare clean and adversarial performance** of a regular DNN and perturbed DNN.
- **characterise** parameters as fragile, robust, and antifragile

Ojha et al. Artificial Intelligence, Elsevier. (2024)

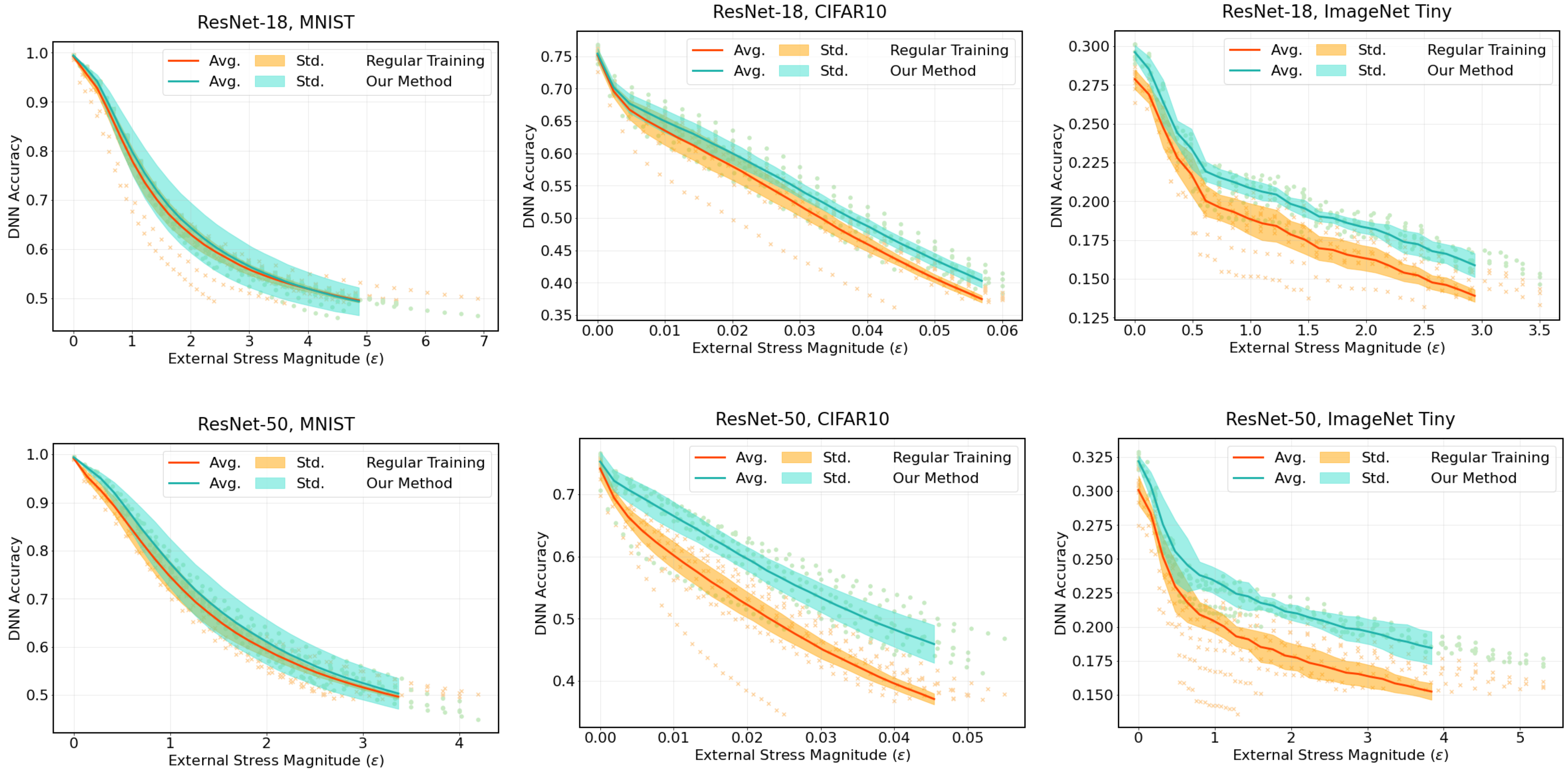
Robustness scores (layer-wise)



Periodic parameter characterisation shown for some networks.

We say that fragile parameters are important to network performance.

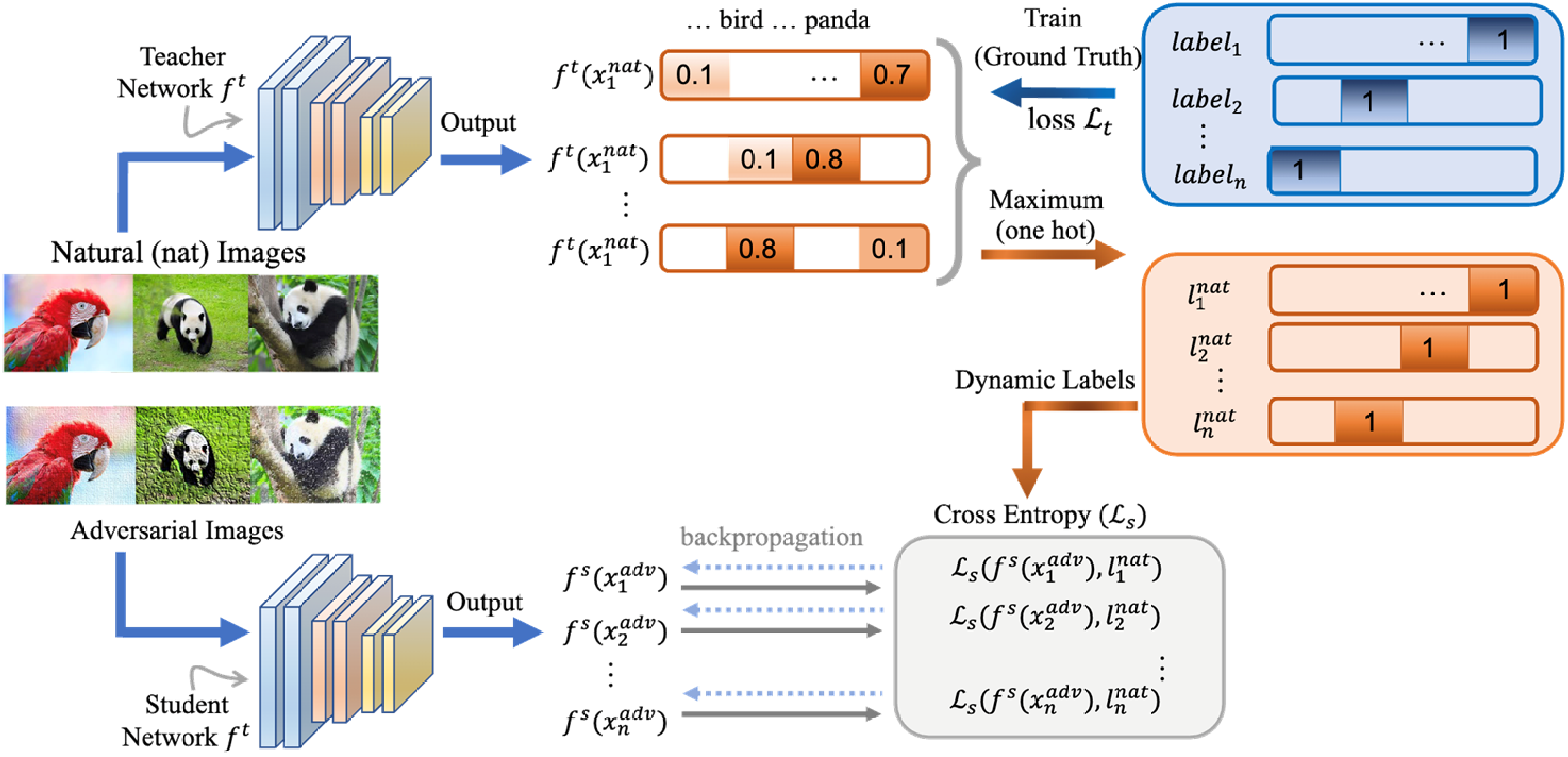
Making model robust against attacks



When we **retrain** networks at periodic intervals using only the characterised ***robust* and *antifragile*** layer parameters (**selective backpropagation**), we observe an **increase** in **adversarial performance**, and clean performance for some networks and datasets.

Securing an AI model against attacks

Dynamic Label Adversarial Training



Liu et al. ICONIP (2024)

National Edge AI Hub

References

- Fragility, Robustness and Antifragility in Deep Learning
  Artificial Intelligence, Elsevier. (2024)
  Pravin C, Martino I, Nicosia G, Ojha V
- Backpropagation neural tree
  Neural Networks (2022)
  Ojha, V., & Nicosia, G
- Adversarial robustness in deep learning: Attacks on fragile neurons
  30th Int. Conf. on Artificial Neural Net., ICANN (pp 16-28), Springer, LNCS, Bratislava (2021)
  Pravin C, Martino I, Nicosia G, Ojha V
- Security Assessment of Hierarchical Federated Deep Learning
  33rd International Conference on Artificial Neural Networks (ICANN). (2024)
  Alqattan D, Sun R, Liang H, Nicosia G, Snasel V, Ranjan R, and Ojha V
- Dynamic Label Adversarial Training for Deep Learning Robustness Against Adversarial Attacks
  31st International Conference on Neural Information Processing (ICONIP). (2024)
  Liu Z, Duan H, Liang H, Long Y, Snasel V, Nicosia G, Ranjan R and Ojha V
- On Learnable Parameters of Optimal and Suboptimal Deep Learning Models
  31st International Conference on Neural Information Processing (ICONIP). (2024)
  Zheng Z, Liang H, Snasel V, Latora V, Pardalos P, Nicosia G, and Ojha V

Edge AI Hub @ APRIL Hub

Dr Varun Ojha

Edge AI Hub AI Theme Lead
Newcastle University
varun.ojha@newcastle.ac.uk
ojhavk.github.io/
https://github.com/vojha-code