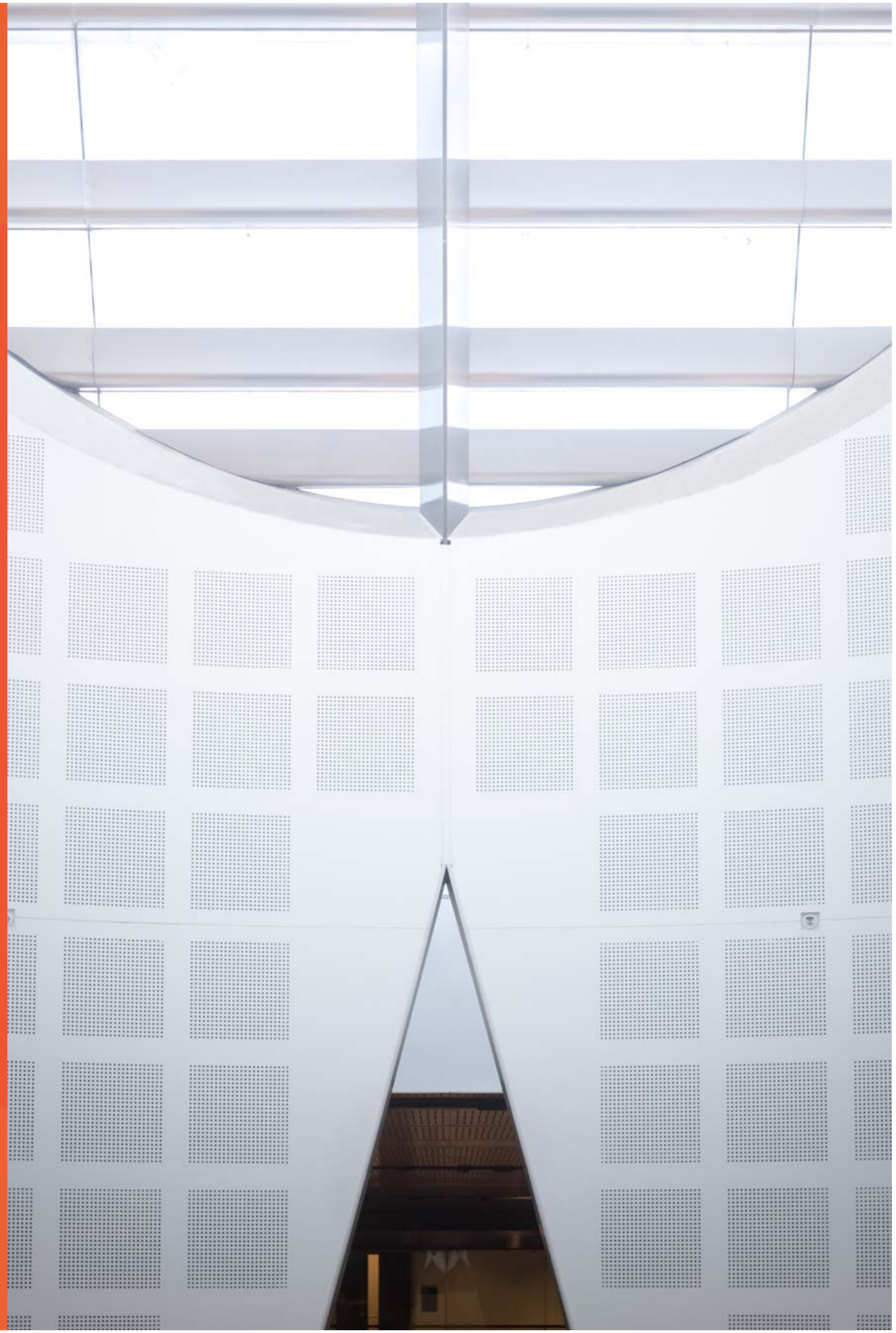
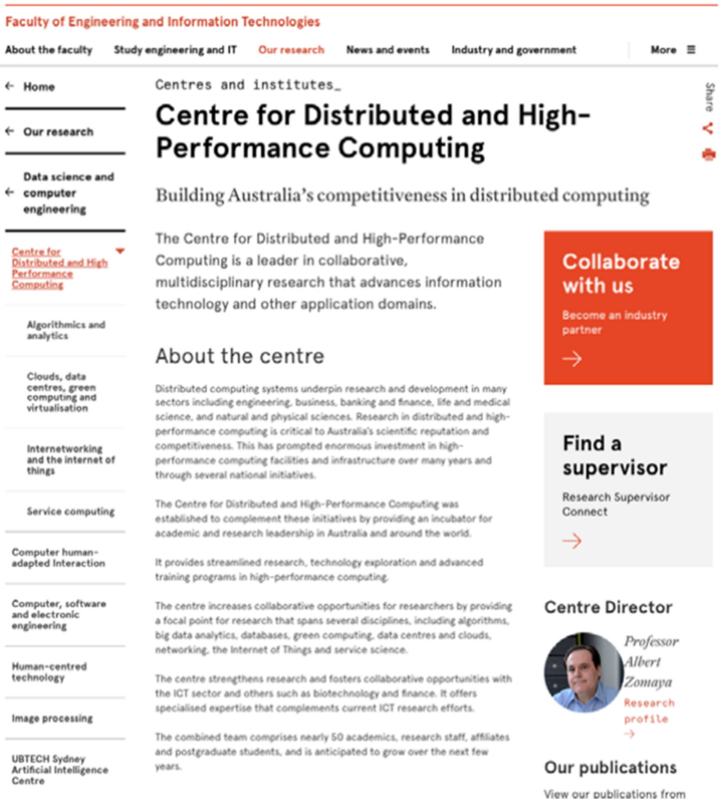
Lightweight Short-term Photovoltaic Power Prediction for Mobile Edge Computing

Albert Y. Zomaya

Chair Professor and Director
Centre for Distributed & High Performance Computing
School of Computer Science
The University of Sydney
Sydney, NSW, Australia
https://albertzomaya.github.io/My-WebPage/

THE UNIVERSITY OF SYDNEY





Faculty of Engineering and Information Technologies

About the faculty | Study engineering and IT | Our research | News and events | Industry and government | More

← Home

← Our research

← Data science and computer engineering

Centre for Distributed and High Performance Computing

Algorithmics and analytics

Clouds, data centres, green computing and virtualisation

Internetworking and the internet of things

Service computing

Computer human-adapted Interaction

Computer, software and electronic engineering

Human-centred technology

Image processing

UBTECH Sydney Artificial Intelligence Centre

Centres and institutes_

Centre for Distributed and High-Performance Computing

Building Australia's competitiveness in distributed computing

Share

The Centre for Distributed and High-Performance Computing is a leader in collaborative, multidisciplinary research that advances information technology and other application domains.

About the centre

Distributed computing systems underpin research and development in many sectors including engineering, business, banking and finance, life and medical science, and natural and physical sciences. Research in distributed and high-performance computing is critical to Australia's scientific reputation and competitiveness. This has prompted enormous investment in high-performance computing facilities and infrastructure over many years and through several national initiatives.

The Centre for Distributed and High-Performance Computing was established to complement these initiatives by providing an incubator for academic and research leadership in Australia and around the world.

It provides streamlined research, technology exploration and advanced training programs in high-performance computing.

The centre increases collaborative opportunities for researchers by providing a focal point for research that spans several disciplines, including algorithms, big data analytics, databases, green computing, data centres and clouds, networking, the Internet of Things and service science.

The centre strengthens research and fosters collaborative opportunities with the ICT sector and others such as biotechnology and finance. It offers specialised expertise that complements current ICT research efforts.

The combined team comprises nearly 50 academics, research staff, affiliates and postgraduate students, and is anticipated to grow over the next few years.

Collaborate with us

Become an industry partner →

Find a supervisor

Research Supervisor Connect →

Centre Director

Professor Albert Zomaya

Research profile →

Our publications

View our publications from

Australian Government

Department of Industry, Innovation and Science

Australia-China Centre for Energy Informatics and Demand Response Technologies

The centre aims to supply efficient, reliable and affordable green energy resources to the Australian and Chinese markets.

Economic, environmental and social benefits:

- a more flexible and reliable energy system
- lower energy bills for consumers
- reduced capital cost investments and need for network upgrades by energy providers
- improved energy system operations and expansion

Partners

- Lead Australian partner: University of Sydney
- Lead Chinese Partner: Tianjin University

Other partners:

- Royal Melbourne Institute of Technology
- EnergyAustralia Pty Ltd
- Tsinghua University
- State Grid Corporation of China

– 11 Researchers
– 20 Affiliates (Australia, Asia, Europe, South America, USA)
– 40+ graduate students

FACEBOOK Research

Australian Government Department of Defence Defence Science and Technology Group

DATA 61 CSIRO

DEFENCE INNOVATION NETWORK To go far, go together

Australian Government Australian Research Council

TOKENONE

amazon webservices

CISCO

国家电网 STATE GRID

BEIJING 2022

Microsoft

EnergyAustralia LIGHT THE WAY

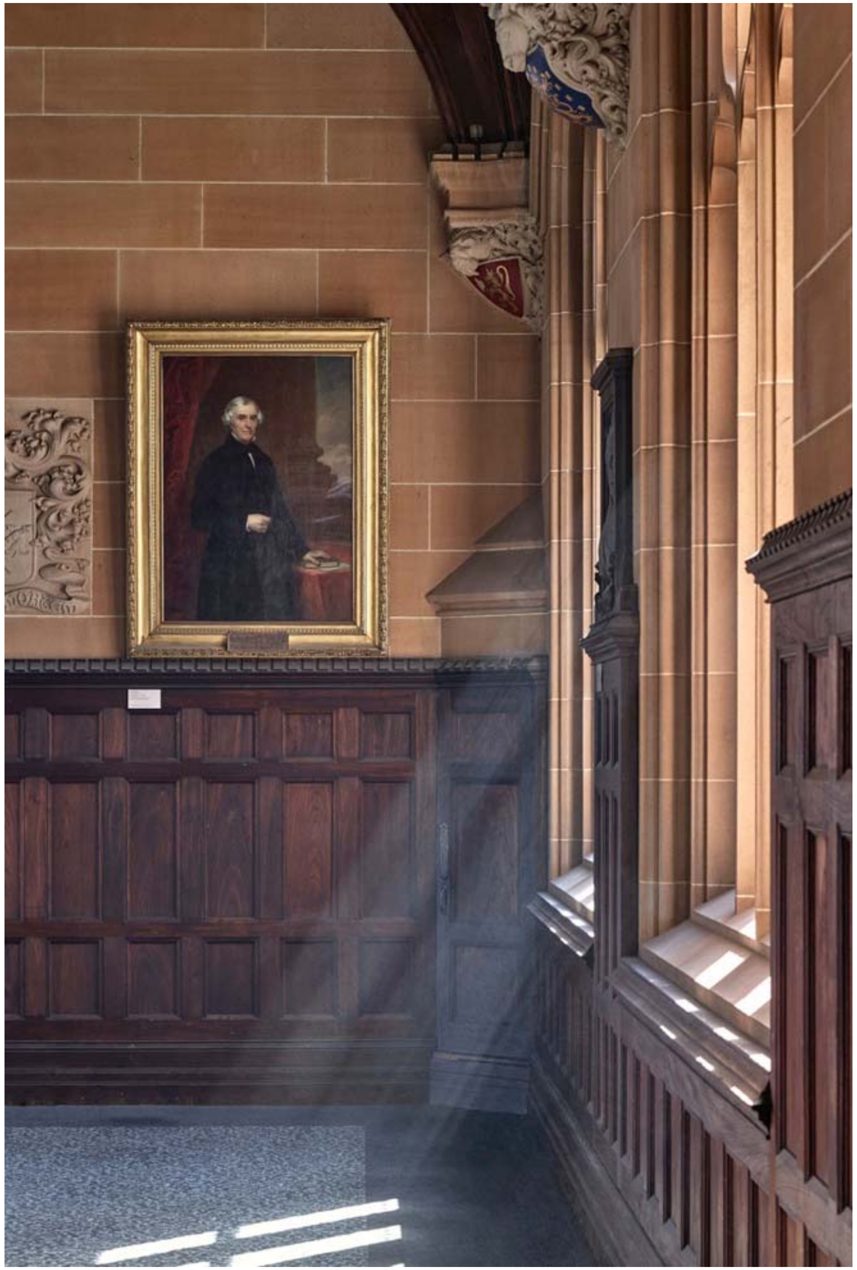
Australian Government Department of Industry, Innovation and Science

Our Research

Research article abstracts of last 5 years…..

Agenda

- Background & Motivation
- Methodology & Implementation
- Experiment & Result Analysis
- Summary
- Future Work



THE UNIVERSITY OF SYDNEY

Background (Massive Investments in Clouds)

1984 – John Gage, Sun Microsystems

“The network is the computer”

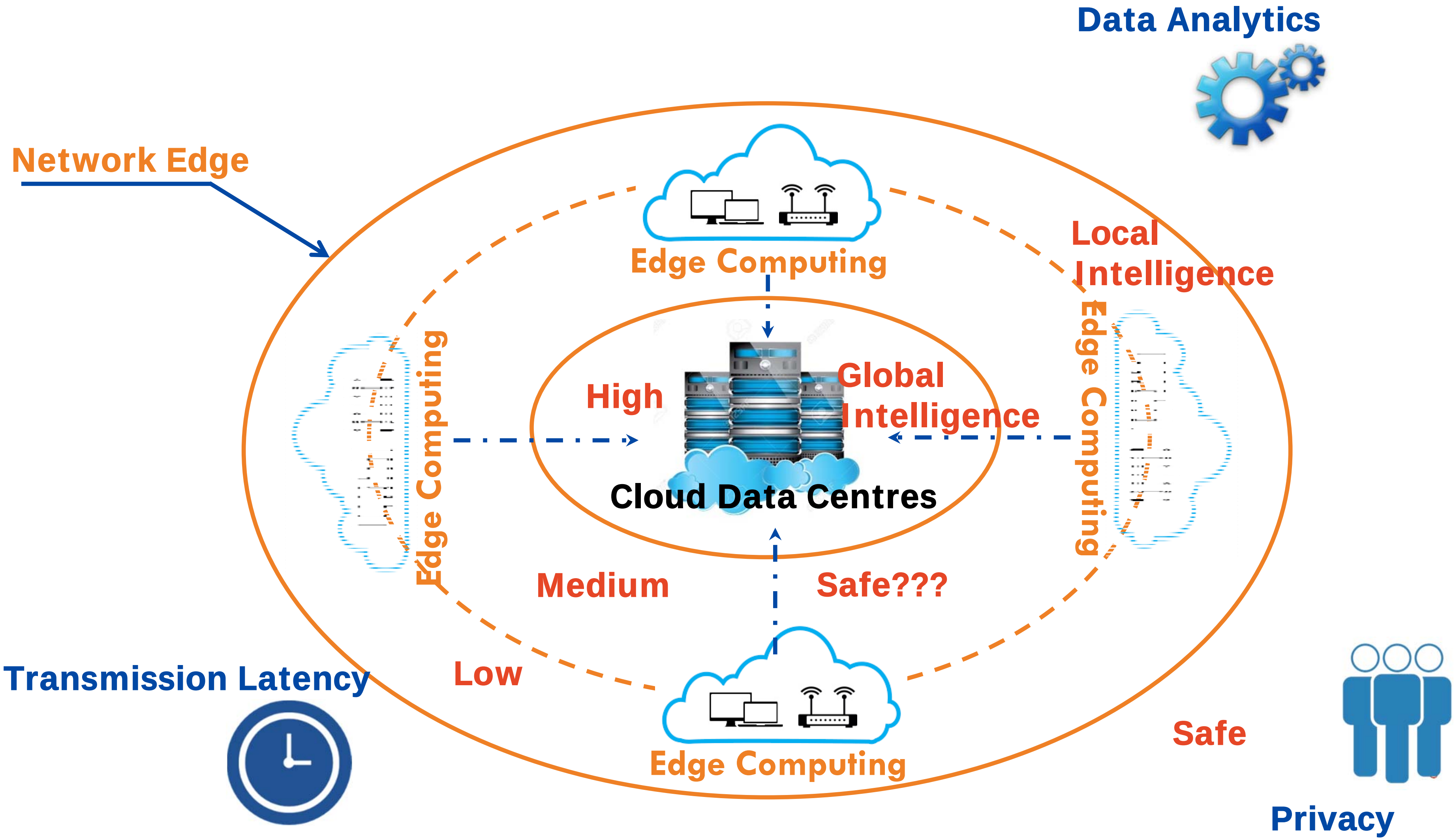
1993 – Eric Schmidt, Google “When the network becomes as fast as

the processor, the computer hollows out and spreads across the network.”

2008 – David Patterson, U. C. Berkeley

“The data center is the computer”

Edge Computing Milieu



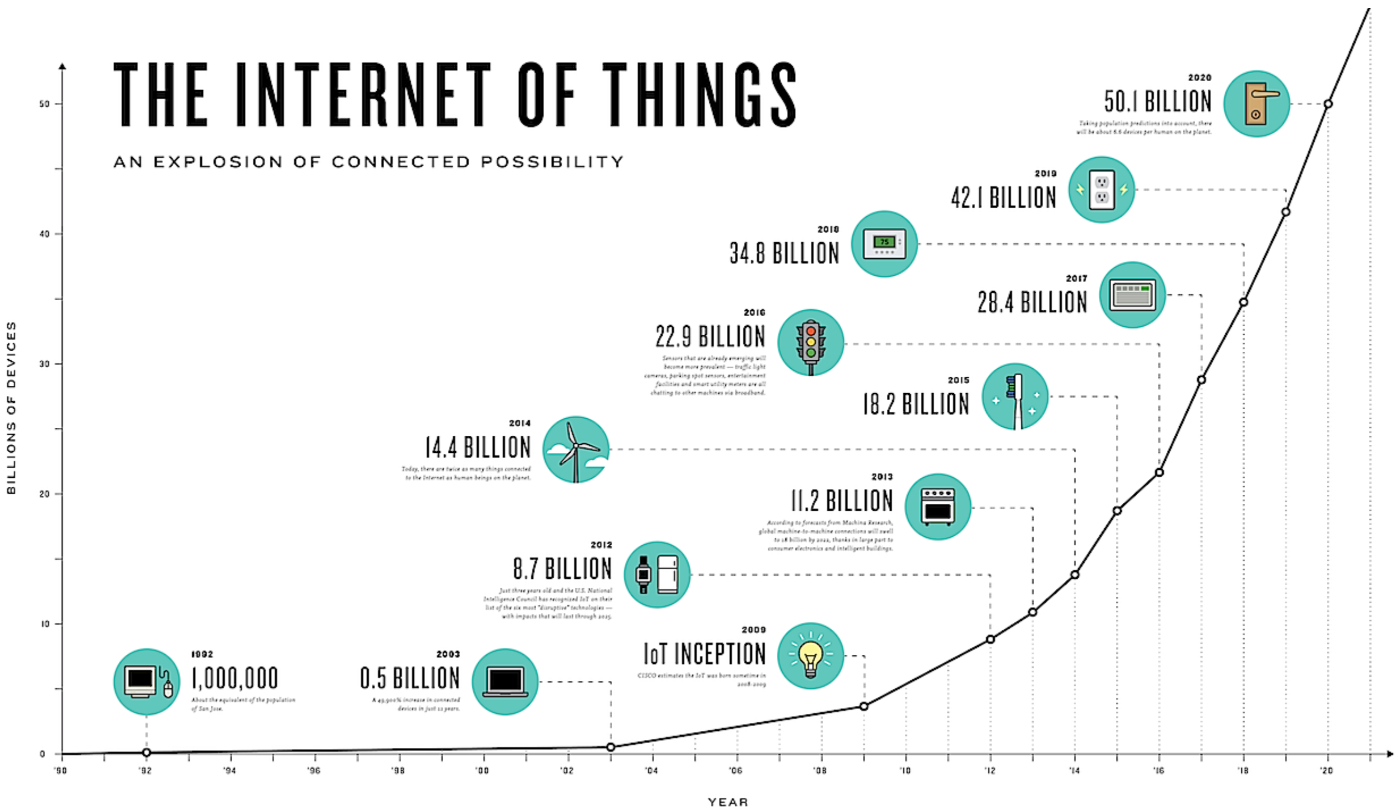
Background (Edge Computing, Fog Computing, …)

Table : Similarities and differences between edge computing and cloud computing

	Edge Computing				Cloud Computing
	MEC	MCC	Cloudlet	FC	
Ownership	Telco	Third-parties, private entities and individuals			Cloud providers
Node location	Edge	Edge, devices	Near edge	Near edge, edge	Network core
Context awareness	High	High	Low	Medium	Low
Latency and jitter	Low	Low	Medium	Low	High
Scalability	High	High	Medium	High	Low

- local processing power close to the source of data
- Traditional edge processing power is given to the IoT device itself
- While in fog computing, computing nodes (e.g., Dockers and VMs) are placed very close the source of data.
- The 'edge computing' paradigm depends on programmability of IoT devices to directly communicate with each other and run user defined codes.
- Unfortunately, standard APIs that provide such functionality are not fully adopted by current IoT sensors/actuators, etc.

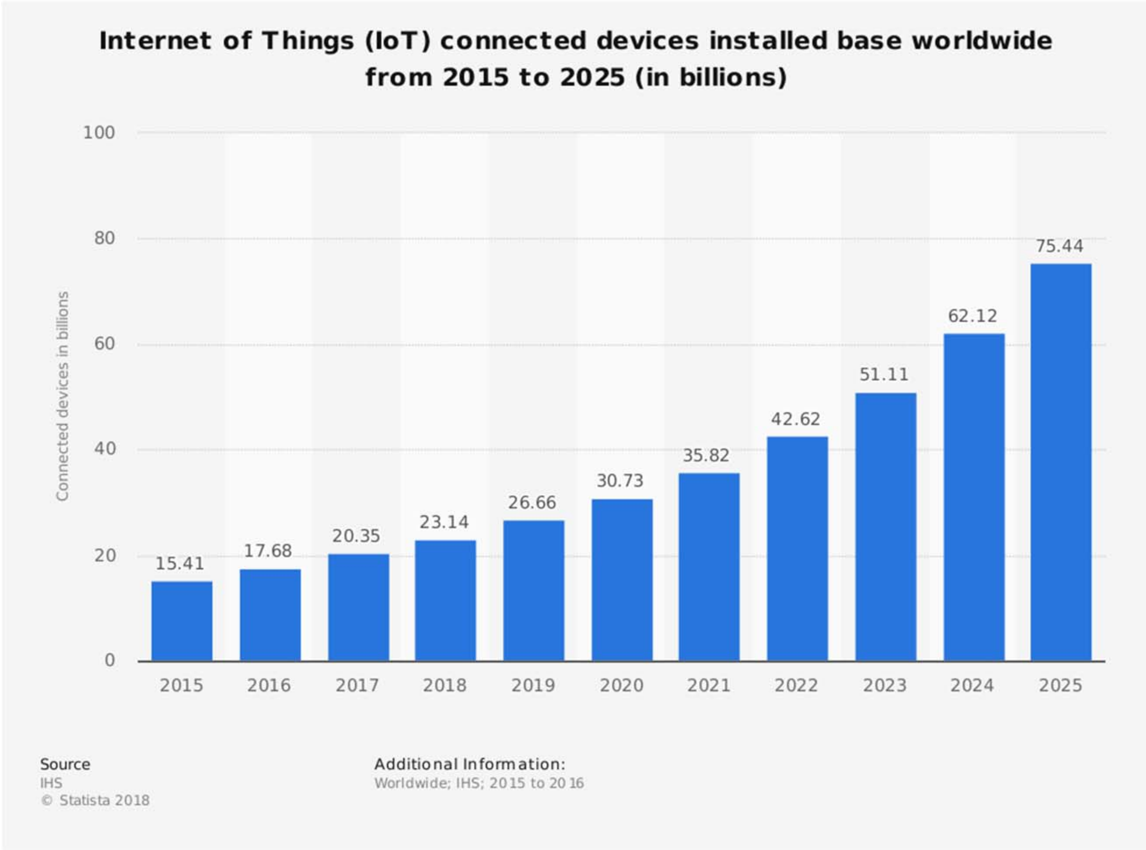
Background (IoT)



IoT: the term coined in1999 by **Kevin Ashton**, co-founder of Auto-ID Center at MIT. IoT as enabler of a world where physical objects are tagged and uniquely identified by RFID transponders.

Rapid Data Generation

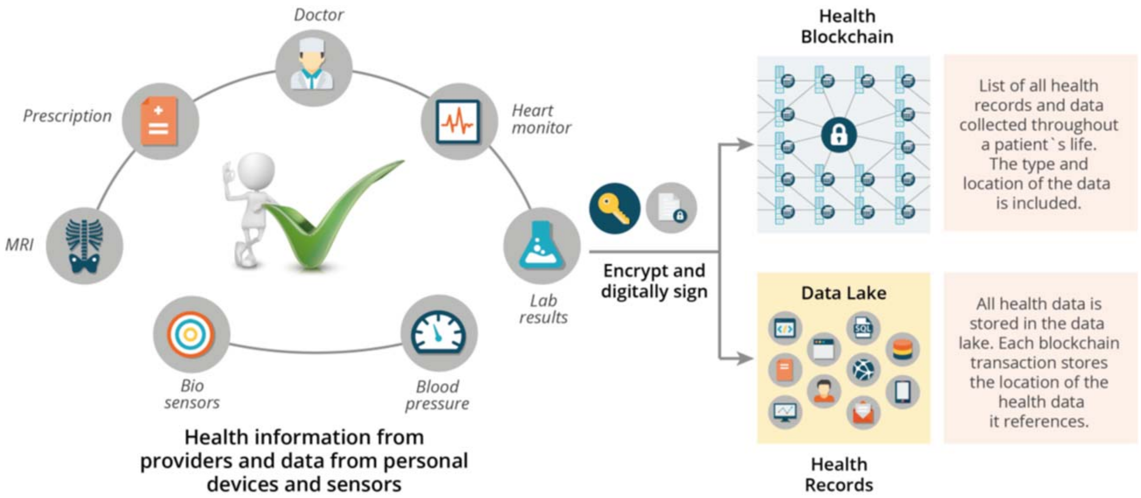
- Internet of Things
  - Growing at an ever-increasing rate
  - By 2025 – over 70 billion Internet of Things devices connected
- Big Data
  - Doubles every 2 years
  - By 2020 – 44Zettabytes or 44 trillion gigabytes
- A big beneficiary of Edge Computing is IoT.

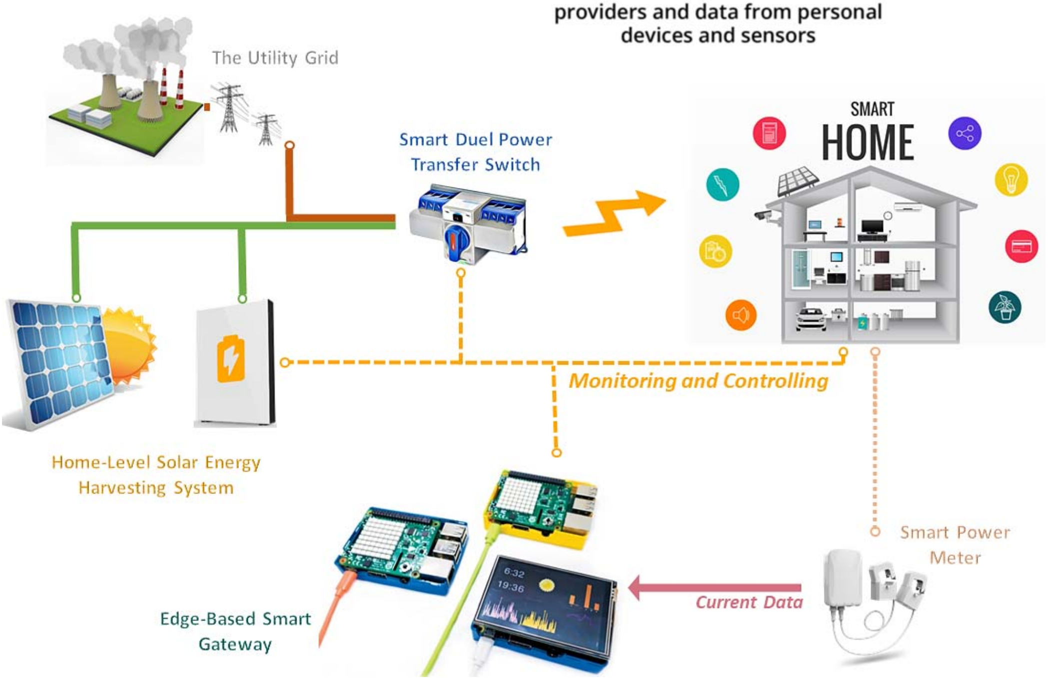


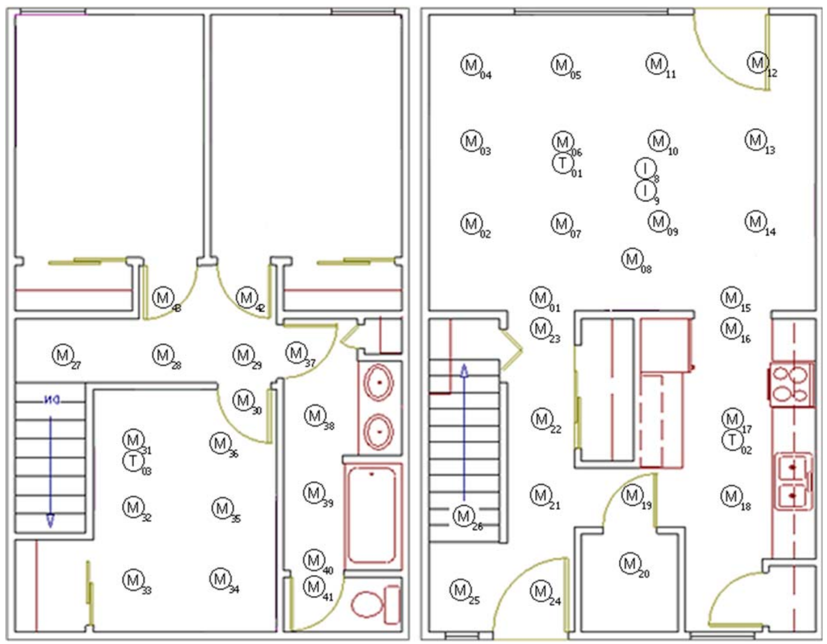
Source: https://www.statista.com/statistics/471264/iot-number-of-connected-devices-worldwide/

[Optimal Application Deployment in Resource Constrained Distributed Edges, *IEEE Transactions on Mobile Computing*, 2020]

Motivation

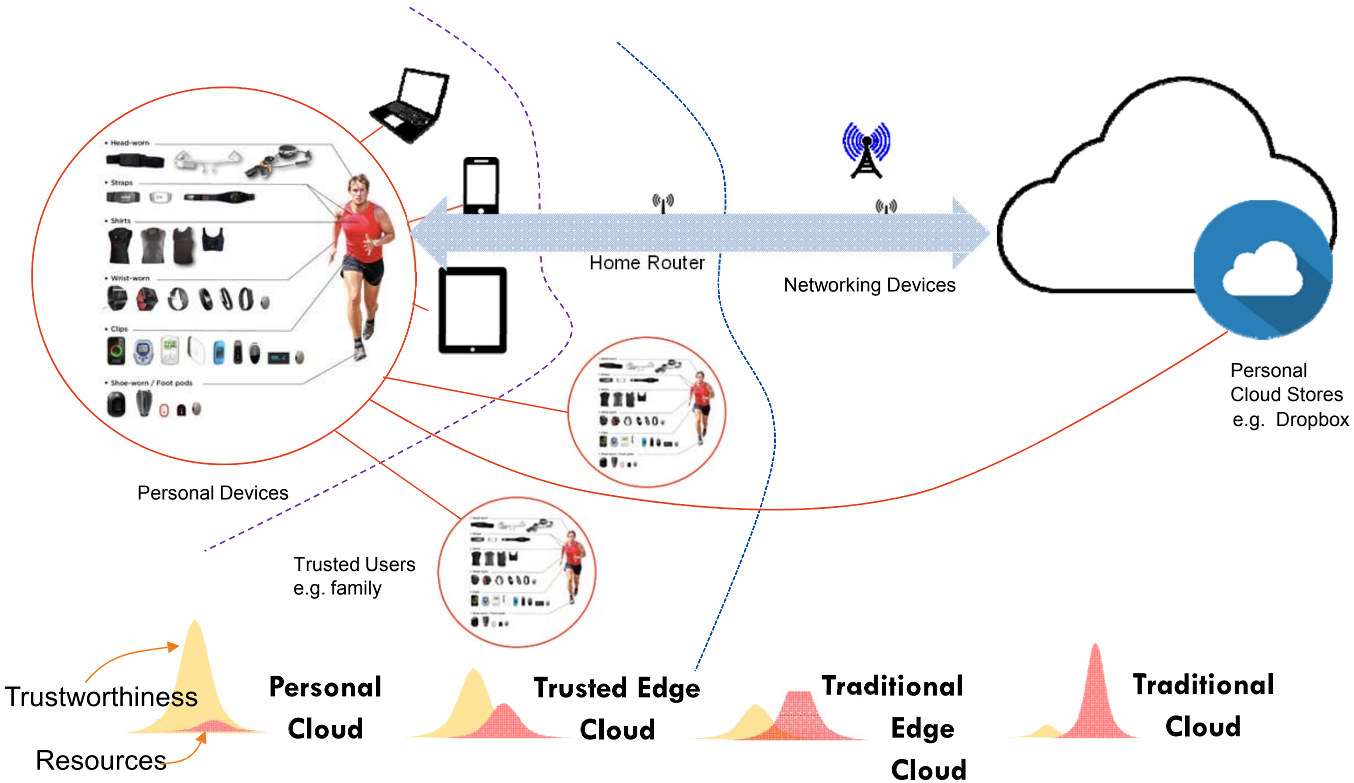






Deployments of edge computing can lead to high energy consumption

More Opportunities

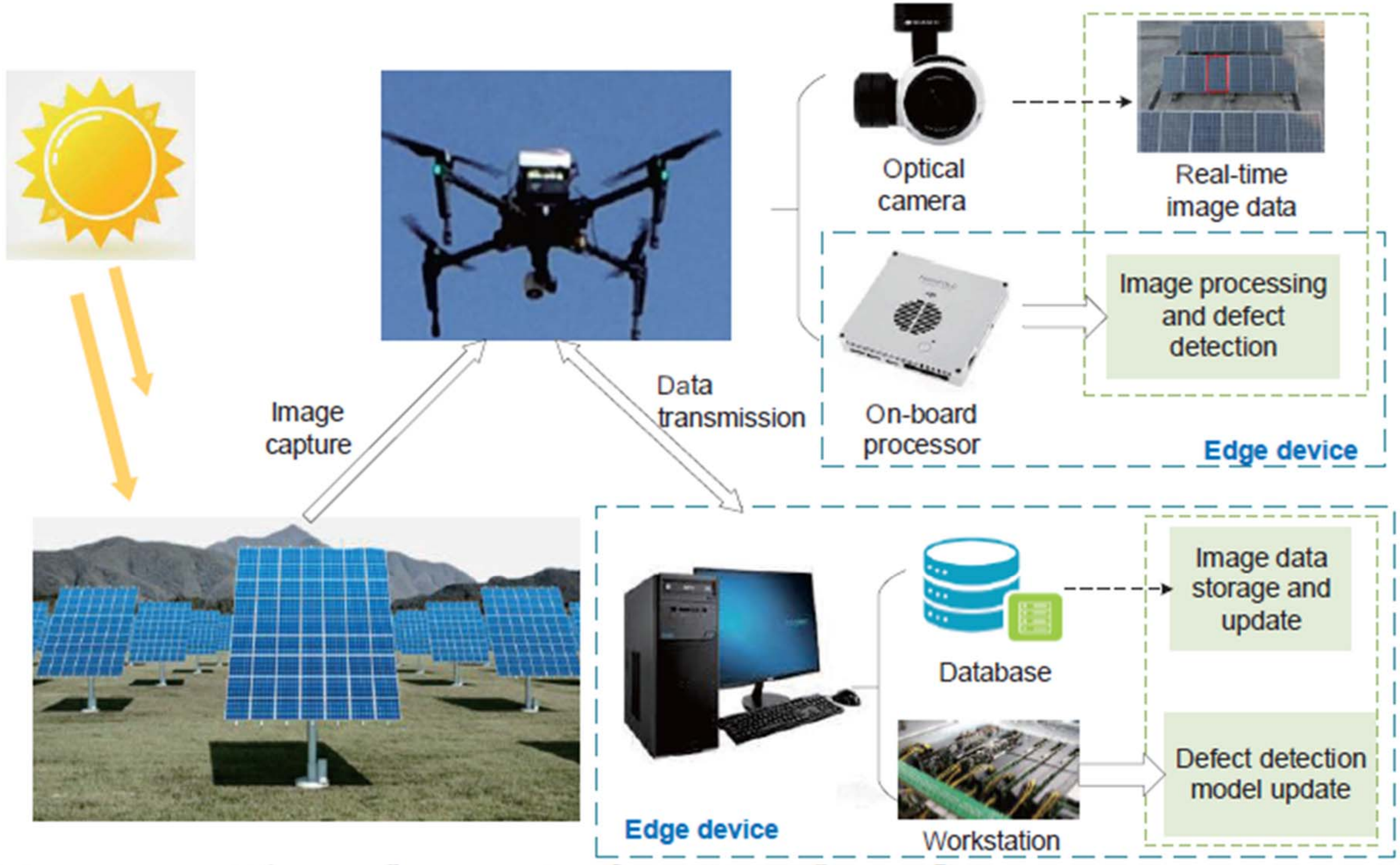


Spatio-temporal dependent characteristics of networking infrastructure

Elastic Computing Platforms for Efficient and Trustworthy Mobile Data Processing and Delivery

[Building an Online Defect Detection System for Large-scale Photovoltaic Plants, *ACM BuildSys*, 2019]

Example

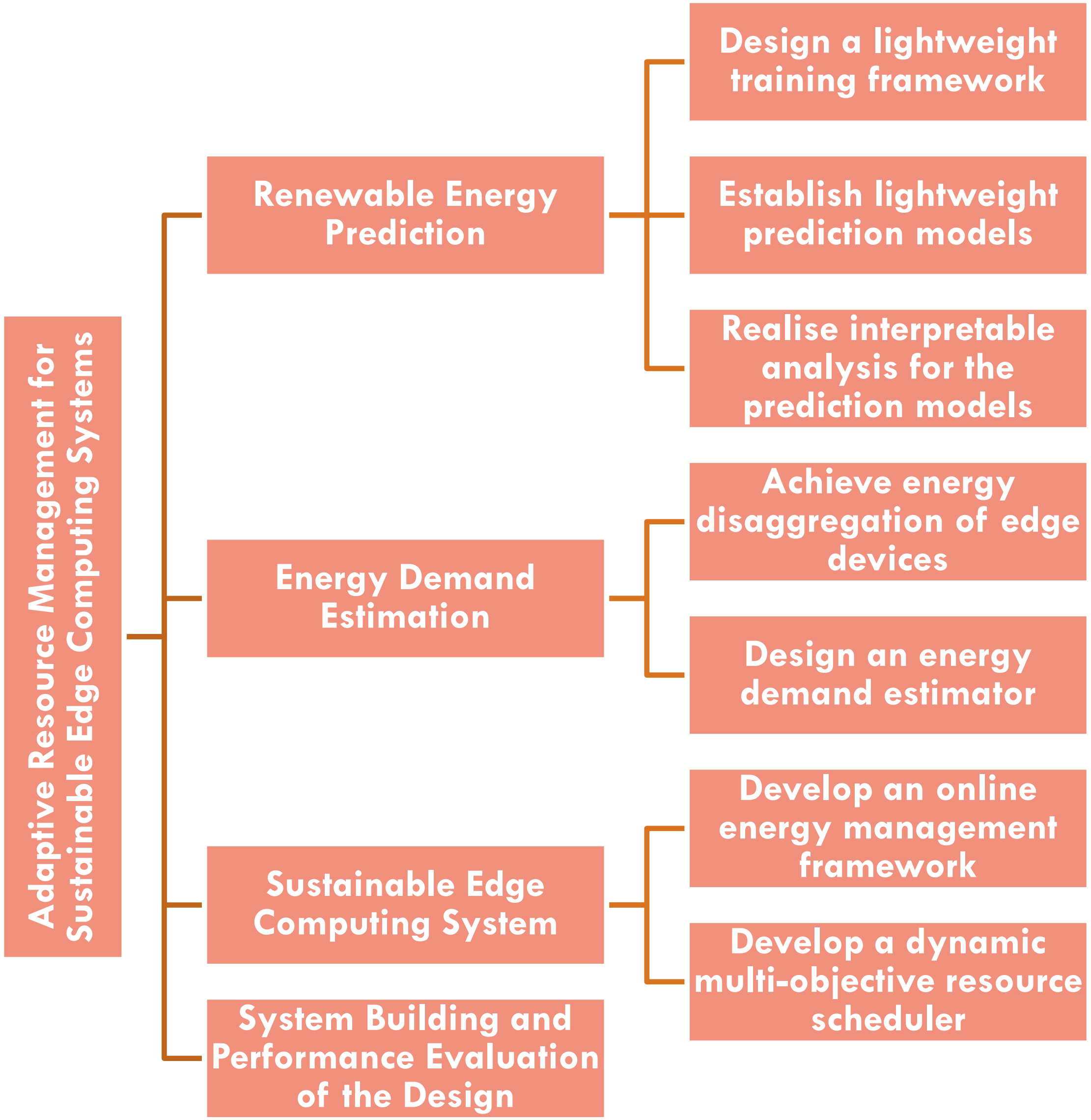


Potential Problems

- Hard to utilize solar energy as stable source as opposed to utility grids.
- The uncertainty of load demand leads to difficulty in resource allocation.
- Degradation in system performance due to lack proper resource management.

[Request Dispatching for Minimizing Service Response Time in Edge Cloud Systems, *ICCCN*, 2018. (Invited Paper)]

A Complex Problem



[Optimal Application Deployment in Resource Constrained Distributed Edges, *IEEE Transactions on Mobile Computing*, 2020]

Challenges of PV Power Prediction

- Lack of a cost-effective model for PV output prediction, which not only can improve accuracy but lower the computation overhead.
- Few predictive models take temporal patterns of solar energy generation into consideration. (*A temporal pattern is defined as a segment of signals that recurs frequently in the whole temporal signal sequence*).
- Models with high accuracy mostly require high computation and storage resources in runtime, which will fit edge computing systems with limited resource.
- The existing lightweight models cannot provide accurate prediction as expected.

[Interpretable Machine Learning In Sustainable Edge Computing: A Case Study of Short-Term Photovoltaic Power Output Prediction. *ICASSP*, 2020. (Invited paper)]

Current State of the Art for Solar Power Prediction

- **Classical Statistical Algorithms**

I. *Numerical weather prediction (NWP)*

II. *Auto-Regressive and Moving Average Model (ARMA)*

- **Machine Learning Algorithms**

I. *Extreme learning machine (ELM)*

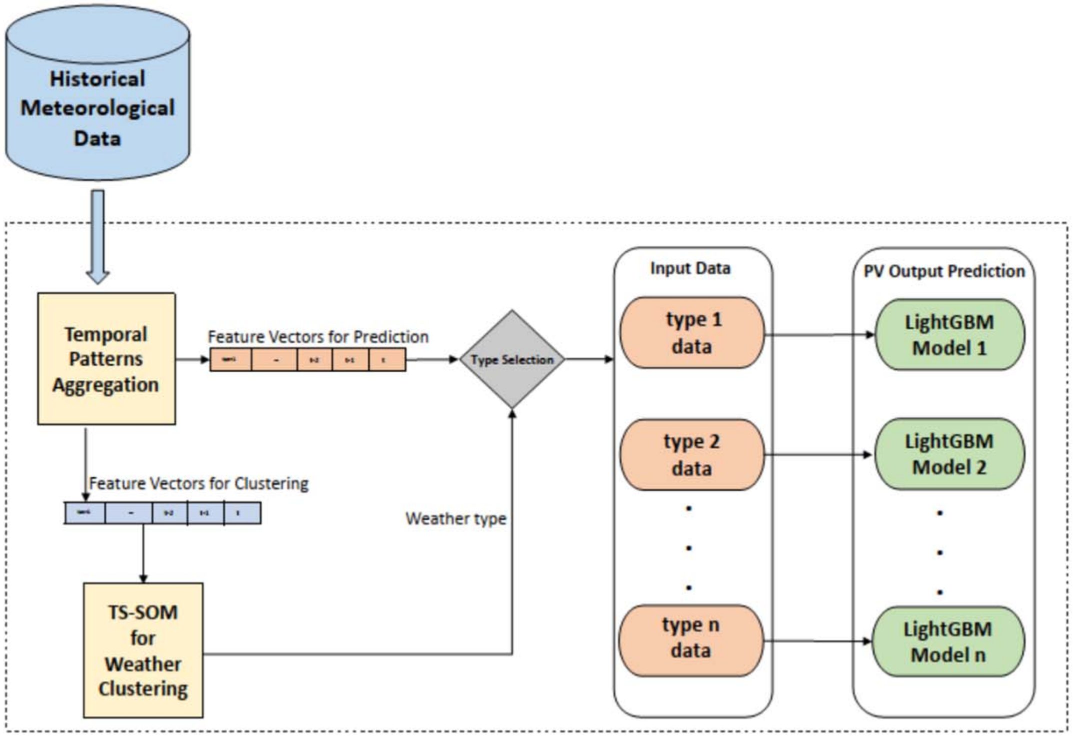
II. *Support vector regression (SVR)*

III. *General regression neural network (GRNN)*

IV. *Recurrent Neural Network (RNN)*

Model Training and Prediction

- *Temporal Patterns Aggregation*
- *Weather Clustering*
- *Model Establishment*
- *Interpretability Evaluation*



Framework of clustering-based prediction model

[Edge Intelligence: The Confluence of Edge Computing and Artificial Intelligence. *IEEE Internet of Things Journal*, 2020]

Methodology & Implementation

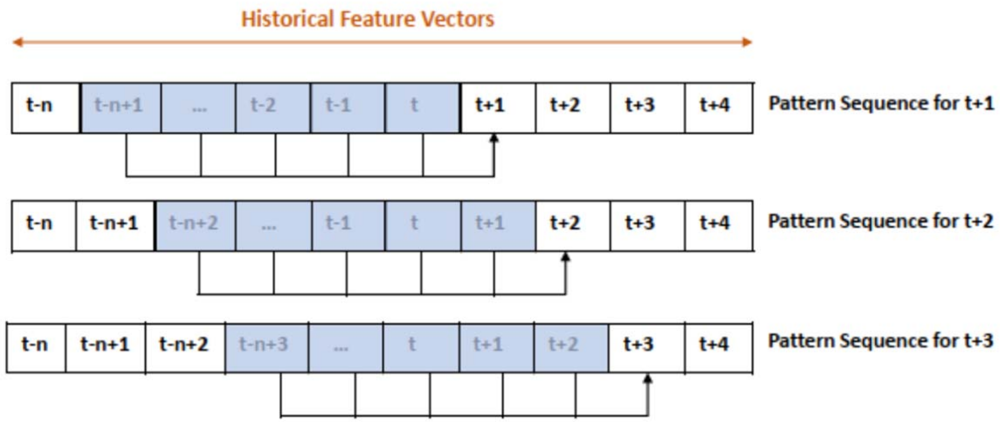
Meteorological Features

Field	Description	Pearson correlation coefficient
ghi	Global horizontal irradiance for centre value	0.9767
ghi10	Global horizontal irradiance for 10% value	0.9629
ghi90	Global horizontal irradiance for 90% value	0.9744
dni	Direct normal irradiance for centre value	0.9275
dni10	Direct normal irradiance for 10% value	0.9077
dni90	Direct normal irradiance for 90% value	0.8644
dhi	Diffuse horizontal irradiance	0.9314
ebh	Direct horizontal irradiance	0.6617
air_temp	Air temperature	0.3350
zenith	Solar zenith angle. Range: 0∼180	-0.8013
azimuth	Solar azimuth angle. Range: -180∼180	-0.0092
cloud_opacity	The quantity of cloud.	-0.2184

The *ghi* is highly influenced by the *weather* condition. (e.g., sunny, partially cloudy, and cloudy)

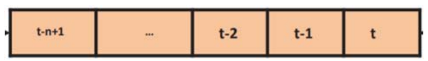
Methodology & Implementation

Temporal Patterns Aggregation



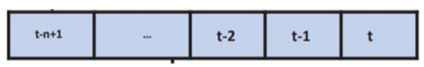
The iterative strategy for temporal pattern aggregation

Feature Vector for Clustering



$$C_{t+1} = \{W_{t-n+1}, \ldots, W_{t-1}, W_t\}$$

Feature Vector for Prediction

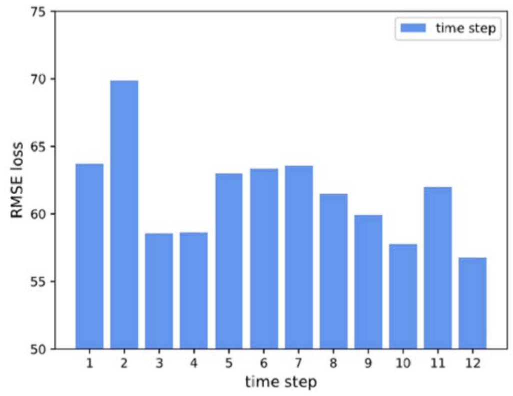


$$F_{t+1} = \{I_{t-n+1}, p_{t-n+1}, \ldots, I_{t-1}, p_{t-1}, I_t, p_t\}$$

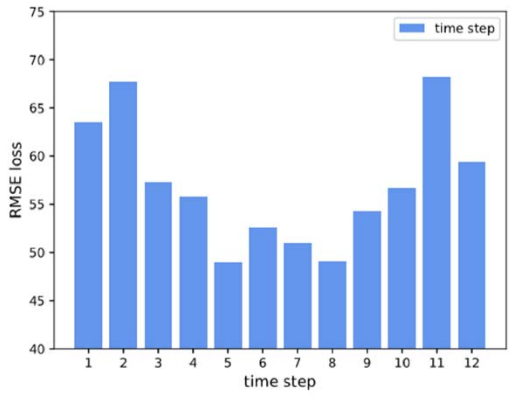
C_t: aggregated feature vectors at time t for weather clustering
W_t: meteorological patterns at time t
F_t: aggregated feature vectors at time t for model training and PV output prediction
I_t: irradiance factors at time t, including ghi, dhi and dni
p_t: actual PV power output at time t

Methodology & Implementation

Temporal Patterns Aggregation

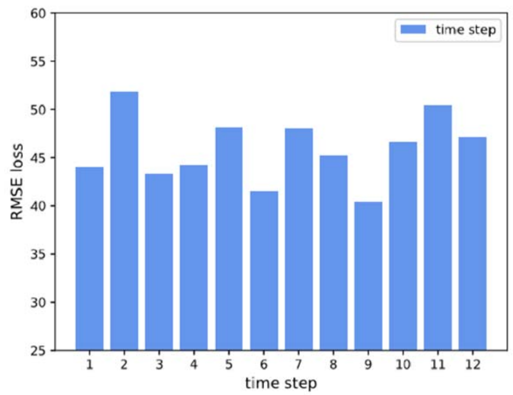


SVR

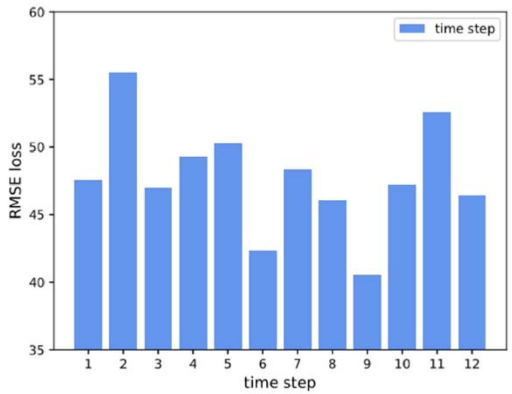


GRNN

RMSE(root mean square error)



XGBoost



LightGBM

Accuracy of four prediction models with different sizes of time step (which is non-linear)

Methodology & Implementation

Weather Clustering

- **Motivation**

I. Reduces spikes and falls in the PV output curves.

II. Builds up a clustering-based model to reduce prediction errors.

III. Copes with the potential issue of data loss.

- **Approach**

I. Tree-structured Self-organizing Map (TS-SOM)

II. Divides the data set into multiple groups and each node of the tree is designed as a traditional SOM neural network.

III. Enables detailed weather clustering gradually from the root to the leaf.

IV. Fit for clustering of the hierarchical data.

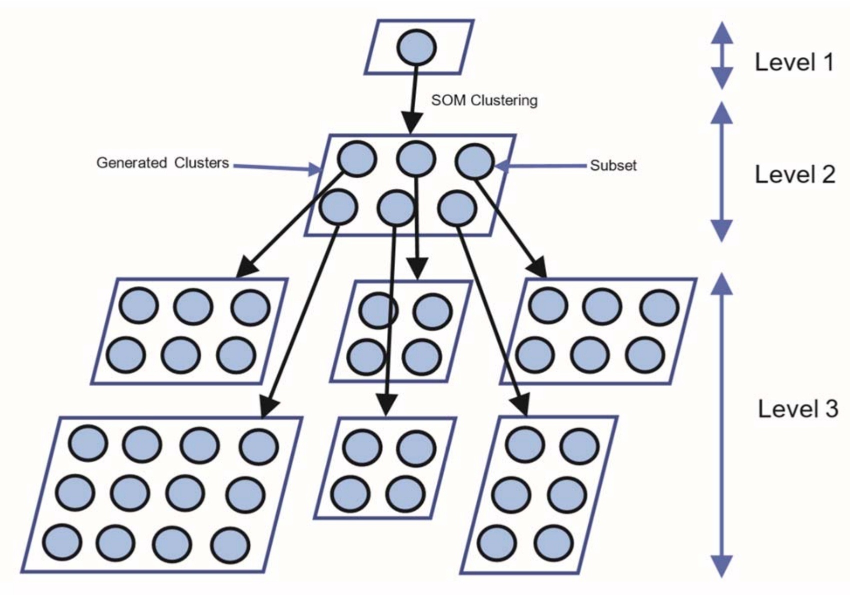
Methodology & Implementation

TS-SOM for Weather Clustering

Set threshold for the tree to stop the TS-SOM recursion

Enable detailed weather clustering recursively

Update global information of the tree (tree attributes) after each split

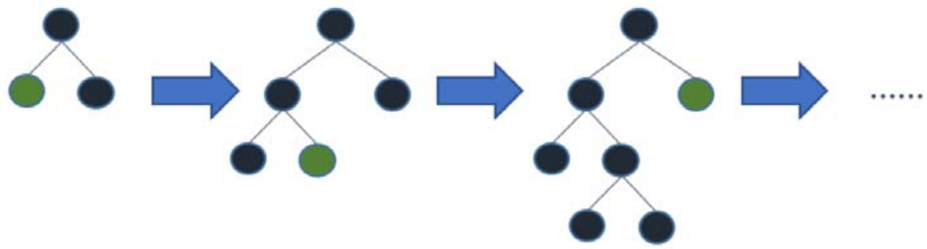


The Framework of Tree-based SOM

Methodology & Implementation

Power Output Prediction Model

- **Tree-based ensemble regression method: LightGBM**
- Ensemble a set of weak learners generated at different training time steps, mostly using Classification and Regression Tree (CART).
- Sums up their results as the final prediction output iteratively.
- Establish a prediction model for each weather cluster.



Leaf-wise tree growth

Methodology & Implementation

Power Output Prediction Model

Novel techniques of LightGBM

- Histogram-based split algorithm
- Gradient-based one-side sampling (GOSS)
- Exclusive feature bundling

Advantages of LightGBM

- Faster training speed and higher efficiency
- Lower memory usage
- Better accuracy
- Compatibility with Large Datasets
- Parallel learning supported

Interpretability Analysis for Prediction Models

Definition

Interpretable machine learning is a technique used to give machine learning models the ability to explain or to present their behaviors in understandable terms to humans.

Background & Motivation

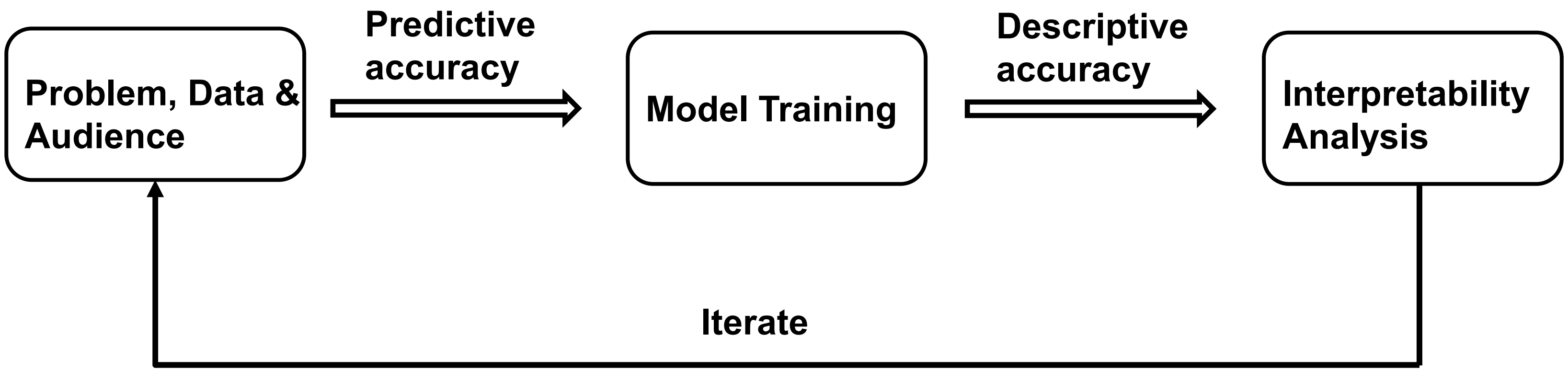
- Machine-learning models have demonstrated great success in learning complex patterns and making predictions about unobserved data.
- However, complex models lack transparency behind their behaviors, which leaves users with little understanding of how particular decisions are made by these models.
- The concerns about the black-box nature of complex models have hampered their further applications.
- Interpretable machine learning would be an effective tool to mitigate these problems and it has recently received considerable attentions.

[Edge-Computing-Enabled Unmanned Module Defect Detection and Diagnosis System for Large-Scale Photovoltaic Plants. *IEEE Internet of Things Journal*, 2020]

Interpretability Analysis for Prediction Models

Traditional Data Science Life Cycle

Interpretation In Data science Life Cycle



[Realising Edge Analytics for Early Prediction of Readmission: A Case Study, *IC2E*, 2020 (Invited paper)]

Interpretability Analysis for Ensemble Models

Feature importance is a simple yet effective explanatory measure to indicate statistical contribution of each feature to the underlying model.

Traditional Feature Attribution Methods:

- Weight: *Count the number of times a feature used to perform the node partitions across the forest.*
- Gain: *Measure the average training loss reduction as the gain when using a feature for node partition.*
- Cover: *Calculate the average coverage of node splits for a certain feature.*

These traditional attribution methods violate the consistency principle of feature attribution

Interpretability Analysis for Ensemble Models

SHapley Additive Explanation (SHAP)

SHAP is an additive feature attribution method that explains a model's output as a sum of real values attributed to each input feature.

Advantages of SHAP

- **SHAP** averages over all possible orderings of the features, rather than just the ordering specified by their position in the tree, which is consistent.
- **SHAP** contains the information of both global explanation and localized explanation for individual prediction.

Experiments & Results Analysis

Experimental Setup

- **Powerful Desktop**
  - i7-7700 (6c/12t) CPU
  - 32GB DDR3 Memory
  - 2T SSD
  - 7200rpm Hard Disk
  - Ubuntu 16.04

- **Raspberry Cluster**
  - Raspberry Pi 3B
  - Quad-core CPU
  - 1GB RAM
  - 64-bit Ubuntu Mate
  - Dask Framework

- **Software Package**

Python 3.7, Scikit-learn 0.21.3 and Tensorflow 1.13.2.

Experiments & Results Analysis

Evaluation Metrics

- **Mean Absolute Error (MAE)**

$$\text{MAE} = \frac{1}{N}\sum_{t=1}^{N}|\widehat{y}(t) - y(t)|$$

- **Root Mean Squared Error (RMSE)**

$$\text{RMSE} = \sqrt{\frac{1}{N} \times \sum_{t=1}^{N}(\widehat{y}(t) - y(t))^2}$$

- **Training Time Cost**

- **Prediction Time Cost**

- **Coefficient of determination (denoted as "R2 ")**

$$R^2 = 1 - \frac{\sum_{t=1}^{N}(y(t) - \widehat{y}(t))^2}{\sum_{t=1}^{N}(y(t) - \overline{y})^2}$$

Proportion of the variance in the dependent variable that is predictable from the independent variable.

Experiments & Results Analysis

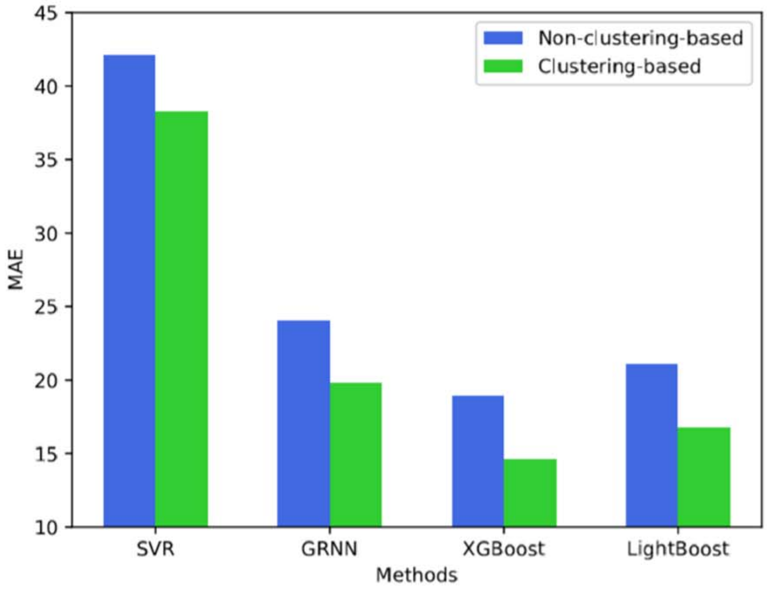
Prediction Performance Evaluation

Algorithm	MAE	RMSE	R^2
Group 1: Non-clustering-based			
SVR	42.13	56.74	0.9469
GRNN	24.02	48.94	0.9536
XGBoost	18.94	40.55	0.9712
LightGBM	21.12	40.57	0.9712
Group 2: Clustering-based			
SVR	38.31	51.42	0.9418
GRNN	19.82	46.85	0.9602
XGBoost	14.60	32.58	0.9825
LightGBM	16.79	35.49	0.9792
Group 3: Recurrent neural network			
LSTM	25.12	45.90	0.9652
GRU	25.37	47.85	0.9610

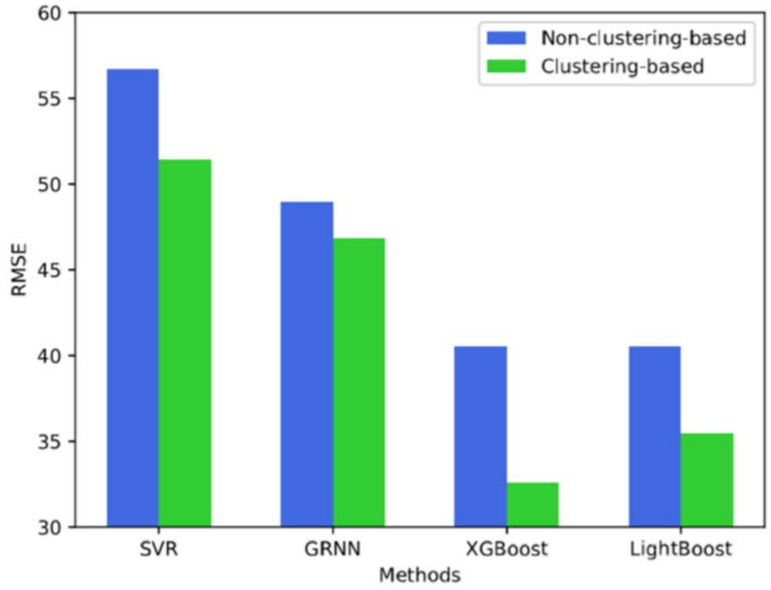
Two ensemble methods (**LightGBM** and **XGBoost**) dominate on all metrics and provide better performance on the test data set with higher accuracy and a lower error rate compared to the others. Besides, with more than 0.97 of R2, these two models indicate that they can also provide better generalization performance and fit to diverse weather conditions.

Experiments & Results Analysis

Prediction Performance Evaluation

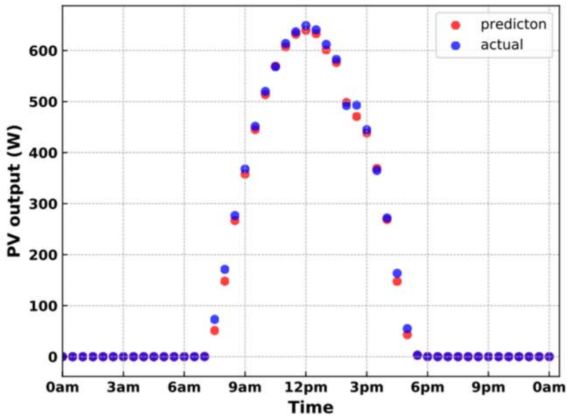


MAE Evaluation

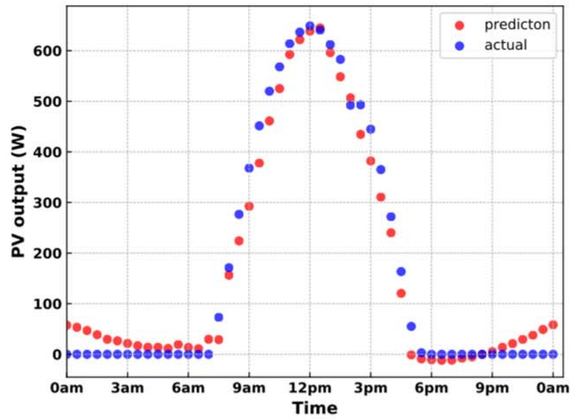


RMSE Evaluation

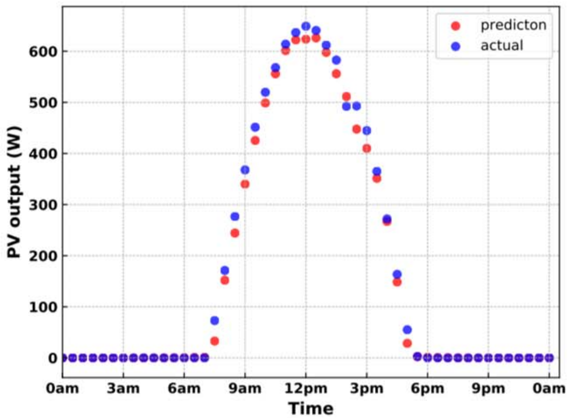
Experiments & Results Analysis



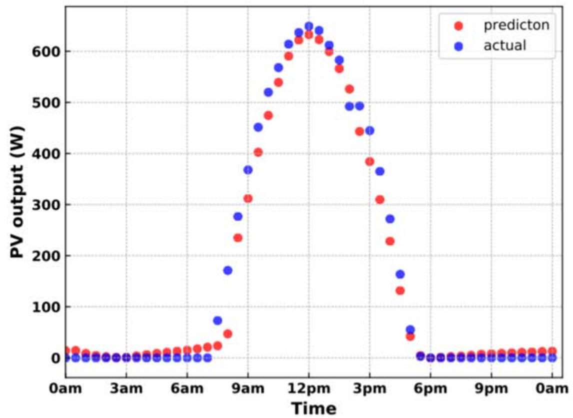
LightGBM



SVR



GRNN



LSTM

PV power output prediction for a typical day

Experiments & Results Analysis

Cost on Powerful Desktop

Algorithm	Training(s)	Execution(s)
Group 1: Non-clustering-based		
SVR	0.4343	0.013
GRNN	0.5513	0.207
XGBoost	0.4735	0.00076
LightGBM	0.0540	0.00029
Group 2: Clustering-based		
SVR	0.3130	0.0093
GRNN	0.1960	0.1016
XGBoost	1.0033	0.0067
LightGBM	0.1669	0.0022
Group 3: Recurrent neural network		
LSTM	299.26	20.52
GRU	219.56	21.22

Cost on Raspberry Pi Cluster

Algorithm	Training time(s)	Execution time(s)
SVR	3.90	0.15
GRNN	2.82	0.407
XGBoost	16.83	0.023
LightGBM	1.39	0.020

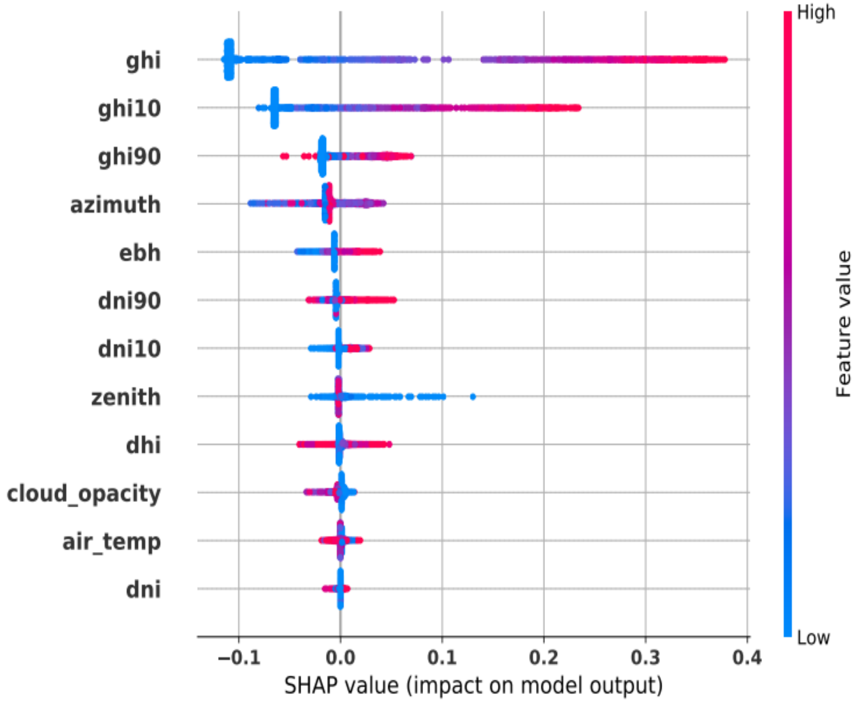
Experiments & Results Analysis

Interpretability Evaluation

The feature importance of ”ghi” is the highest, which means it makes the most contribution to the prediction.

It is noteworthy that ”ghi” makes the most impacts on most predictions.

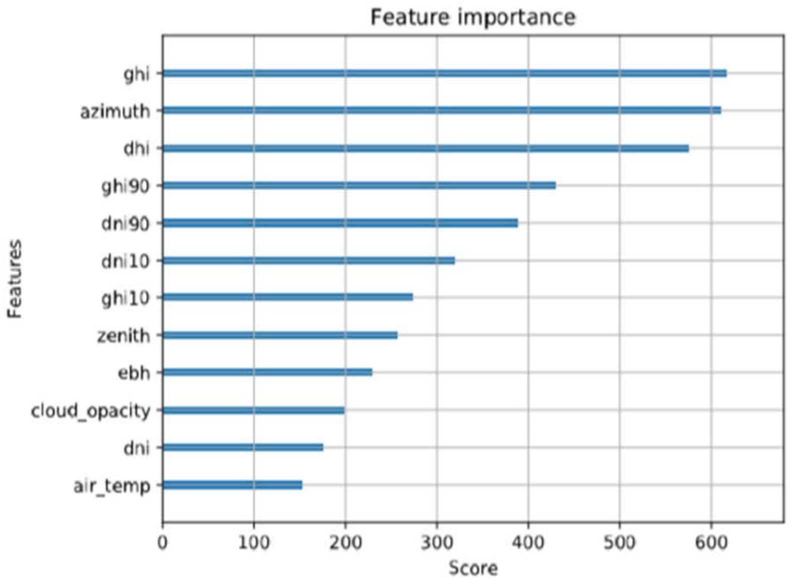
SHAP indeed guarantees the consistency in prediction model by comparing results of **Gain** and **Weight.**



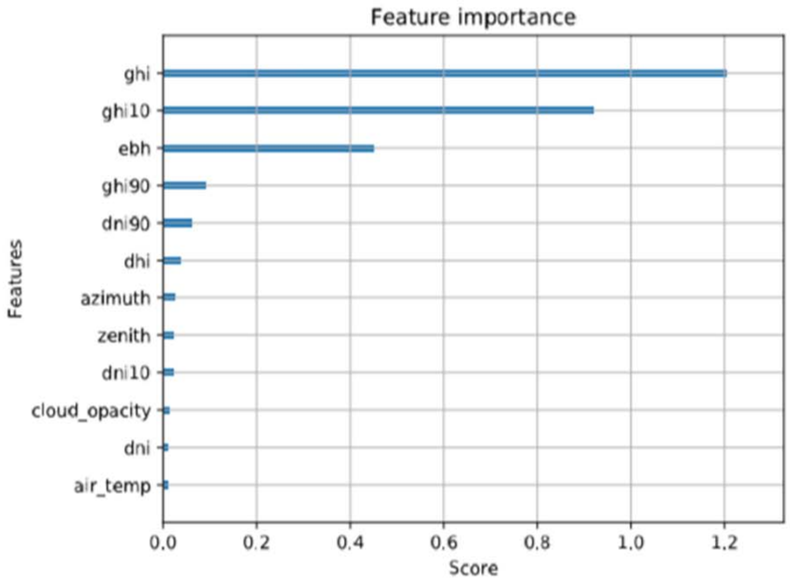
The *ghi* is highly influenced by the *weather* condition. (e.g., sunny, partially cloudy, and cloudy)

Experiments & Results Analysis

Interpretability Evaluation



Results of Gain



Results of Weight

Summary

Our proposed clustering-based model surely provides high-level performance on tackling the problem of solar power prediction and lowers the resource overhead when compared to other widely used regression algorithms.

The feature ”ghi” makes the most contribution on PV output prediction, and the characteristics of solar power generation are also reflected in the orderings of feature importance.

[A Lightweight Short-term Photovoltaic Power Prediction for Edge Computing, *IEEE Trans on Green Communications and Networking*, 2020]

Future Work

Enhance generalization ability of the proposed PV output prediction

Overcome multiple challenges related to resource management on edge devices

Develop a dynamic multi-objective resource scheduler

Finally....

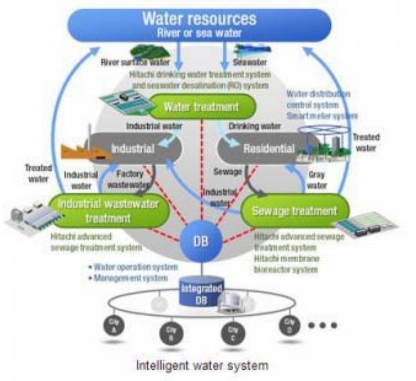
- Edge Computing is a rapidly evolving technology with many opportunities, and many more challenges.
- **Performance predictability** is a major concern in current Edge systems where **heterogeneous resources** are to be allocated.
- **Realistic performance metrics** are needed at the application level to collectively and truly represent **performance variances** occurring across all system-level components.
- **New optimisation models** that integrate a variety of resources with strict real-time/capability constraints.
- **Trade-off frontiers** between an Edge system performance and associated costs. This is necessary to inform resource allocation across a platform as well as well to cap expenditure.

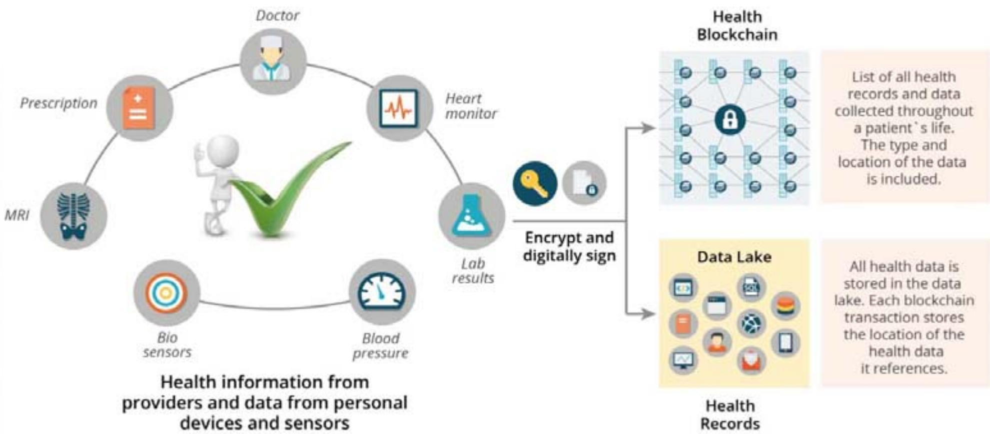
[Federated Learning over Wireless Networks: Convergence Analysis and Resource Allocation, *IEEE/ACM Transactions on Networking*, 2021]

Things in the Pipeline

– Online machine learning for real-time energy disaggregation for electricity distribution feeders
– Low latency smart meter (computation and communication) – implemented and tailored to end users
– Residential demand response/thermal loss minimization studies in Eco capsules
– Smart water systems management
– Mobile blockchain + edge computing on e-health applications
– Multiple data sources (visible/infrared/electric sensing) integration/analytic for the pitfalls detection of PV panels

The University of Sydney





Thank You

Questions?

THE UNIVERSITY OF SYDNEY