Interpretable Machine Learning at the EuXFEL

Amna Majid[1] **Danilo Ferreira de Lima**[1] with many collaborators from DA, CTRL, XPD, VAC, THE, SCS, SQS at the European XFEL, and CFEL at DESY

[1]European X-Ray Free Electron Laser

Machine Learning at the EuXFEL

- Automate data analysis activities.
- But ... not all approaches are equal.
- Users have the last word on how to do their experiments.

Solutions must conform to:

- *interpretability* → what do the results mean?
- *explainability* → methods science-aware?
- *quality control* → conditions for operation?

How to achieve it?

- Clarify how the method works.
- Shape methods based on scientific content.
- Estimate uncertainties and data quality.

Use-case: Enhancing non-invasive X-ray diagnostics

Two beam diagnostics devices in SASE3: Grating Spectrometer and Photo-Electron Spectrometer.

Grating Spectrometer (GS)

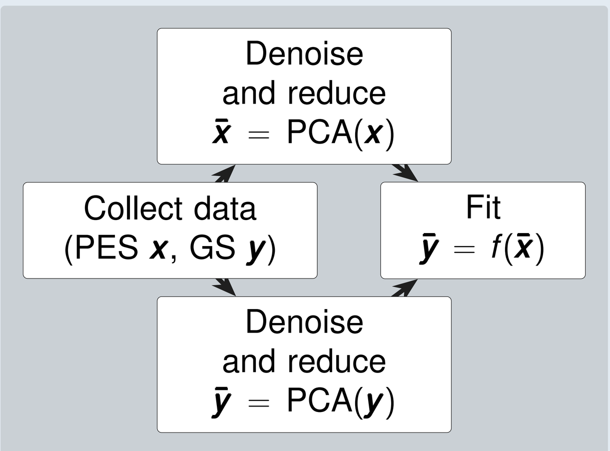
- High resolution.
- Simple calibration
- Invasive.
- Train-resolved.

Photo-Electron Spectrometer (PES)

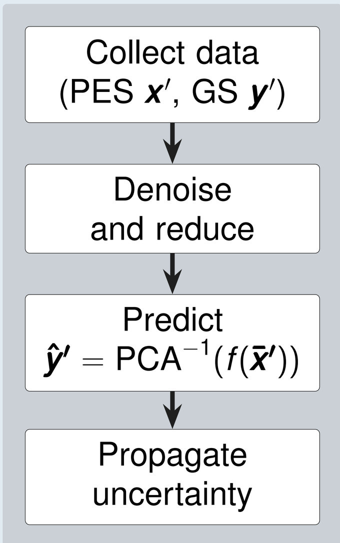
- Low resolution.
- Complex calibration.
- Non-invasive.
- Pulse-resolved.

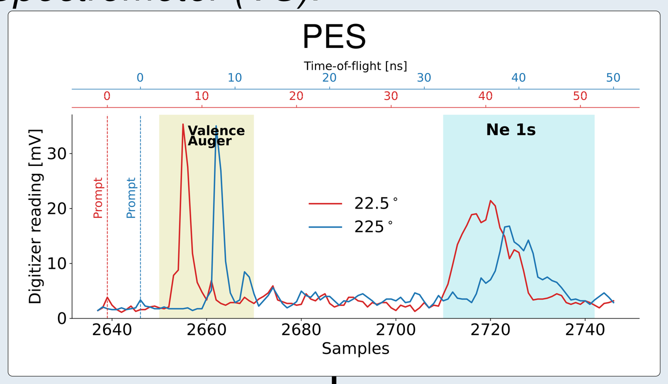
Idea: Map PES to GS to get best of both worlds → *Virtual Spectrometer (VS)*.

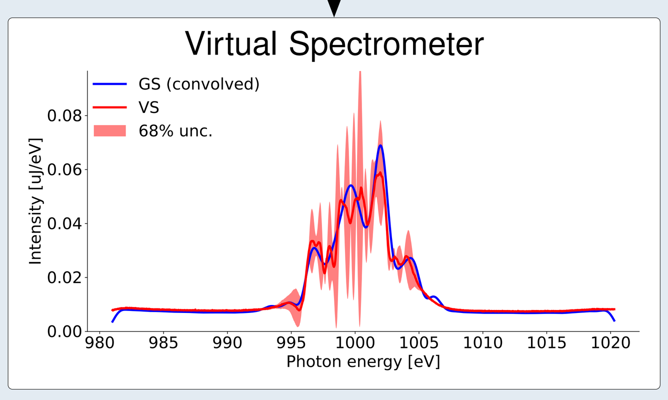
Training phase



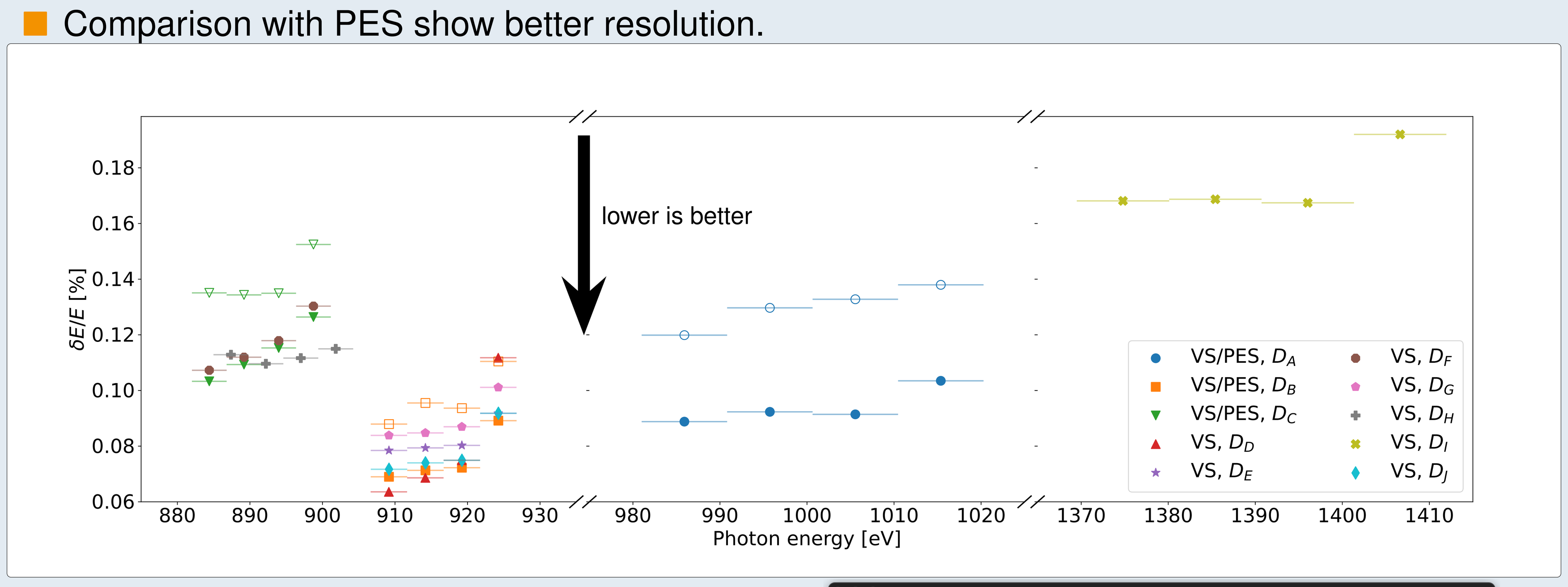
Inference phase



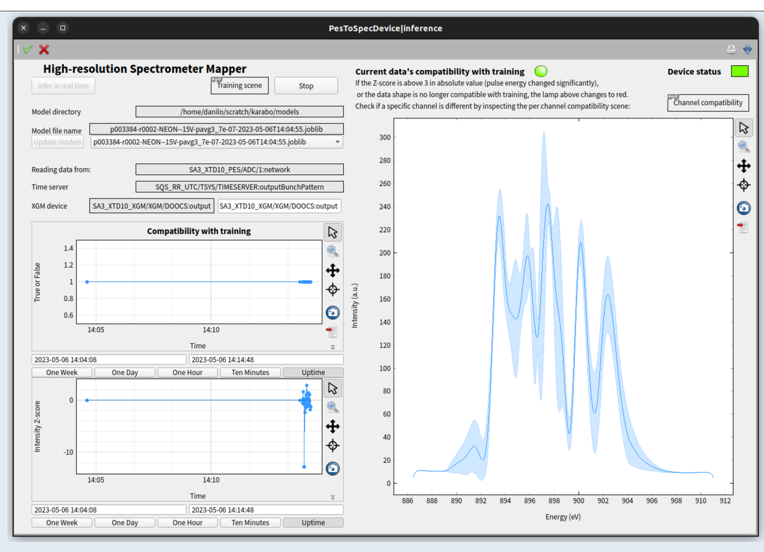




- Systematic resolution studies under several conditions done.
- Comparison with PES show better resolution.

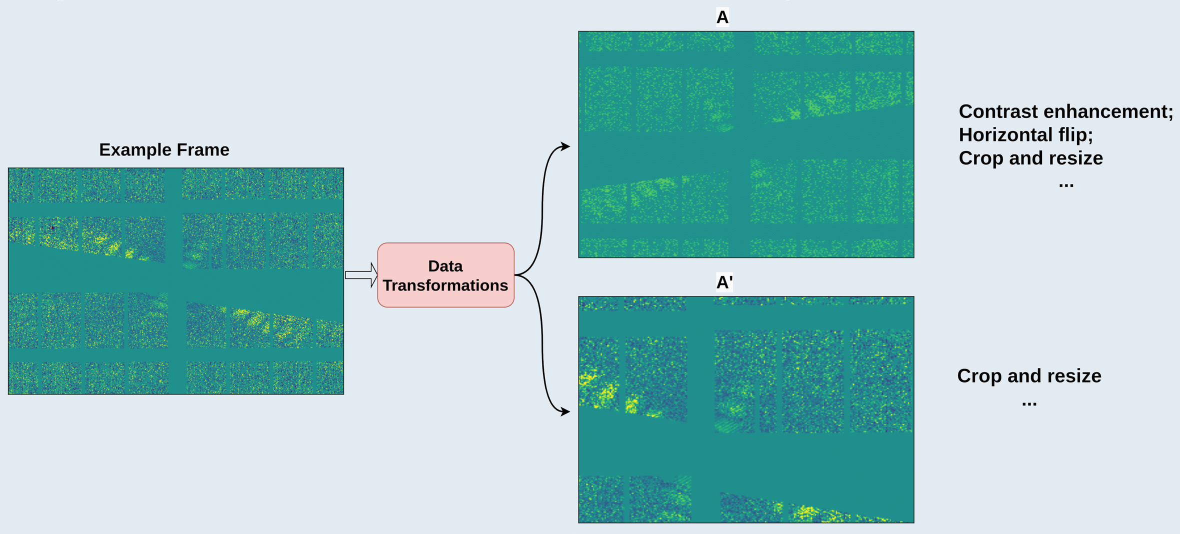


- Software deployed.
- Requires measurement conditions to be unchanged.
- *Quality control* and *explainability*:
  - Uncertainty informs on results quality.
  - Conditions are monitored and alarmed on.
  - Resolution estimates provided.
  - Interface informs on method and procedure.

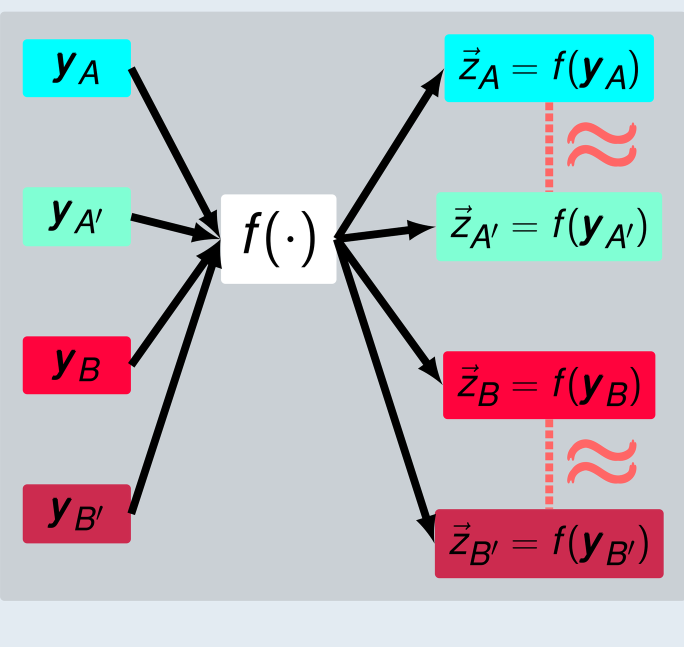


Use-case: How do we Google data?

- How can we make data findable as soon as we collect it?
- **Concept**: *Change* the data *view* and enforce their similarity.

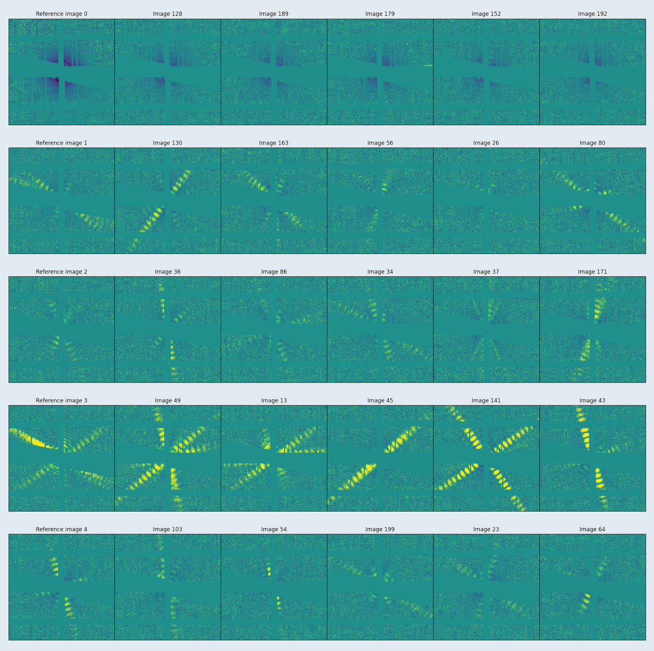


- *Equivalent views* ⇒ variations to ignore.



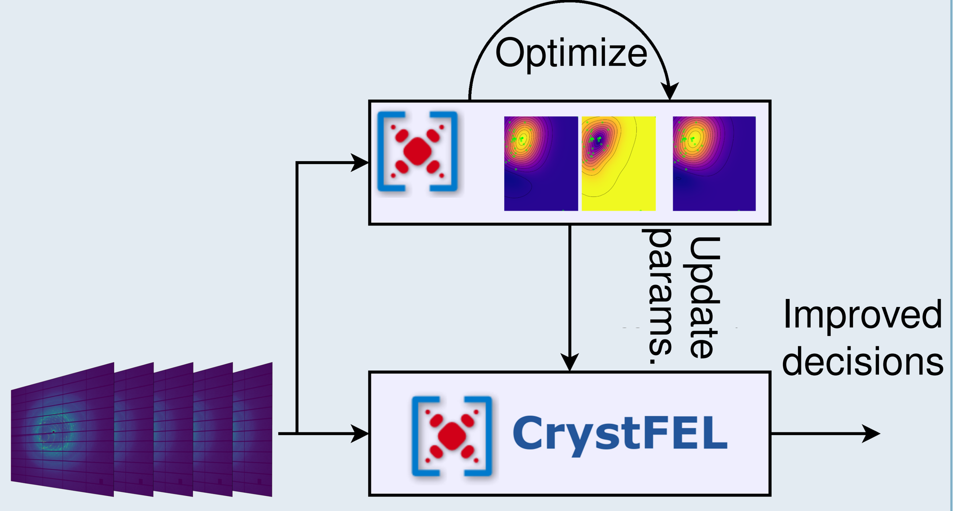
Example
⇓

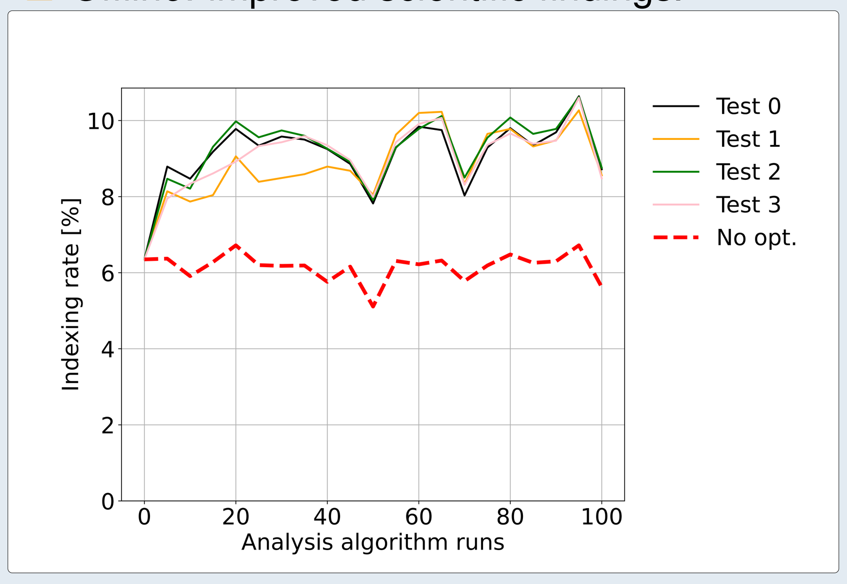
Similar ⇒
Similar ⇒
Similar ⇒
Similar ⇒
Similar ⇒

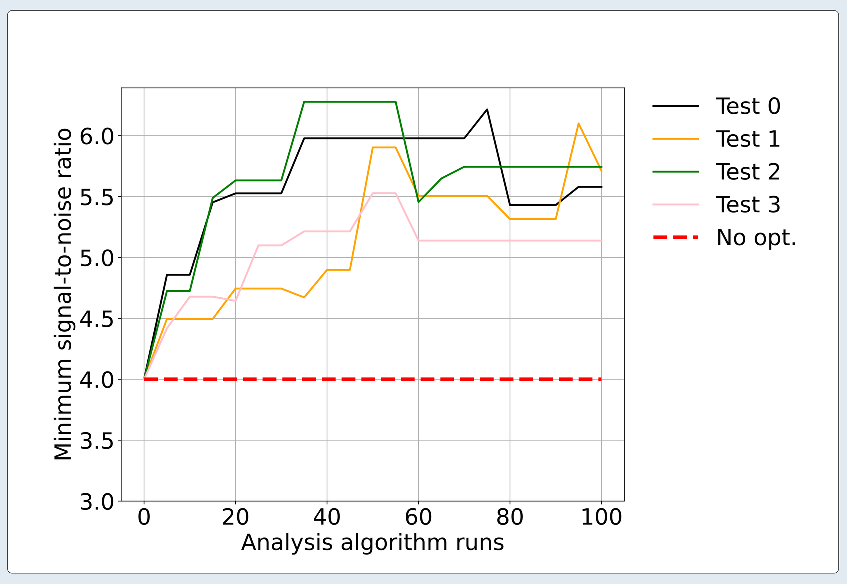


Use-case: Streamlining data analysis using ML

- Often data analysis pipelines have parameters.
- **Idea:** Simplify data analysis for non-experts.
- **Goal:** Tune parameters to maximize a *metric*.
- This example: maximize indexed frames fraction.
- *Online*: fast feedback, higher success chances.
- *Offline*: improved scientific findings.



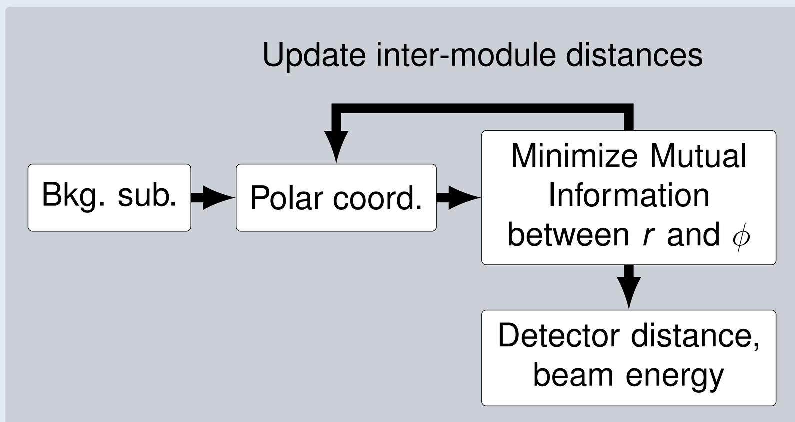


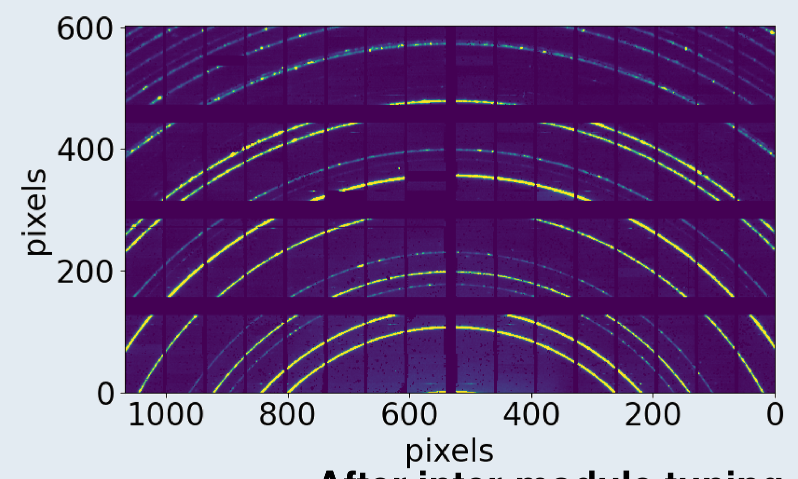


- Hen Egg-White (HEW) Lysozyme with the AGIPD detector at EuXFEL SPB/SFX.

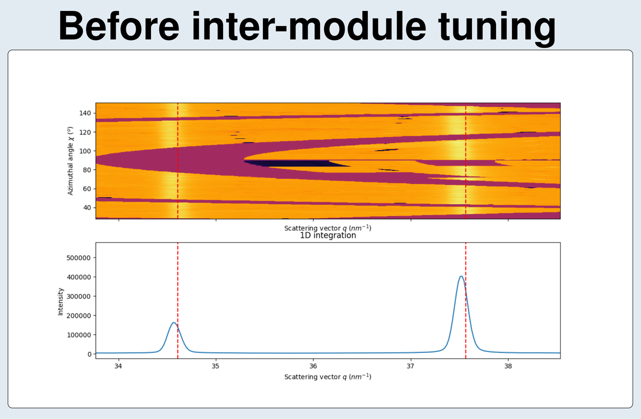
Use-case: Multi-modular geometry tuning

- Misalignment of module positions.
  - Manual alignment: requires lots of time.
  - Powder diffraction-based methods require many parameters and often manual tuning.
- Let's start with powder diffraction: can we improve and *automate* it?





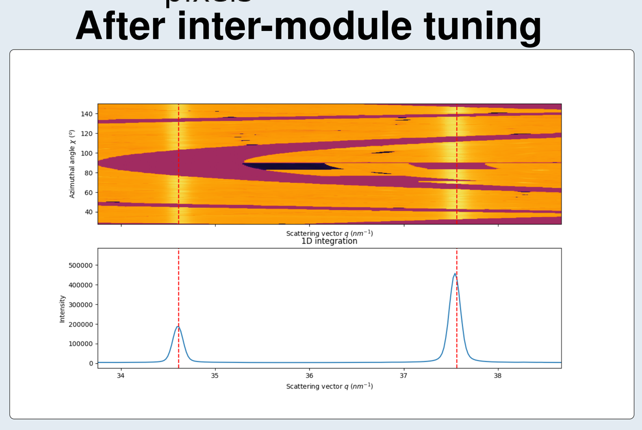
Before inter-module tuning



(Tuning only full detector position and rotation with pyFAI)

Impose independence between r and ϕ

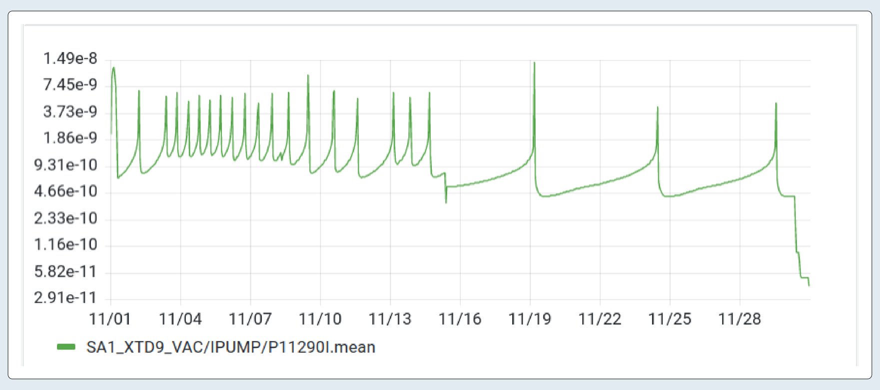
After inter-module tuning

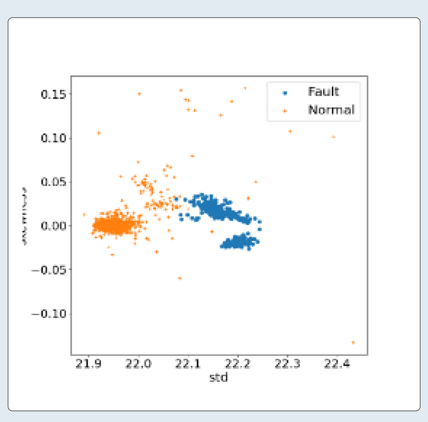


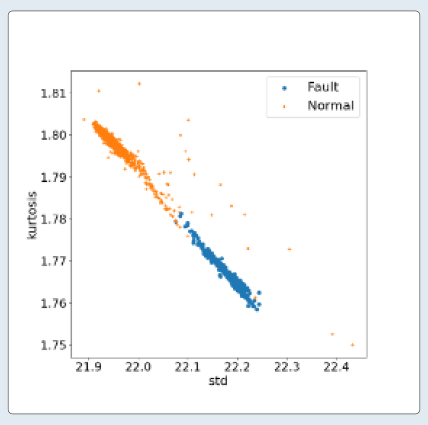
(Including post-processing of full detector position and rotation with pyFAI)

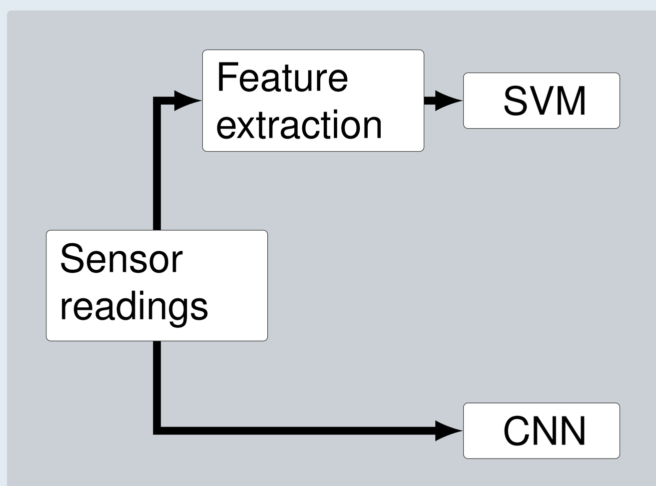
Use-case: Predictive Maintenance (in Ion Pumps)

- Faults may lead to loss of beam time.
- Important to detect them early.
- Difficulty: complex system makes it hard for humans to monitor everything.
- Example: Ion pump faults have lead to significant downtime.
- Detection mechanism: frequent surges in pressure level.









Method	Accuracy [%]	Precision	Recall
SVM	99.98	1.00	0.96
CNN	99.95	0.99	0.99

- Two methods researched with similar performance.
- SVM makes a linear cut in the feature space of peak characteristics → easy interpretation.
- CNN uses all information.
- Prefer interpretable method!
- Web interface under development.

Summary

- Several approaches to enhance automation at the EuXFEL being researched and developed.
- Control system allows for integration and deployable methods.
- Interpretability, explainability and quality control assets to guide towards adequate solutions.
- Interface design is simple, but highlights those characteristics to guide users.
- Aim for a holistic approach to integrate those features in all applications.

Have a look at our other Data Analysis posters in the 203 EuXFEL User Meeting!
More information: `http://www.xfel.eu/data_analysis`