UNIVERSITEIT
iYUNIVESITHI
STELLENBOSCH
UNIVERSITY

100

1918 · 2018

forward together · saam vorentoe · masiye phambili

Machine Learning Applications in Minerals Engineering

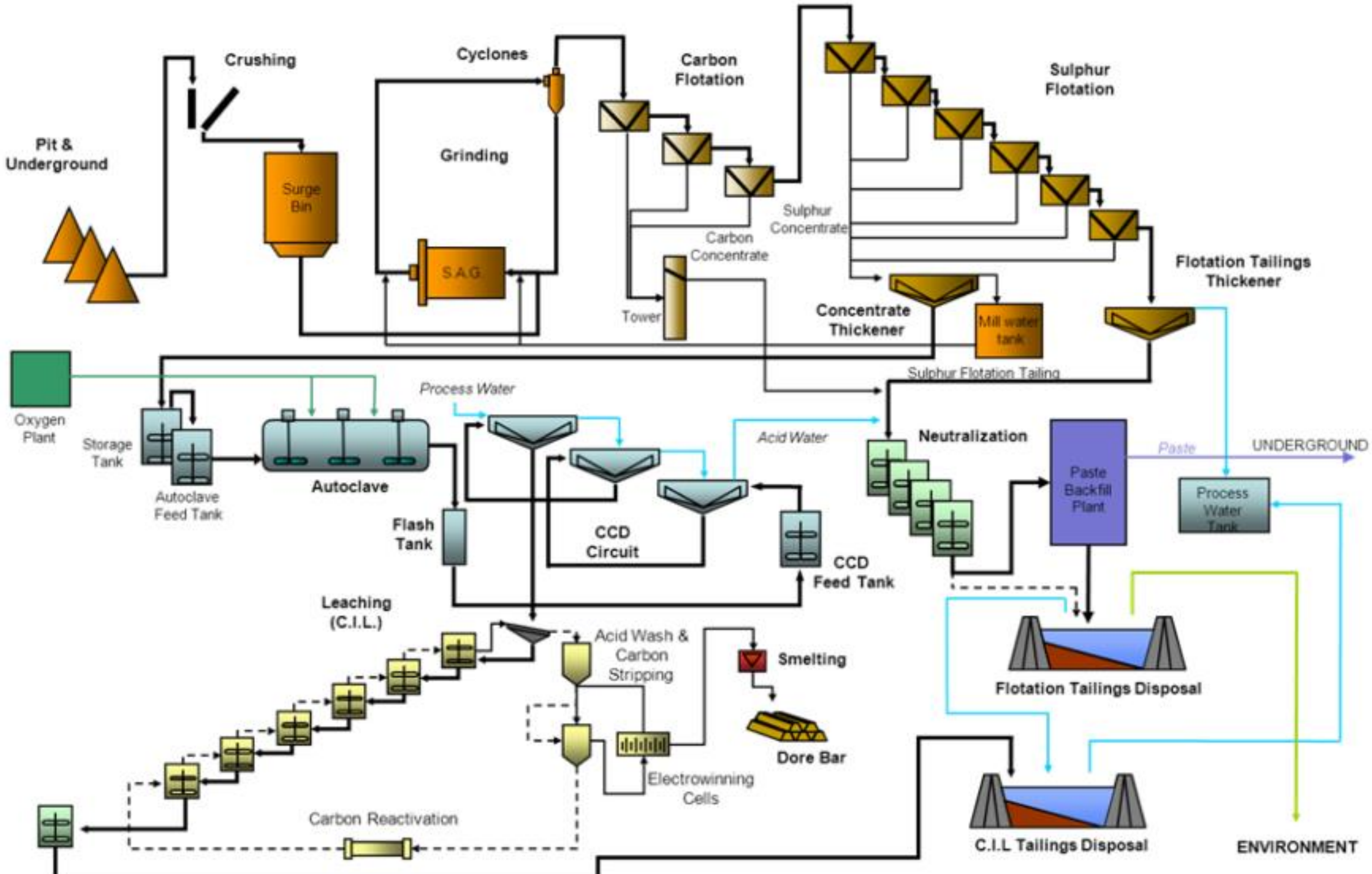
Lidia Auret, John McCoy

Department of Process Engineering, Stellenbosch University, South Africa

IFAC MMM2018, 23-25 August 2018

Mineral processing challenges

Continuous, connected, controlled, circulating, complex, changing



Mineral processing data

- **Online data**
  - Physical property sensors (~seconds)
    - E.g. mass flow rate, density, temperature, pressure
  - Image data (~minutes)
    - E.g. rocks on conveyor belts, flotation froth
- **Offline data**
  - Laboratory data (~hours)
    - E.g. metal content, particle size distribution
  - Image data (~days)
    - E.g. microscopic grain shape and colour
  - Text data (~days)
    - E.g. maintenance logs, metallurgist reports

Machine learning definitions

$$\boldsymbol{Y} = f(\boldsymbol{X}, \boldsymbol{\theta})$$

$\boldsymbol{Y} = output\ variables$	$Y \sim continuous; Y \sim categorical$
$\boldsymbol{X} = input\ variables$	$X \sim continuous\ X \sim categorical$
$\boldsymbol{\theta} = parameters$	$\theta_m = model\ parameters;$
	$\theta_h = hyperparameters$
$f(\quad) = functional\ form$	$Parametric, e.g. linear\ regression$
	$Non-parametric, e.g. neural\ nets$

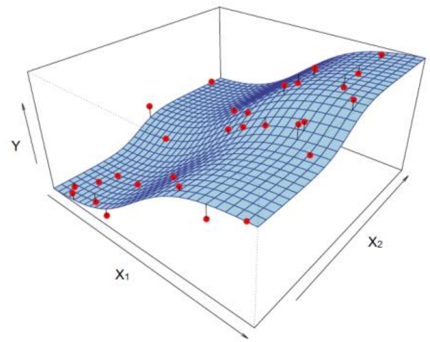
Training: Learn θ_m (e.g. minimize $\sum_i \left(Y_i - \hat{Y}_i\right)^2$)

Validation: Learn θ_h

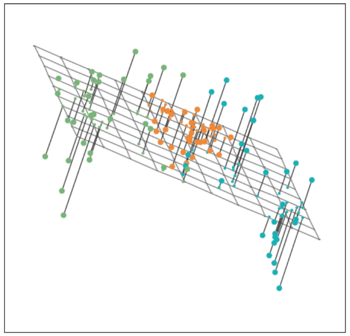
Machine learning definitions

$$Y = f(X, \theta)$$

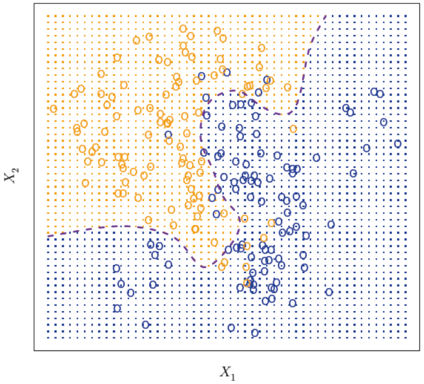
Supervised learning	Unsupervised learning



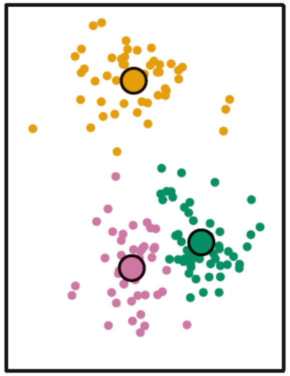
Regression
$Y \sim continuous$



Noise removal, feature extraction
$\hat{X} = f(X, \theta)$
$T = f(X, \theta)$

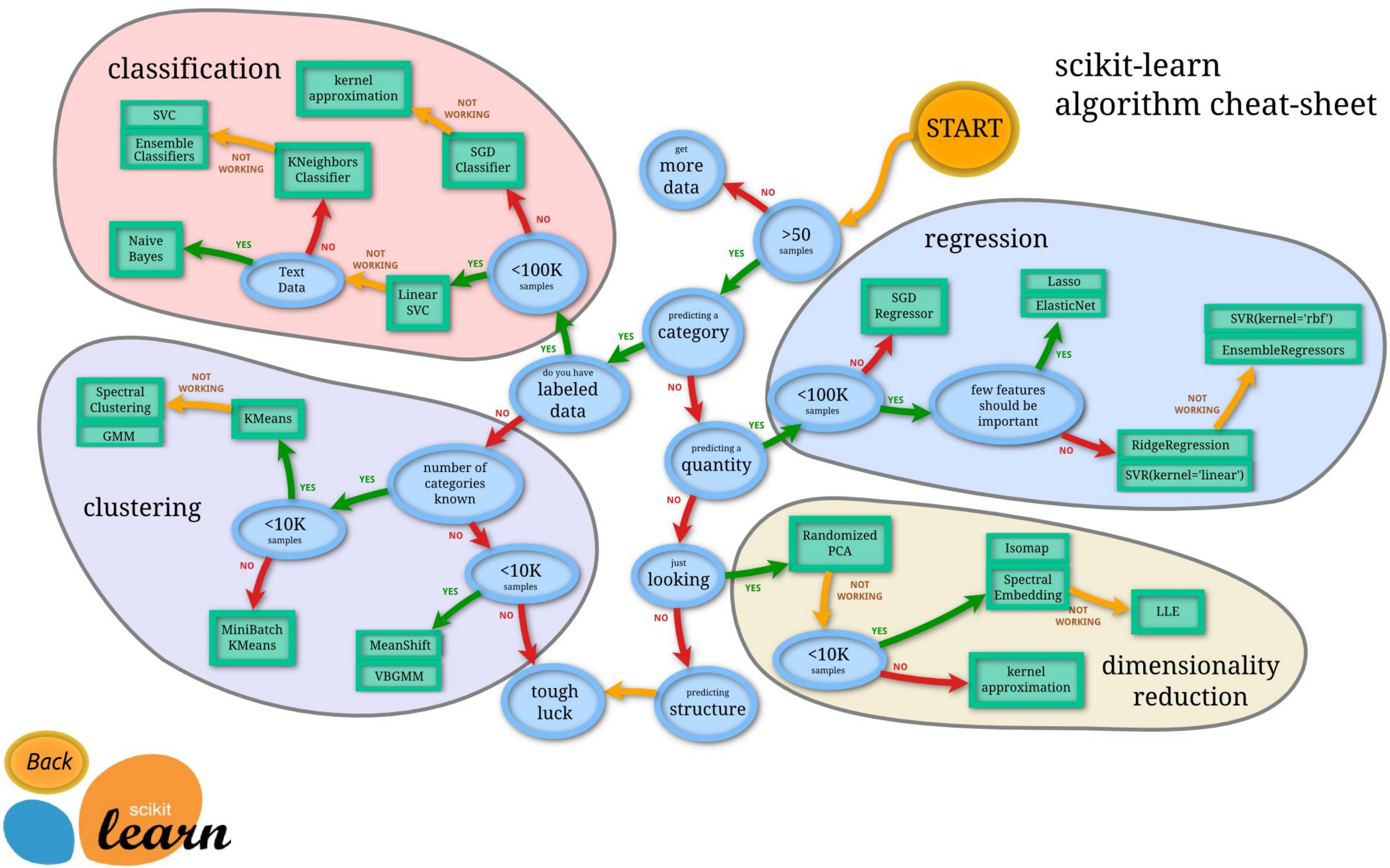


Classification
$Y \sim categorical$



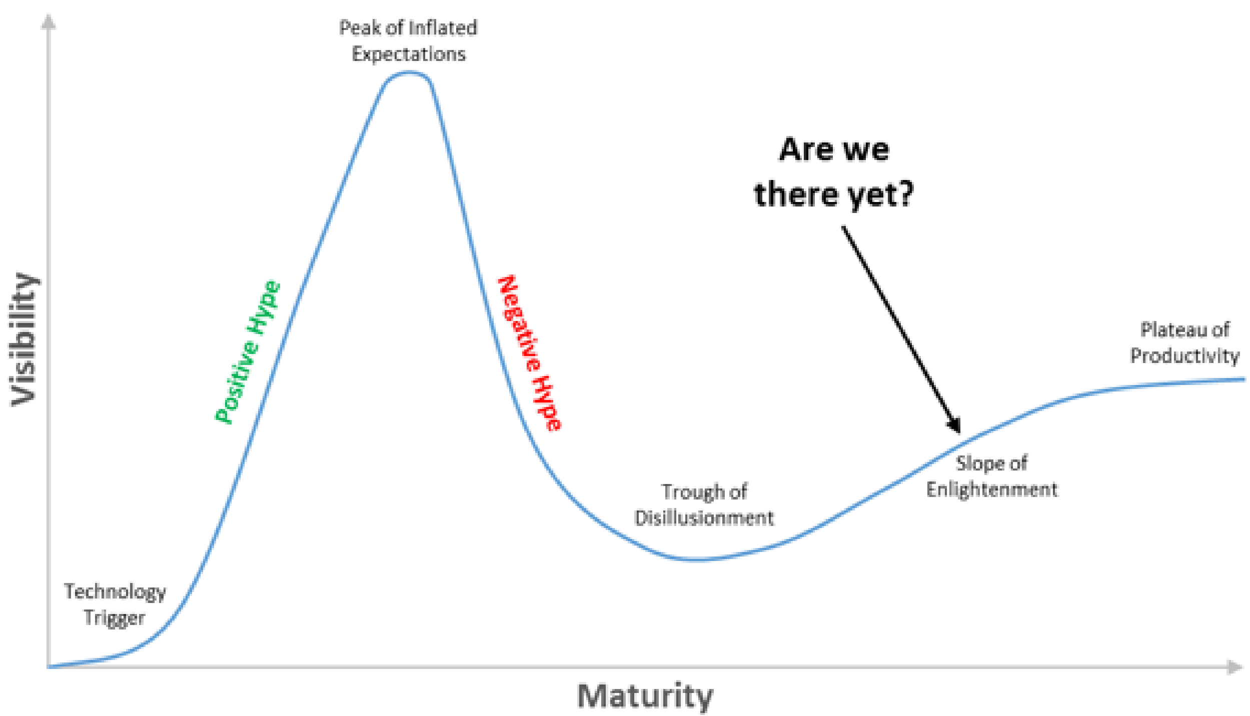
Clustering
$C = f(X, \theta)$

Machine learning definitions



Hype

BIG DATA! INDUSTRY 4.0! ARTIFICIAL INTELLIGENCE! (?)



$Y = f(X, \theta)$ More data + Better computers + Better methods

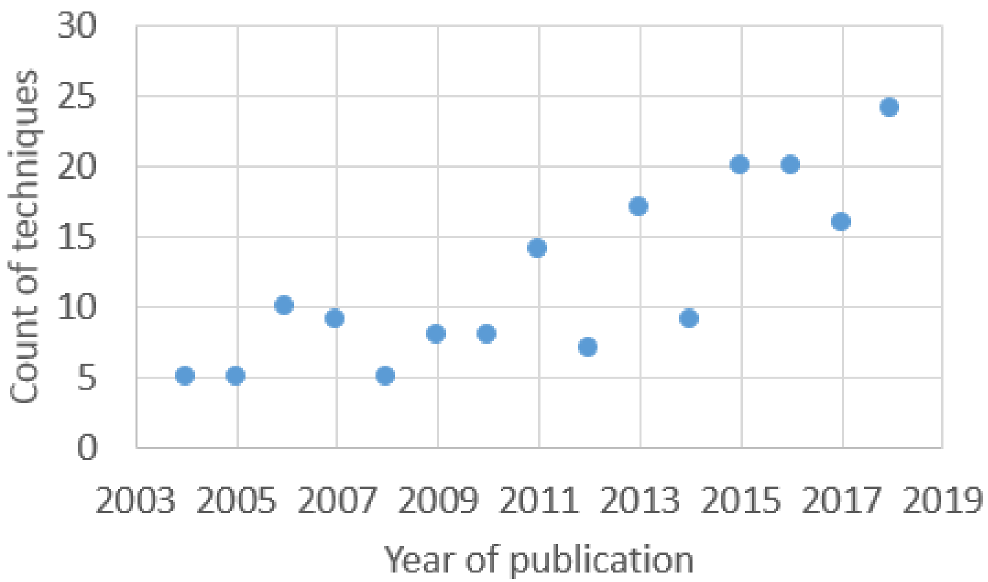
Hype

- Increasing popularity in many applied sciences
  - Special issues in journals of medicine, finance, environmental science, etc.

- Review undertaken: 13 journals and conference proceedings (2004 – 2018):
  - AIChE; Chemometrics and Intelligent Laboratory Systems; Computers and Chemical Engineering; Control Engineering Practice; Engineering Applications of Artificial Intelligence; Journal of Process Control; IFAC MMM; Industrial and Chemical Engineering Research; International Journal of Mineral Processing; International Journal of Mining, Reclamation and Environment; JSAIMM; Minerals and Metallurgical Processing; Minerals Engineering

- Tool for researchers: Searchable summaries
  - Category and application
  - Method, inputs, outputs, hyperparameters
  - Success and implementation

Hype



177 technique applications

Implementation	Count
Experimental data	105
Simulated data	8
Industrial data	40
Industrial implementation	24

Success	Count
Yes	141
Limited	35
No	1

Category	Count
Fault detection and/or diagnosis	30
Data-based modelling	40
Machine vision	107

Data-based modelling

Fast, easy measurements

eg flow rate
pressure
temperature
spectra

→

Supervised learning

Data-based modelling techniques

PCA, PCR, PLS
kNN, k-means clustering
CART, RF
SVM, SVR, ANN, ELM
QTA

→

Slow, difficult measurements

eg composition
size distribution
mill load
equipment failure

Trends

Soft sensors
Industrial implementation: relatively mature

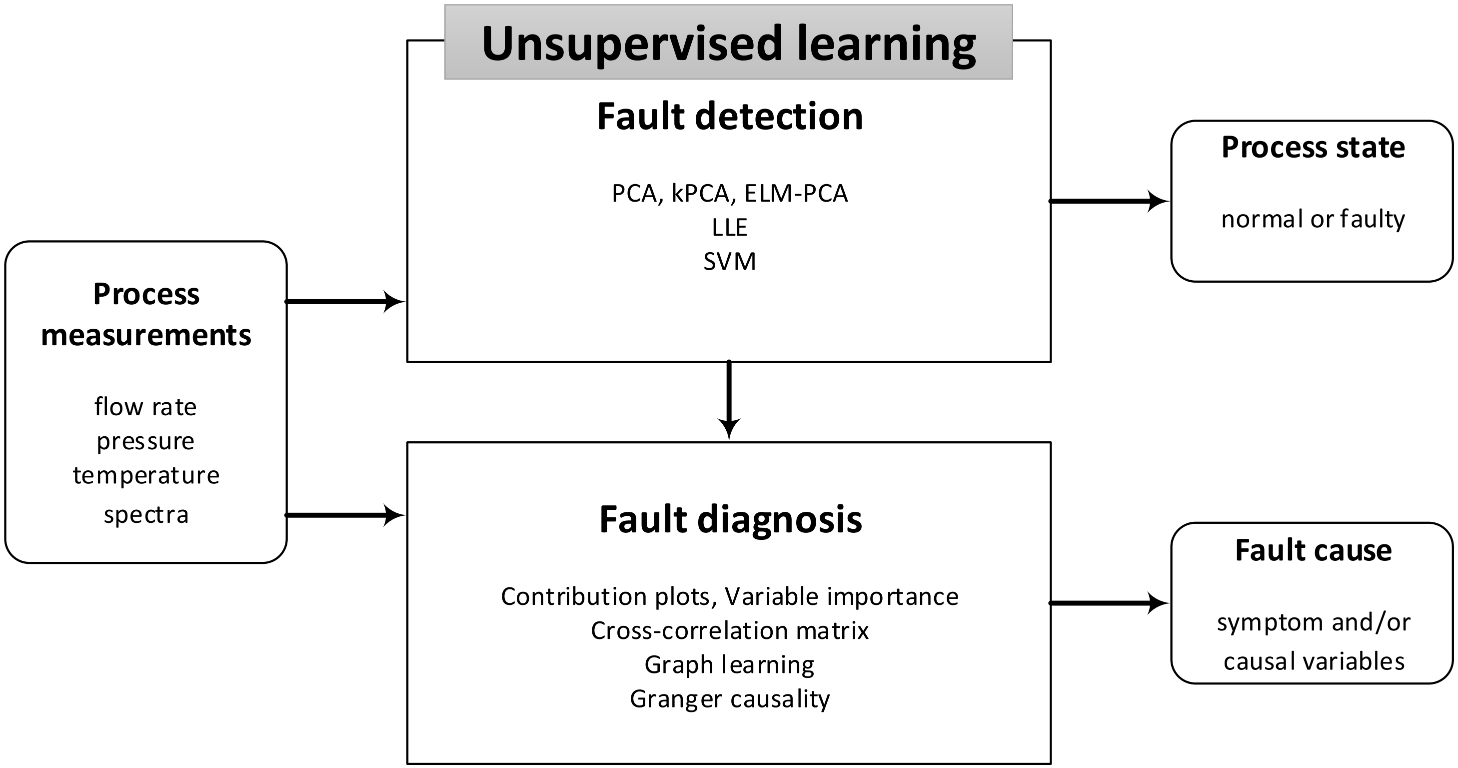
Challenges

Non-stationary
Nonlinear
Multiple modes
Missing values
Faulty and noisy values
Sampling times
Time lags
Correlation

Opportunities

Deep neural networks
Adaptive retraining
Qualitative trend analysis
Dynamic behaviour

Fault detection and diagnosis

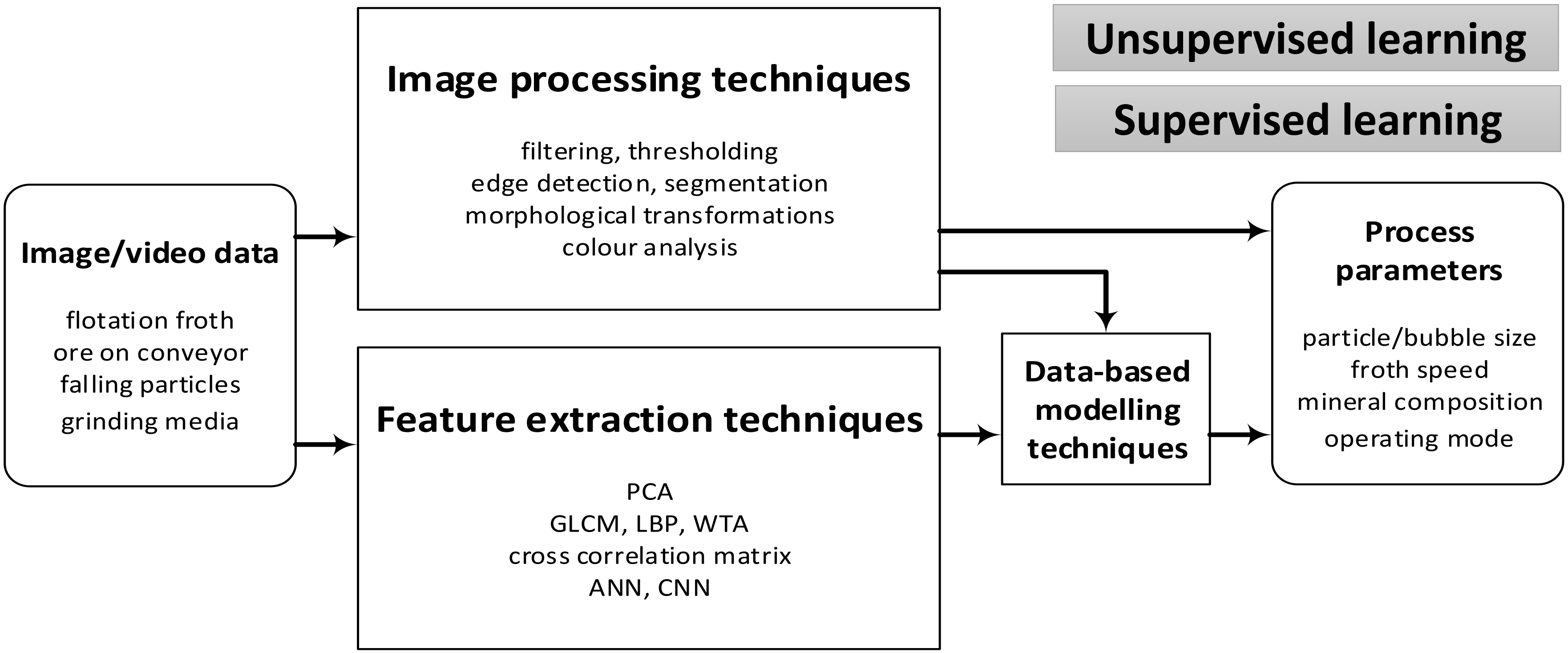


Trends
Industrial data (not implementation)
PCA and modifications

Challenges
Data scarcity
Benefit assessment
Implementation

Opportunities
Root cause analysis
Diverse data
Prognosis

Machine vision



Trends
Simple industrial implementation relatively mature
Texture features

Challenges
Surface-only
Limited state information
Labelled images

Opportunities
Crowd-sourced labelling
Deep nets
Diverse data

Golden rules

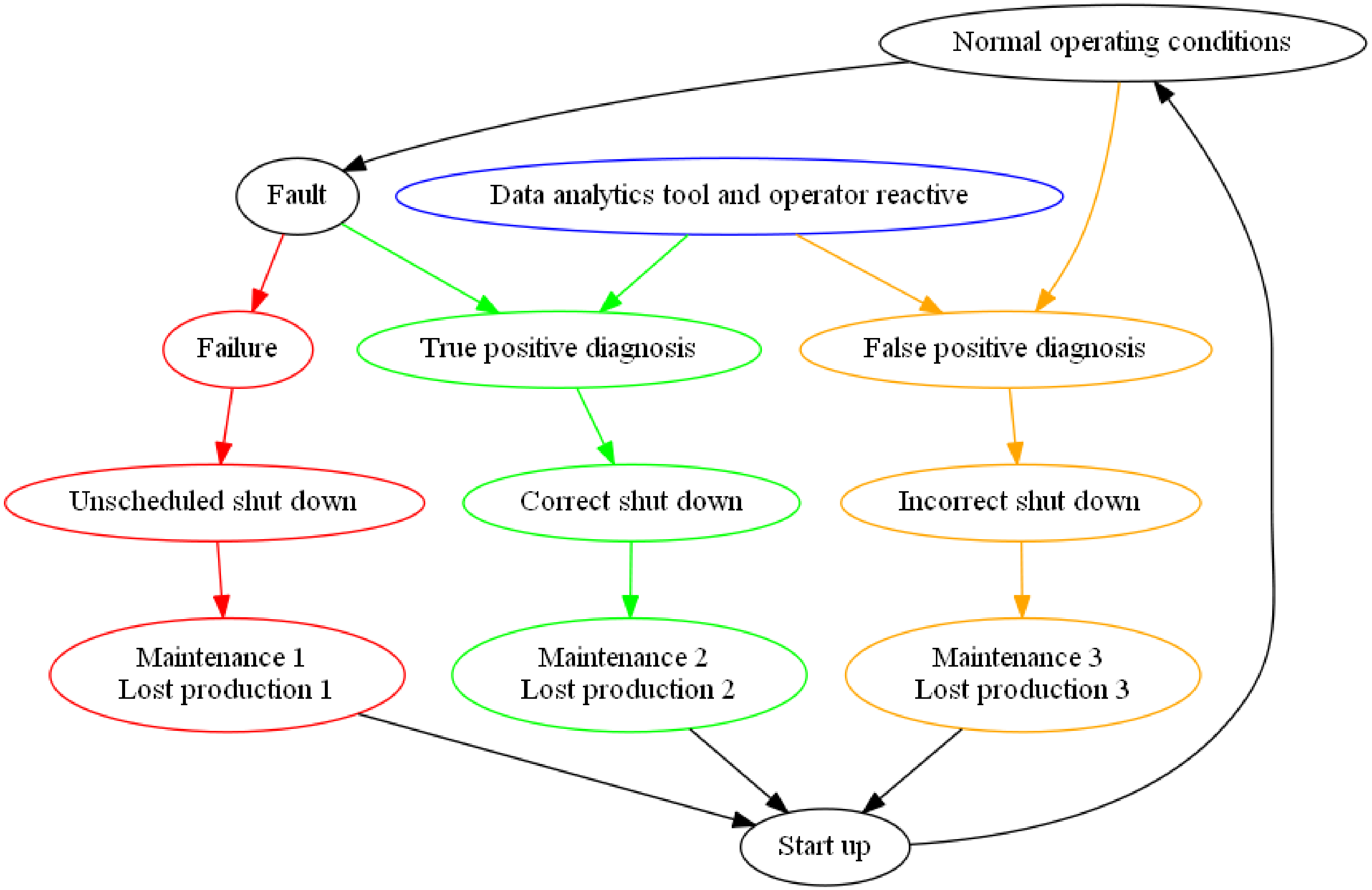
- **Hyperparameter sensitivity and guidelines**
  - Show sensitivity to hyperparameter selection
  - Guidelines relating hyperparameters to industrial context

- **Data diversity and explicit model validity**
  - Training data should include entire expectation of process data variation
  - Model predictions should include metric to indicate level of certainty / extrapolation

- **Comparison to simple and/or fundamental models**
  - Numerical motivation should be given for complex models
    - Compare to simpler techniques
  - Incorporate fundamental knowledge

Future directions

- **Build the business case**
  - Data-based modelling / machine vision: Similar to economic motivation for control
  - Fault detection and diagnosis: More complicated



Future directions

- **De-risk the method**
  - Thorough robustness analysis
  - Availability of benchmark industrial datasets
    - "UCI ML repository" archive.ics.uci.edu/ml for mineral processing
    - Control loop data repository: sacac.org.za/Resources
  - Availability of benchmark simulation datasets
    - "Tennessee Eastman process" for mineral processing
    - Simulation repository: github.com/ProcessMonitoringStellenboschUniversity

Future directions

- **Train the humans**
  - Engineers of today and tomorrow need to be data science literate
  - Not necessary to be an expert in machine learning
    - Basic understanding of goals and types
    - Ability to communicate requirements for new solutions
    - Ability to critically assess the results (check golden rules)
  - At undergraduate, postgraduate and professional levels
  - Challenge: Lack of domain-specific resources (e.g. examples, textbooks)
    - Good place to start: www.statlearning.com

Domain knowledge + Machine learning =
Better solutions

Questions?

www.ifacmmm2019.org