Machine Learning for Smart Industry Applications

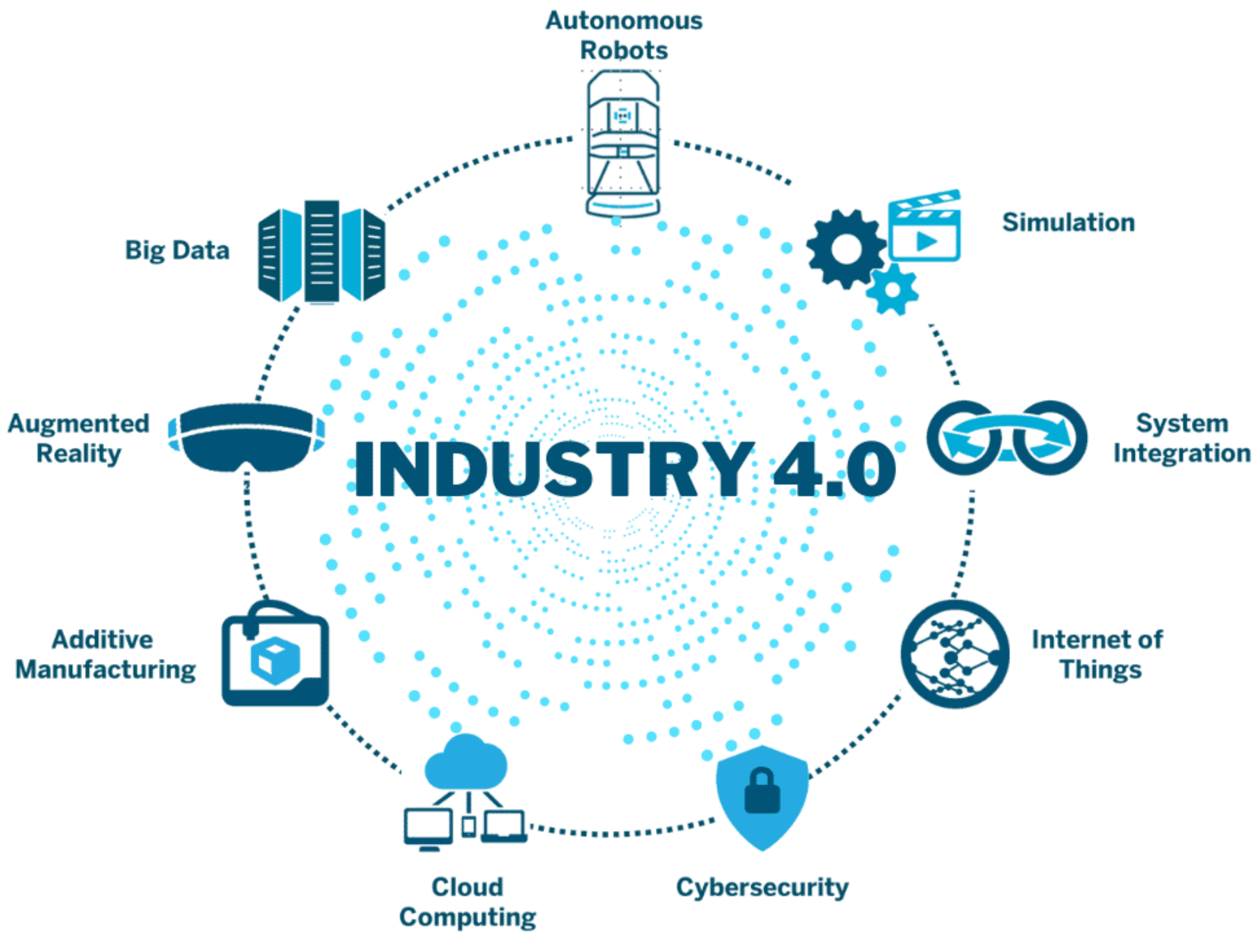
Azqa Nadeem

a.nadeem@utwente.nl

Today we will cover…

- Challenges for Industry 4.0
- Overview of Machine Learning (ML)
- ML models for smart industry applications

Industry 4.0 – Smart automation and more



https://www.calsoft.com/what-is-industry-4-0/

Industry 4.0 – Smart automation and more

- Blurs the line between information technology (IT) and operational technology (OT) to bring digital transformation to the factory floor

- IT
  - Computer equipment, networks, software, middleware systems, …
- OT
  - Sensors, monitors, actuators, generators, programmable logic controllers (PLCs), industrial robots, …

https://www.fortinet.com/resources/cyberglossary/it-vs-ot-cybersecurity

[IT+OT] Challenges

- Asset integration challenges
  - IT and OT operate on fundamentally different infrastructures – system heterogeneity

- Increased security threats
  - OT traditionally does not support security tooling – IT + OT vulnerabilities
  - Stuxnet: https://www.youtube.com/watch?v=DDH4m6M-ZIU

- Increased data volume
  - OT generates massive amounts of data – diverse/unstructured data

[IT+OT] Challenges (Cont.)

- Data imbalance
  - Most data pertains to benign activities

- Patterns over long intervals
  - OT malfunction/attacks may be spread over several weeks or months

- Talent/skills gap
  - IT+OT management requires highly specialized personnel

Smart automation via Machine Learning

- Manual monitoring and maintenance no longer feasible

- Paradigm shift towards data-driven approaches
  - Data is the new gold

- Machine learning as a potential solution

Machine learning – A recap

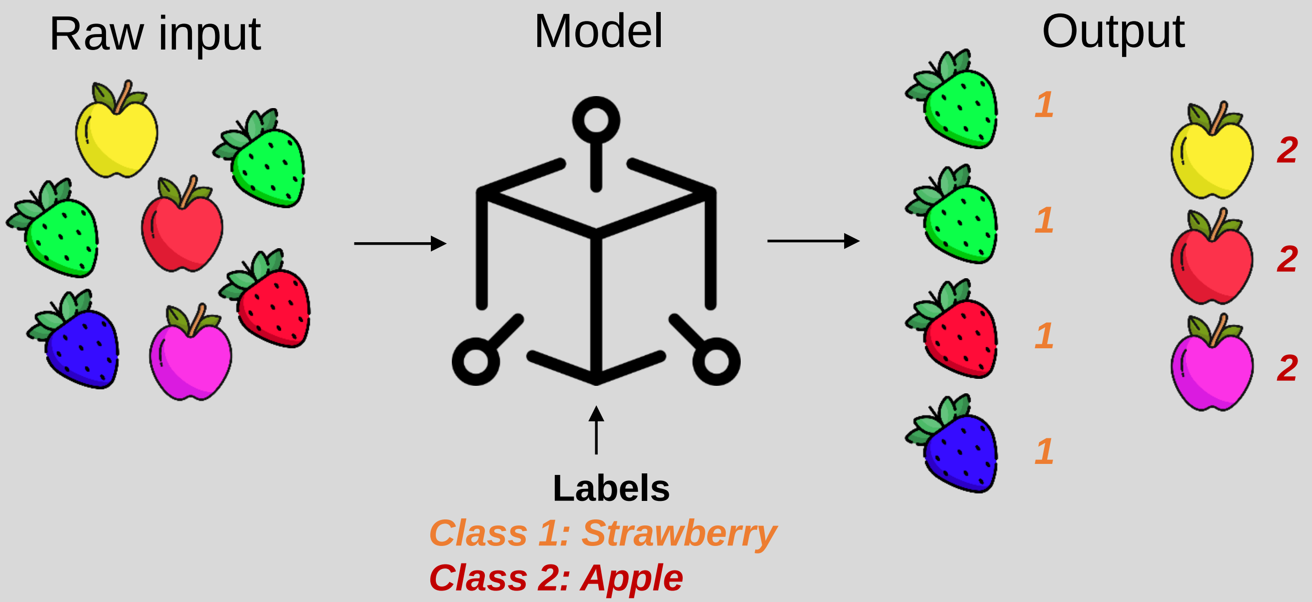
- Machine learning is a field of study that develops statistical models that learns from data and generalizes to unseen data without explicitly programming them

- Requires
    - Data (+ Labels) | Objective function | Evaluation criteria

- Paradigms
    - Supervised learning | Unsupervised learning | Reinforcement learning

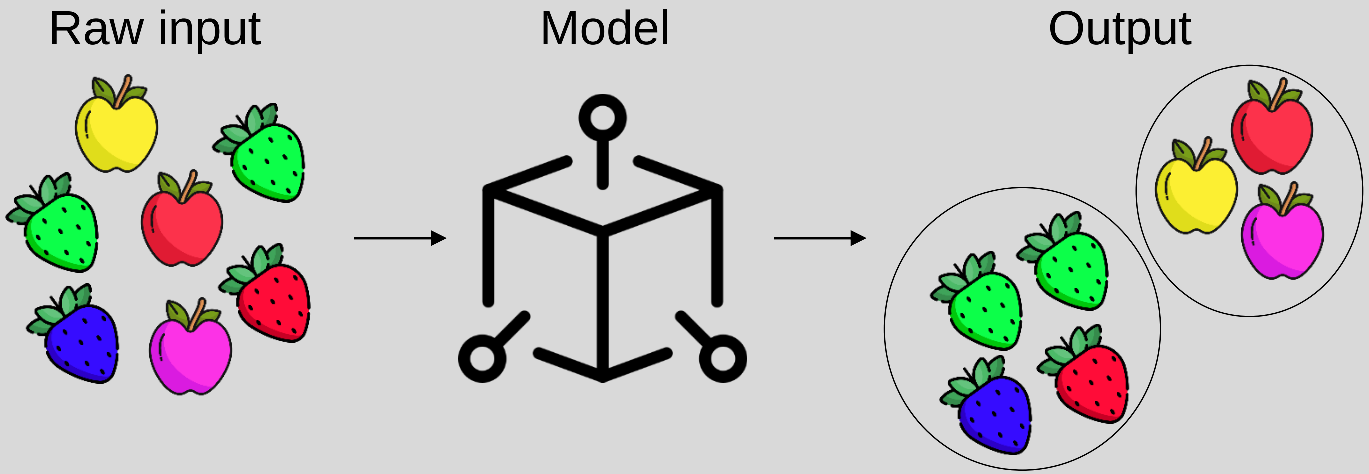
Machine learning – A recap (Cont.)

- Supervised learning → Classification/Regression
  - Learn a decision boundary for classifying/predicting data w.r.t. labels
  - e.g., Support Vector Machine, Decision trees, CNN, …
  - Evaluate using accuracy, F1-score, AUC, …



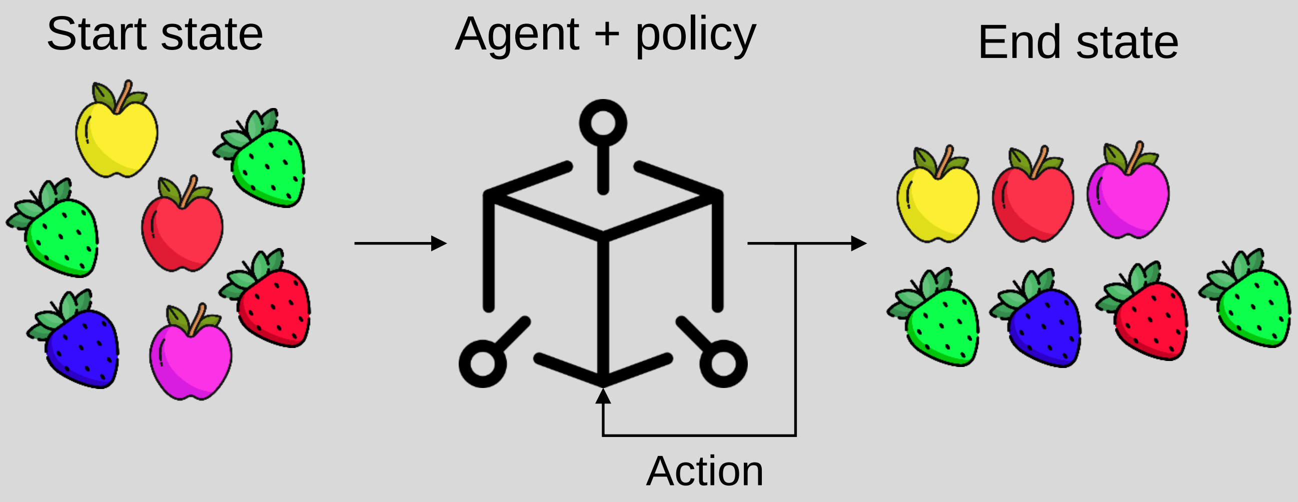
Machine learning – A recap (Cont.)

- Unsupervised learning → Pattern discovery
  - Discover distinct patterns/groups in data w/o labels
  - e.g., Clustering, association rules, auto-encoders, …
  - Evaluate using entropy, cohesion/separation, support, *qualitative*, …

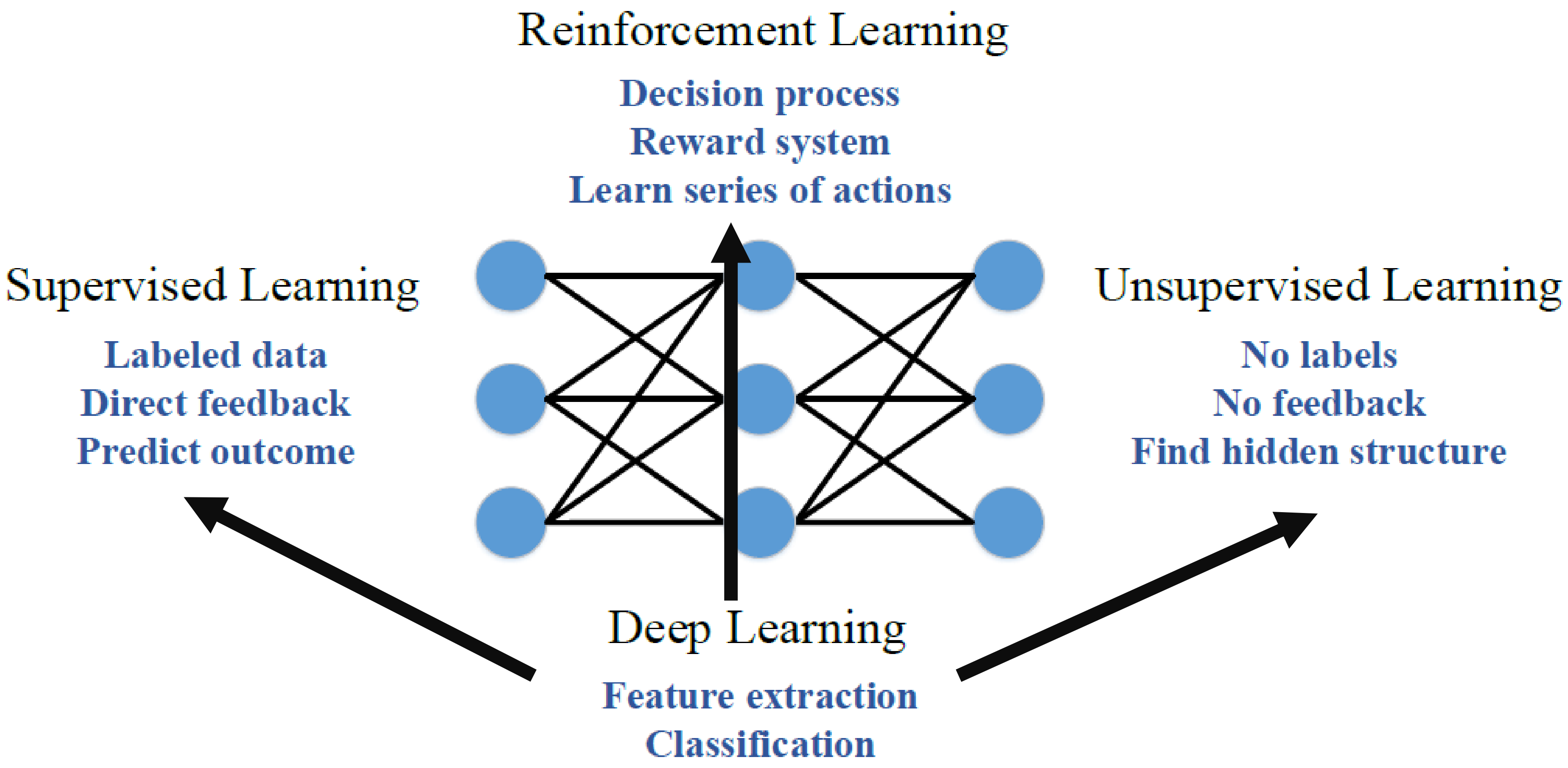


Machine learning – A recap (Cont.)

- Reinforcement learning → Sequential decision making
  - Agent gets from start state to end state by earning rewards
  - e.g., Q-learning, SARSA, DQN, …
  - Evaluate using convergence rate, reward, stability, …



Machine learning – A recap (Cont.)



Angelopoulos, Angelos, et al. "Tackling faults in the industry 4.0 era—a survey of machine-learning solutions and key aspects." *Sensors* 20.1 (2019): 109.

Using Machine Learning for Industry 4.0

- ML-assisted monitoring/maintenance/prevention/optimization, …
  - Using ML to address challenges of industry 4.0

- Concrete problem definition
  - Collect data via sensors, logs, side-channels, … (Labels??)
  - Choose suitable ML algorithm
  - Choose applicable evaluation criteria

- [Caveat] Model output dependent on data quality

https://www.calsoft.com/what-is-industry-4-0/

Data properties

- Volume: Size of data being generated
- Velocity: Speed at which data is generated
- Variety: Diversity of data instances or data sources
- Veracity: Accuracy/precision of the generated data
- Volatility: Timeliness/freshness of the generated data

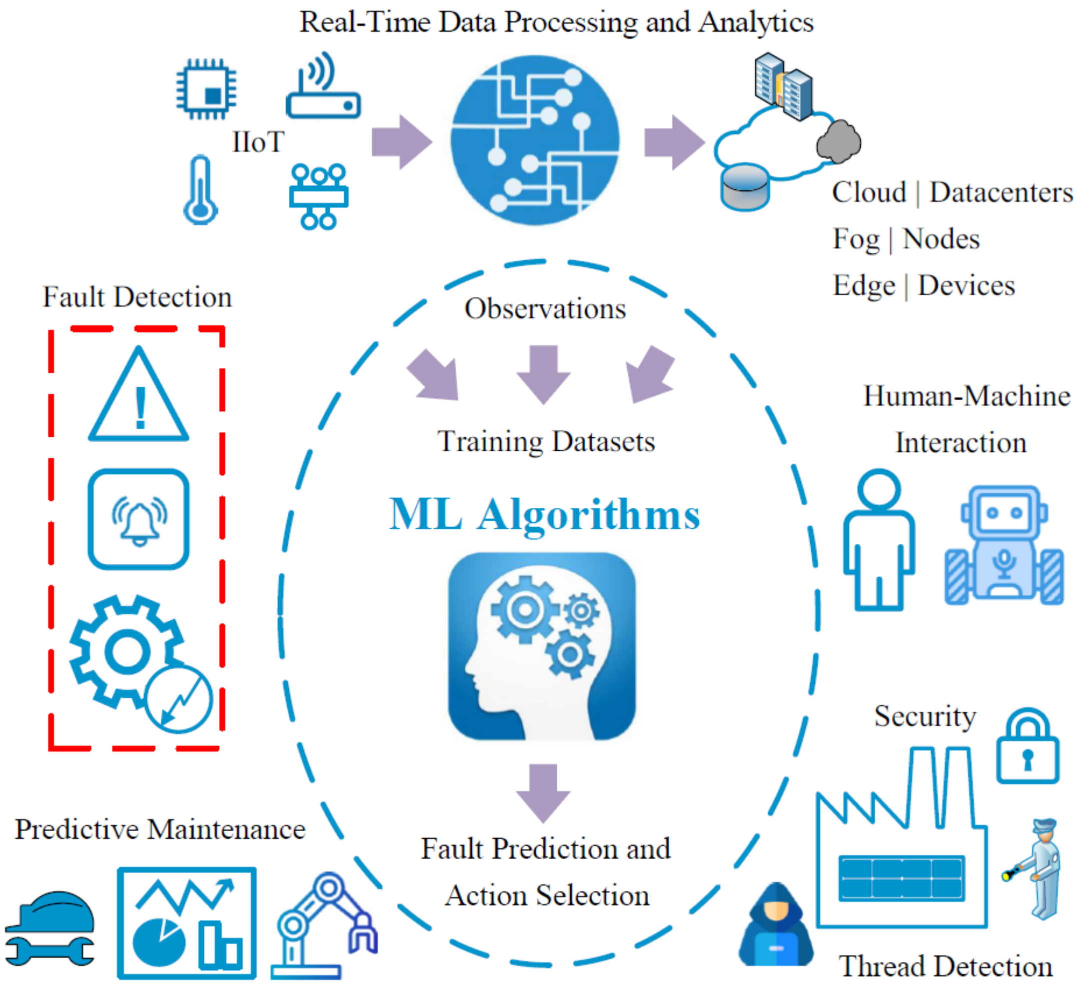
- Imbalance: Classes of vastly different sizes
- Noise: Links to Veracity
- Representativeness: Links to Variety
- Concept drift: Links to Volatility

Where can Machine Learning help?

- Remote monitoring/analytics
- Predictive maintenance
- Fault diagnosis
- Testing and quality control
- Process optimization
- Security threats

Exercise: Consider the case of a water filtration/treatment plant

Where can Machine Learning help?



Angelopoulos, Angelos, et al. "Tackling faults in the industry 4.0 era—a survey of machine-learning solutions and key aspects." *Sensors* 20.1 (2019): 109.

1) Remote monitoring/analytics

Analyze the performance of equipment remotely and in real-time

- Task:
- Data:
- Properties:
- Model:
- Evaluation:

1) Remote monitoring/analytics

Analyze the performance of equipment remotely and in real-time

- Task: *Record and display analytics from various plant modules*
- Data: Sensor/actuator readings
- Properties: Diversity, completeness, volume, …
- Model: Anomaly detection, data summarization, behavior discovery
- Evaluation: Ground truth? Internal metrics? Qualitative?

2) Predictive maintenance

Detect issues in equipment before they lead to failures for scheduling maintenance tasks

- Task:
- Data:
- Properties:
- Model:
- Evaluation:

2) Predictive maintenance

Detect issues in equipment before they lead to failures for scheduling maintenance tasks

- Task: *Monitor plant modules over time to predict wear & tear*
- Data: Sensor/actuator readings from plant modules
- Properties: Timeseries, imbalance, concept drift
- Model: Anomaly detector, Regressor, e.g., SVM
- Evaluation: FPR, PR curve, prediction accuracy

3) Fault diagnosis

Root cause analysis of system failures

- Task:
- Data:
- Properties:
- Model:
- Evaluation:

3) Fault diagnosis

Root cause analysis of system failures

- Task: *Reverse engineer the origin of a plant module failure*
- Data: List of failures, sensor readings over time
- Properties: Forensic analysis, diversity, noise
- Model: Clustering, LSTM & attention
- Evaluation: Top-k accuracy, distance from origin, …

4) Testing and quality control

Validating the correct operation of equipment against specifications

- Task:
- Data:
- Properties:
- Model:
- Evaluation:

4) Testing and quality control

Validating the correct operation of equipment against specifications

- Task: *Validate that the plant modules work as specified*
- Data: Sensor/actuator readings from plant modules, specifications
- Properties: Completeness, volume, concept drift (?)
- Model: Automata or RNN + Model checking
- Evaluation: Number of violations

5) Process optimization

Optimizing the consumption of resources and process scheduling

- Task:
- Data:
- Properties:
- Model:
- Evaluation:

5) Process optimization

Optimizing the consumption of resources and process scheduling

- Task: *Make a plan for the optimal quantities of raw material needed*
- Data: Action/value/rewards specification, start state, end state
- Properties: Clarity of problem specification and end goal
- Model: Q-learning or SARSA
- Evaluation: Convergence time, plan size, plan stability

6) Security threats

Detecting and mitigating security threats faced by industrial equipment

- Task:
- Data:
- Properties:
- Model:
- Evaluation:

6) Security threats

Detecting and mitigating security threats faced by industrial equipment

- Task: *Detect deviation in plant's behavior due to a cyber attack*
- Data: Sensor/actuator readings, network traffic, …
- Properties: imbalanced, concept drift, noise vs. diversity
- Model: Anomaly detector, LSTM, decision tree
- Evaluation: Time to detect, FPR, PR curve

Impact of Computing Architectures

Determines the amount of data generated and level of remote monitoring possible

- Cloud computing
  - Centralized model where all computing happens on the cloud
- Fog computing
  - Distributed architecture that adds a layer between edge and cloud with additional computing resources
- Edge computing
  - Decentralized model where computing happens closest to the source

Impact of Computing Architectures

- Limited device resources or central compute option
  - Opt for cloud computing

- Low latency or custom compute on each device
  - Opt for edge computing

- Offload some compute from device
  - Opt for fog computing

Example: A cloud-assisted smart factory

Application layer

ML-assisted predictive maintenance, context-aware services, process optimization, analytics, …

Cloud layer

ML-assisted resource management, data processing, …

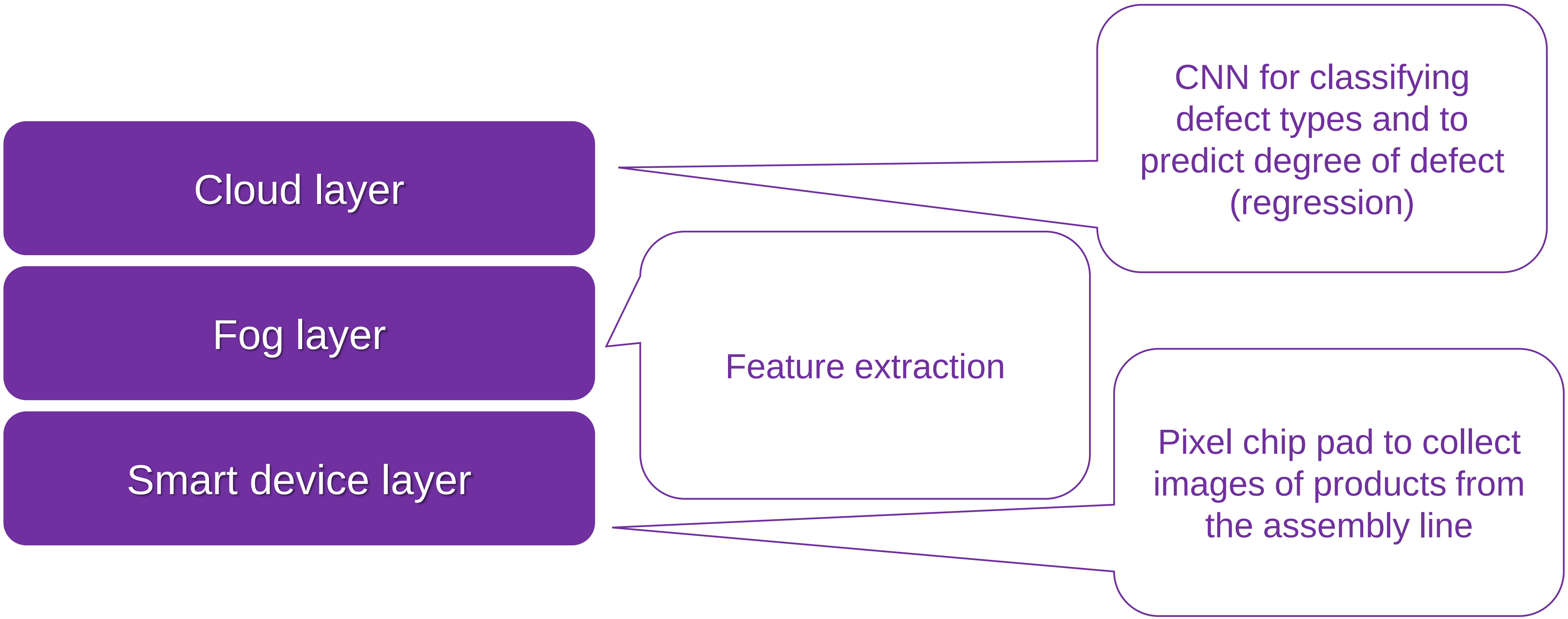
Network layer

ML-assisted communication optimization – device to device, device to cloud, load balancing, resource allocation, …

Smart device layer

Data collection, smart robotic arms, automated guided vehicles, computer vision, path planning, …

Example: A fog-based defect detector



Takeaways

- Industry 4.0 blurs the line between IT and OT
- Machine learning facilitates Industry 4.0
  - Monitoring, predictive maintenance, fault detection, security threats, …
- Design considerations for smart industrial applications on case-by-case basis
  - Choice of data, model, evaluation, …
- Computing architecture determines data granularity for models

Reading material

- Angelopoulos, Angelos, et al. "Tackling faults in the industry 4.0 era—a survey of machine-learning solutions and key aspects." Sensors 20.1 (2019): 109.
- Candanedo, Inés Sittón, et al. "Machine learning predictive model for industry 4.0." Knowledge Management in Organizations: 13th International Conference, KMO 2018, Žilina, Slovakia, August 6–10, 2018, Proceedings 13. Springer International Publishing, 2018.
- Rai, Rahul, and Chandan K. Sahu. "Driven by data or derived through physics? a review of hybrid physics guided machine learning techniques with cyber-physical system (cps) focus." IEEE Access 8 (2020): 71050-71073.
- https://www.fortinet.com/resources/cyberglossary/it-vs-ot-cybersecurity

Questions?