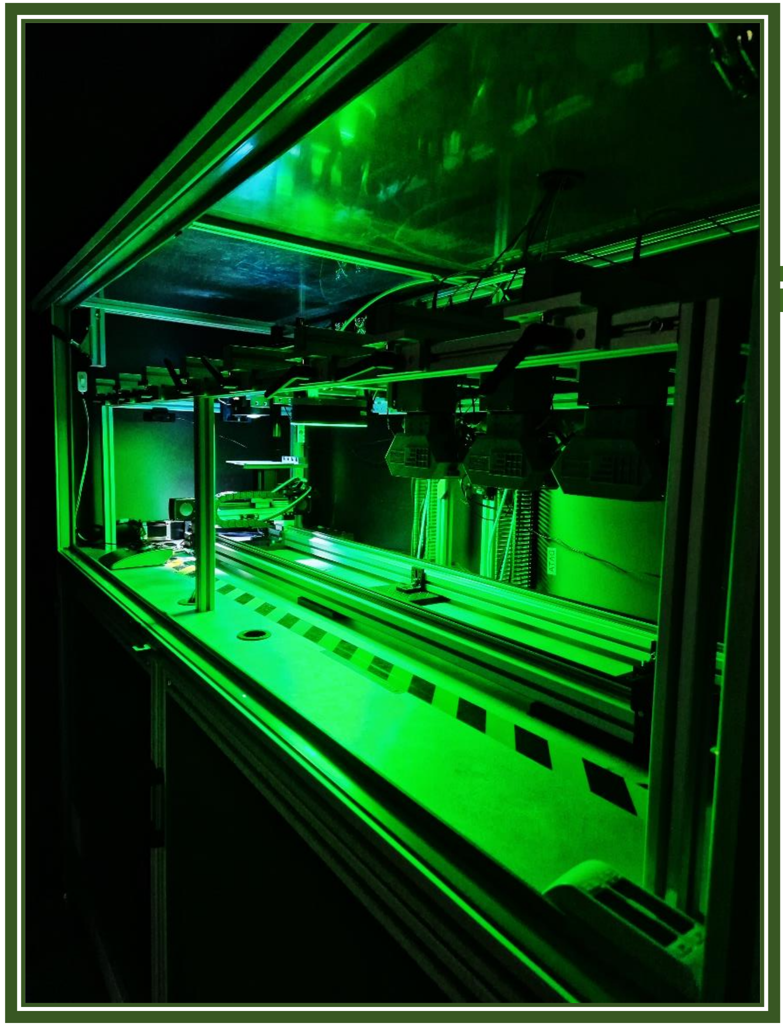


Explainable Artificial Intelligence driven methodology for accelerated research of thin film PV technologies: Case study of kesterite-based devices

Jon Garí-Galíndez[1,2], Robert Fonoll-Rubio[1], Jacob Andrade-Arvizu[1], Alejandro Perez-Rodriguez[1,3], Pedro Vidal-Fuentes[1], Maxim Guc[1], Victor Izquierdo-Roca[1]

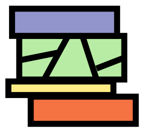
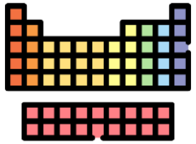
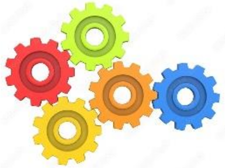
(1) Catalonia Institute for Energy Research – IREC, 08930 Sant Adrià de Besòs, Barcelona, Spain
(2) Departament de Física, Campus Nord B4-B5, Universitat Politècnica de Catalunya, E-08034 Barcelona, Spain
(3) Departament d'Enginyeria Electrònica i Biomèdica, IN2UB, Universitat de Barcelona, C/ Martí i Franqués 1, 08028 Barcelona, Spain

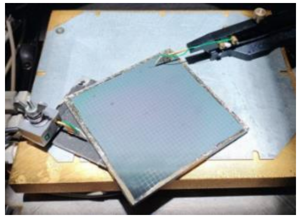
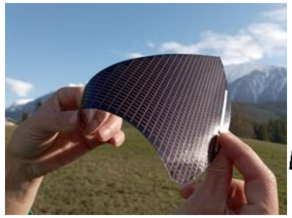
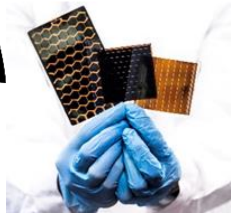
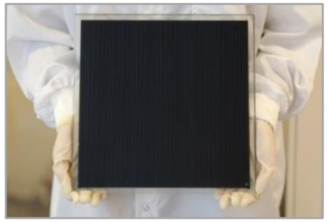
IREC
Shaping Energy for a Sustainable Future

Customizable AI-based in-line process monitoring platform for achieving zero-defect manufacturing in the PV industry (Platform-ZERO)
January 2023 – December 2026

Motivation: systems complexity

Thin-film PV technologies aim to extend their applications in different environments with the objective of efficient electrification of society.

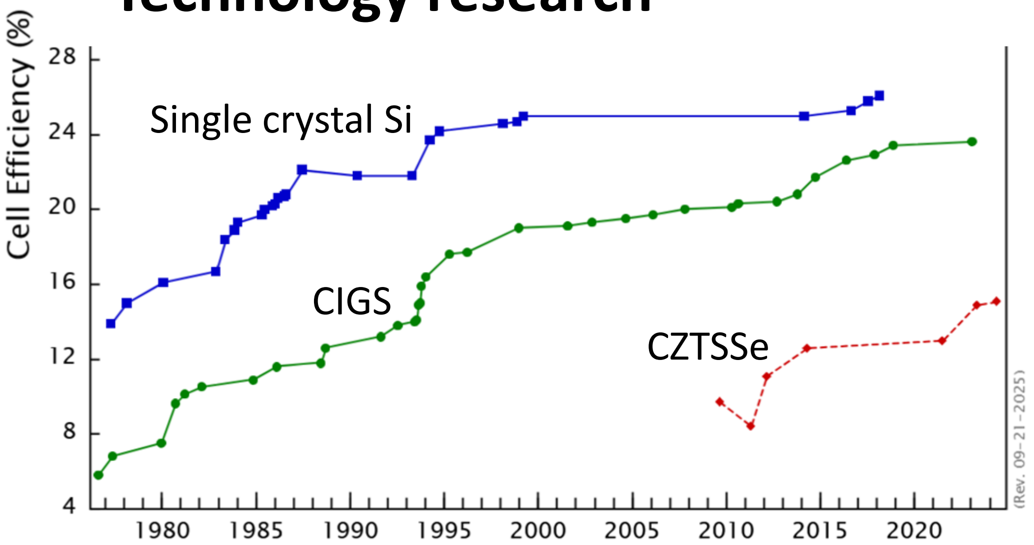
Complex device structures	Complex materials	Complex synthesis process
		

PV Technology	CZTSe-based PV	Customizable/flexible PV			High efficiency CIGSe-based PV
PV device					
Device structure	ITO (500 nm) i-ZnO (50 nm) CdS (50 nm) CZTSe (1.5 µm) MoSe$_2$ (100 nm) Mo (1 µm) Glass (3 mm)	ITO (500 nm) Zn(O,S) (50 nm) CIGSe (1.5 µm) MoSe$_2$ (100 nm) Mo (1 µm) SiO$_x$ (100 µm) Steel (3 mm)	ITO (500 nm) InS (50 nm) CIGSe (1.5 µm) MoSe$_2$ (100 nm) Mo (1 µm) TiW (100 µm) Steel (3 mm)	Carbon (20 µm) P3HT (8 nm) Perovskite (500 nm) SnO$_2$ (10 nm) IZO (270 nm) PET foil (210 µm)	AZO (400 nm) i-ZnO (35 nm) CdS (50 nm) CIGSe (2.0 µm) MoSe$_2$ (40 nm) Mo (600 nm) Glass (3 mm)
Advantage	Critical Raw Materials-Free	Mechanical flexibility and customized solutions			High efficiency
Efficiency	11.1% (AM1.5) [1]	15% (AM1.5) [2]	15% (AM1.5) [3]	30.9% (indoor 1000 lux) [4]	22.6% (AM1.5) [5]

[1] 2025 MRS Spring Oral presentation EN05.06.01, 9 abr 2025, 1:30 p.m., Summit, Level 3, Room 330
[2] DOI: 10.1002/smtd.202301573
[3] DOI: 10.1016/j.tsf.2017.06.031
[4] DOI: 10.1002/adfm.202206761
[5] DOI: 10.1002/pssr.201600199

MOTIVATION: System scale-up

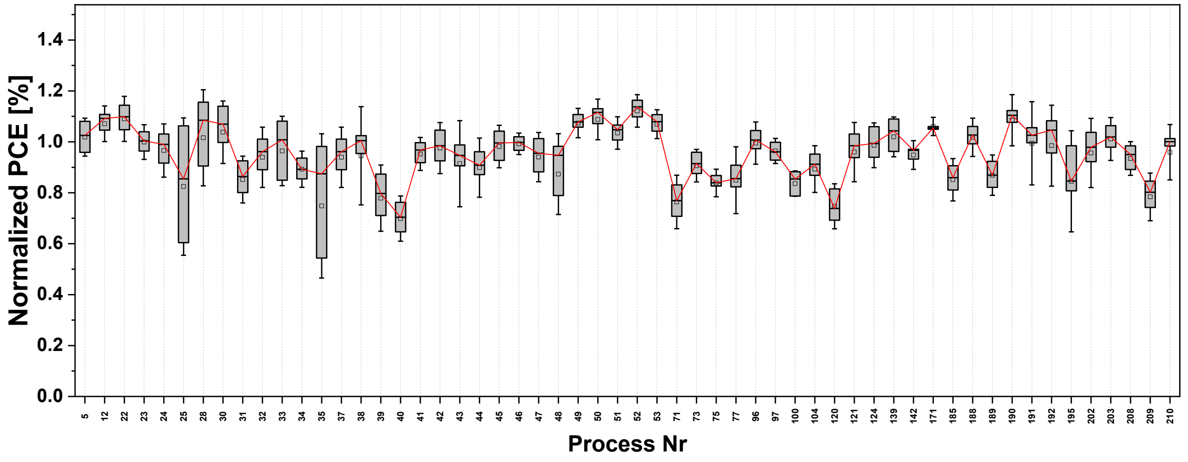
Technology research



Large research time >10 years

➔ limitation in understanding cause-&-effect relationship

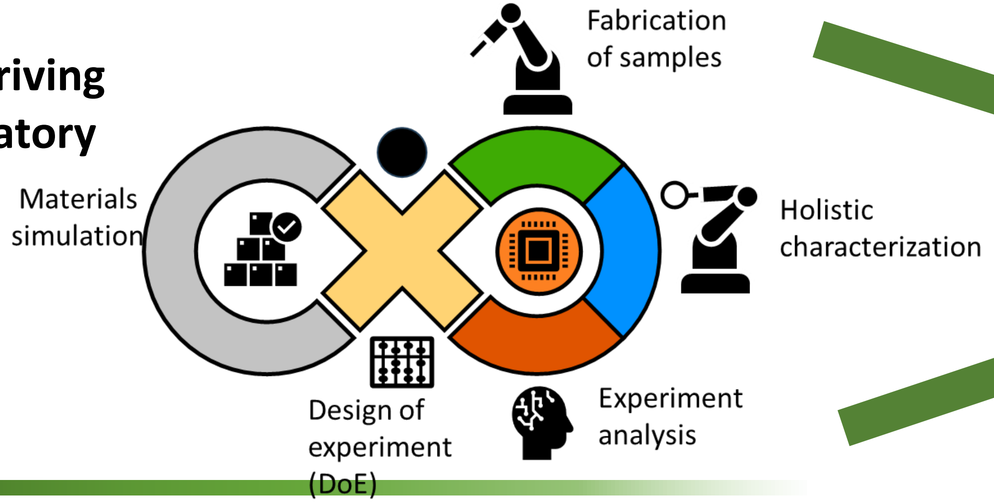
Industrial scale-up



Lab to industry scaling up gap

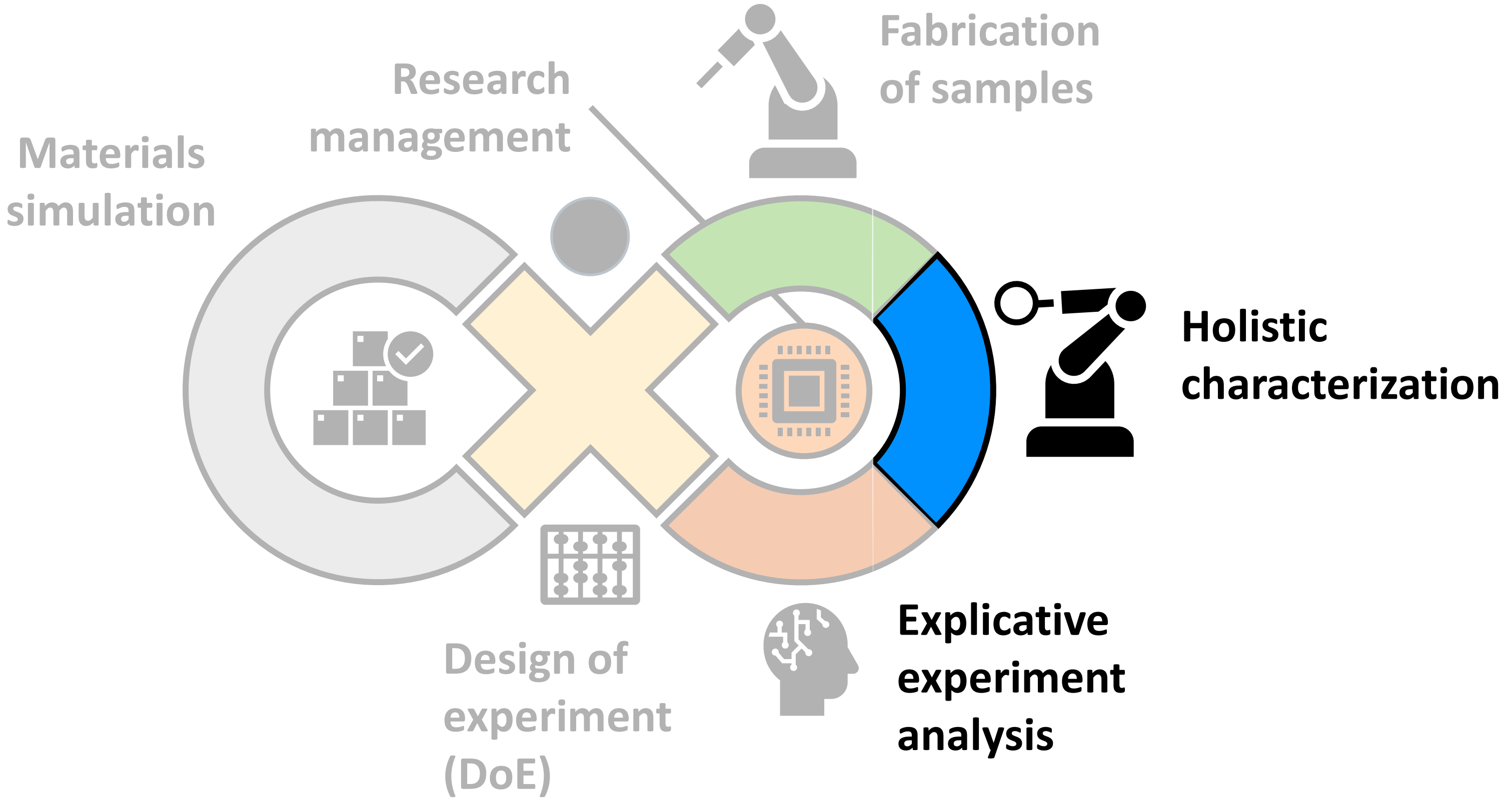
➔ reproducibility problems

Self-driving laboratory



Optimized quality control and process monitoring

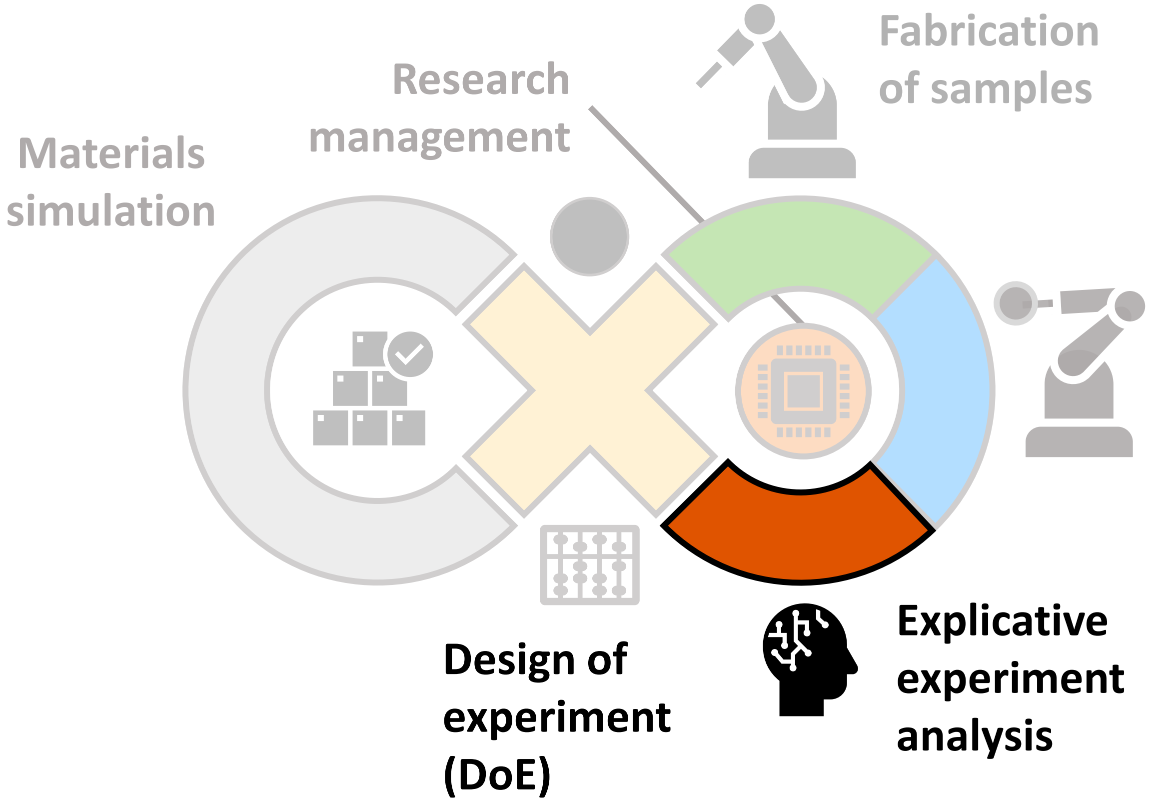
MOTIVATION: Self-driving laboratory



- Robotization of the process
- Combination of different techniques
- Heterogeneous data fusion

TALK 2AO.2.2

MOTIVATION: Self-driving laboratory



- Robotization of the process
- Combination of different techniques
- Heterogeneous data fusion

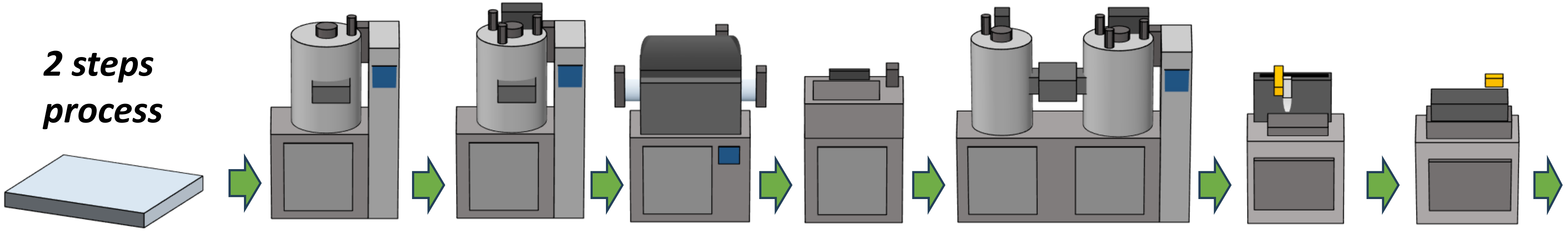
- Data classification
- Multidimensional analysis
- Acceleration of Interpretability

TALK
2AO.2.2

TALK
2AO.2.4

MOTIVATION: Case study of kesterite-based devices

2 steps process



Glass Substrate

Sputtering: Metallic Back contact

Sputtering: Absorber precursor

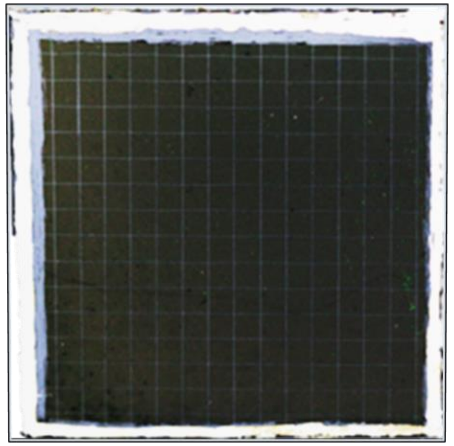
Annealing: Light Absorber synthesis

CBD: Carriers separation stack

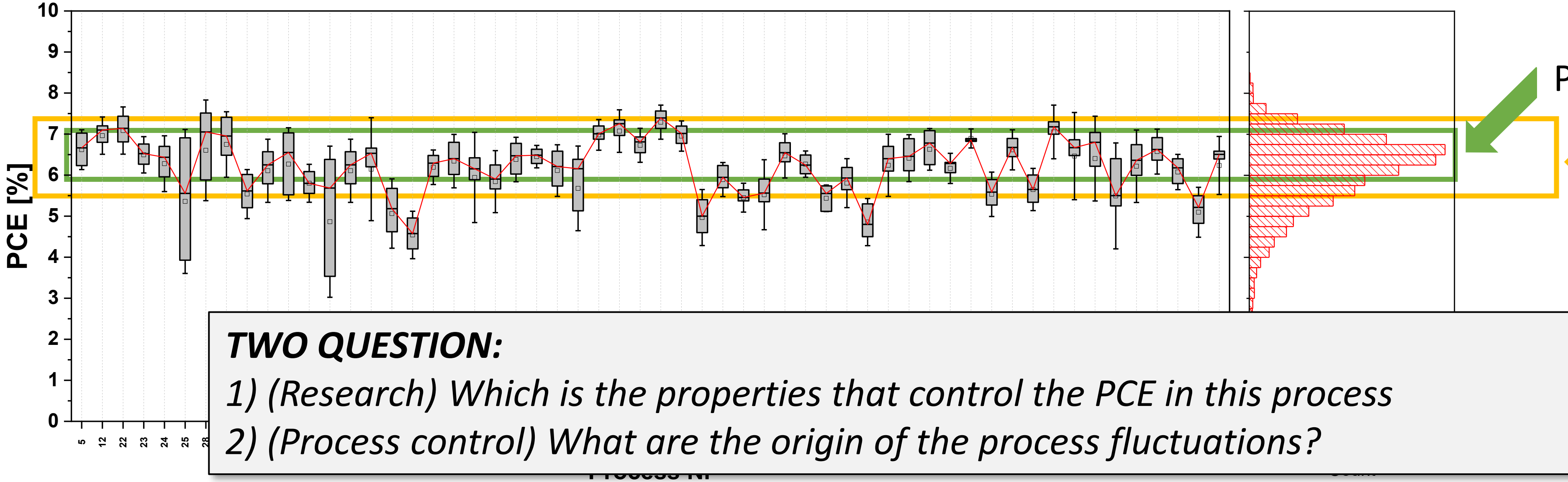
Sputtering: Front Transparent contact stack

Annealing: System stabilization

Scribing: Devices discretization



5x5 cm2 sample
#196 3x3mm^2 Test-cell



PCE: 6.00-7.00 ➔48%

PCE: 5.25-7.25% ➔77%

TWO QUESTION:

1) (Research) Which is the properties that control the PCE in this process

2) (Process control) What are the origin of the process fluctuations?

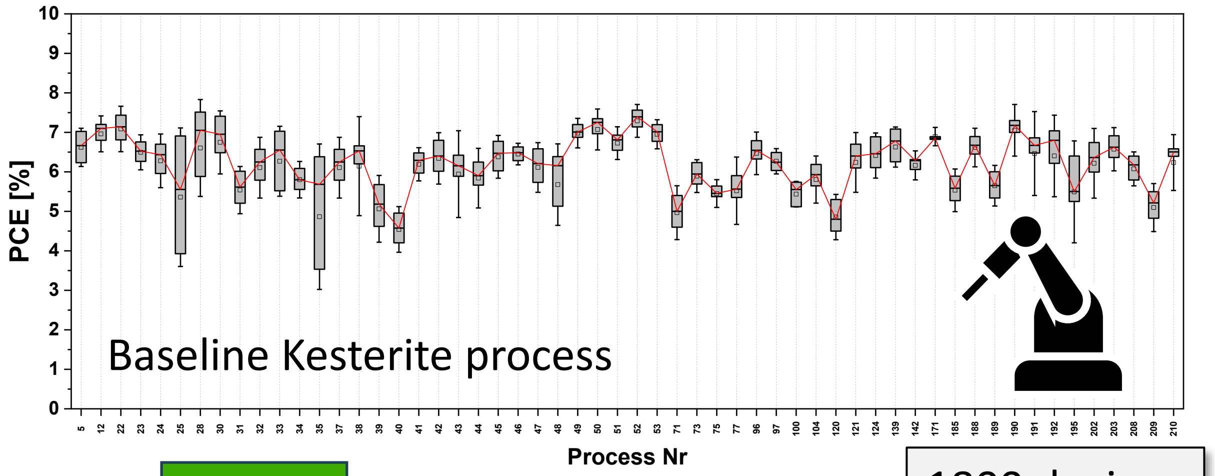
METHODOLOGY

➔ **Holistic characterization**

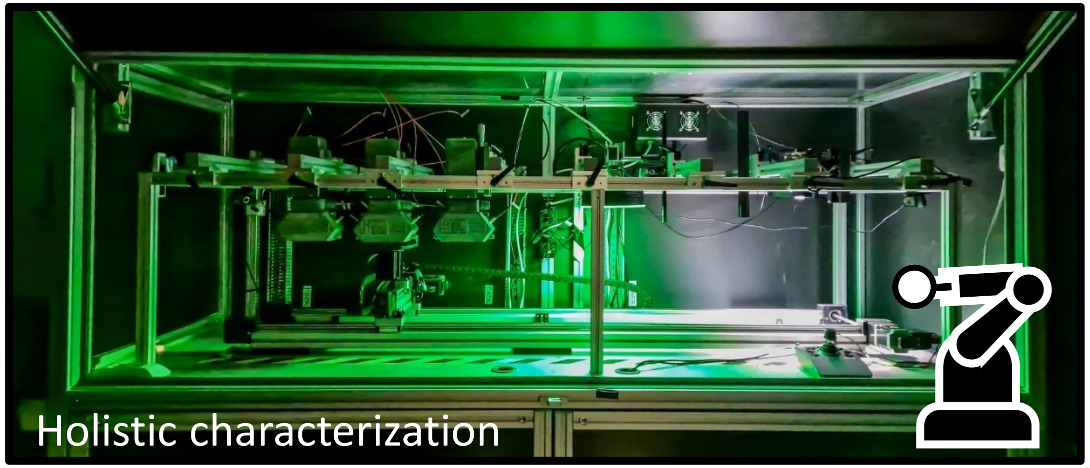
TALK 2AO.2.2

>21.000.000 Features (Dimensions)

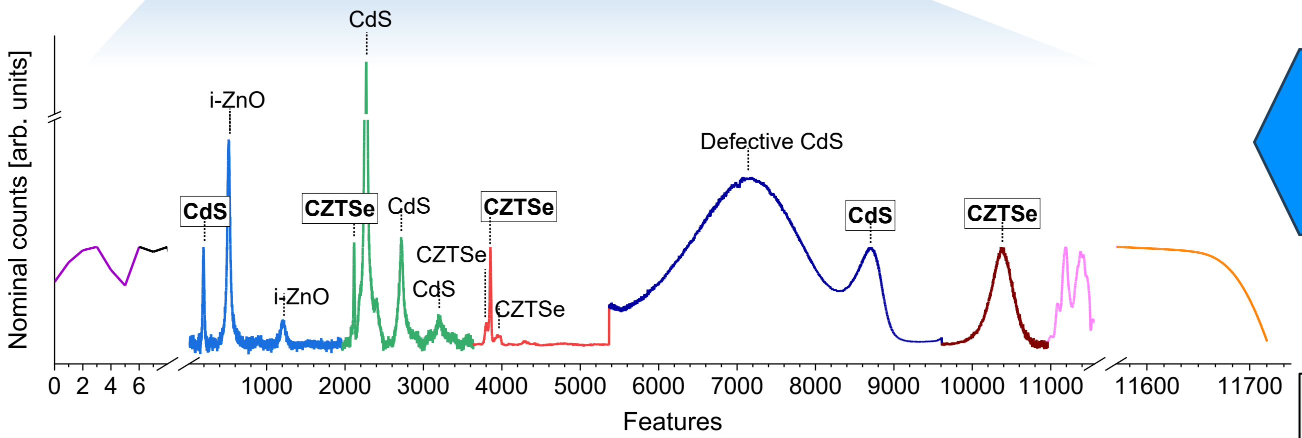
too much data!



1800 devices

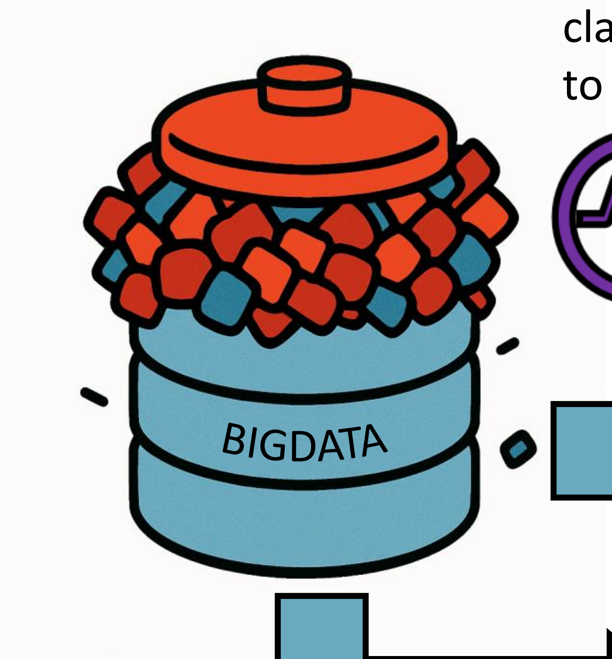


9 techniques with a total of 12000 features per device



METHODOLOGY

➔ **Holistic characterization**



➔**Classification:** Method to classify data information according to the interest of the study

Black-box system model

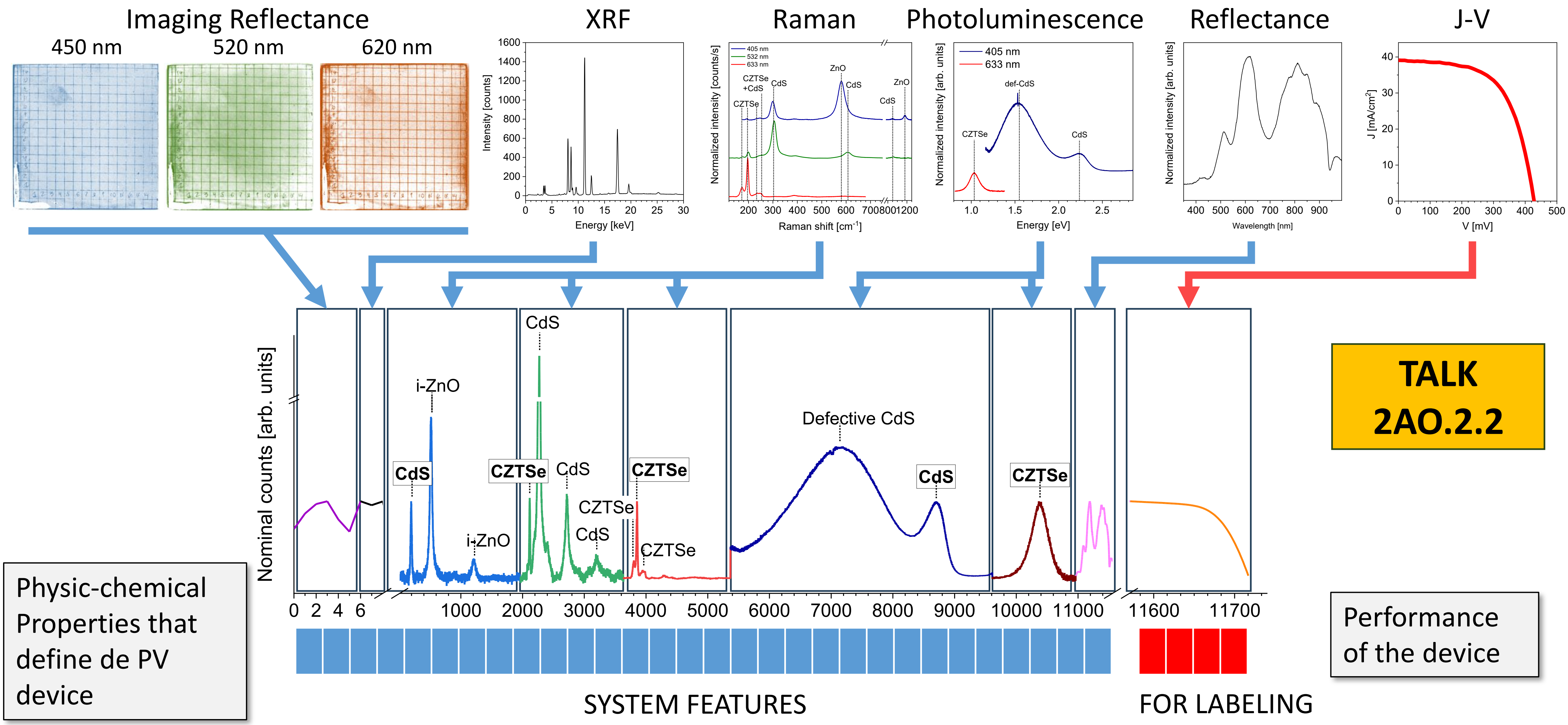
➔ **AI assisted modeling**: Correlation between the characterization information and the proposed classification

Sensitivity analysis

Explainable analysis

➔**Explainability**: Identification of the relevance of the initial information (features) in the model

HOLISTIC CHARACTERIZATION

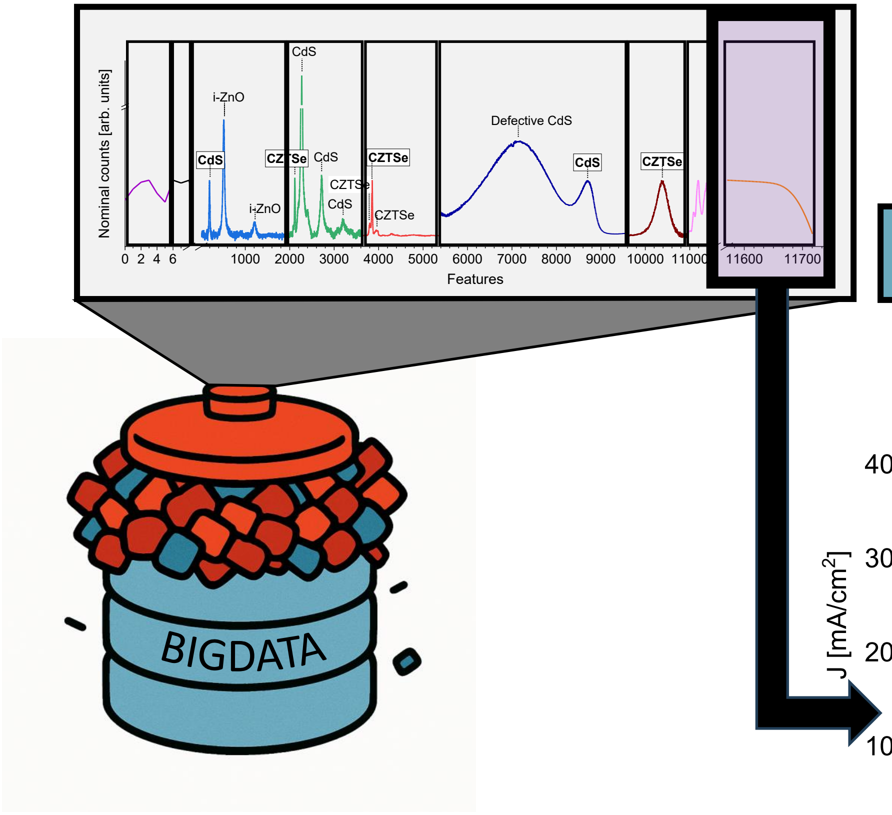


TALK 2AO.2.2

Physic-chemical Properties that define de PV device

Performance of the device

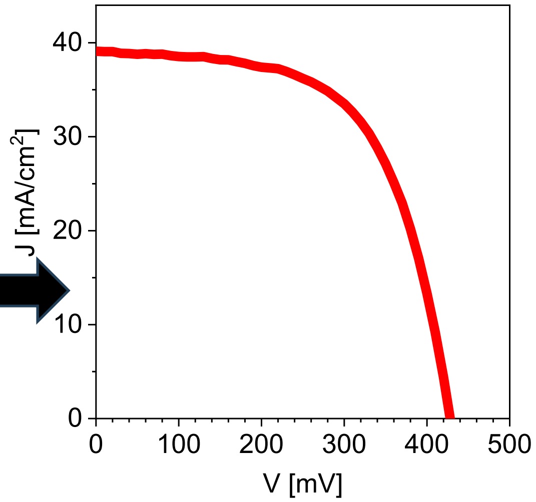
DATA CLASSIFICATION

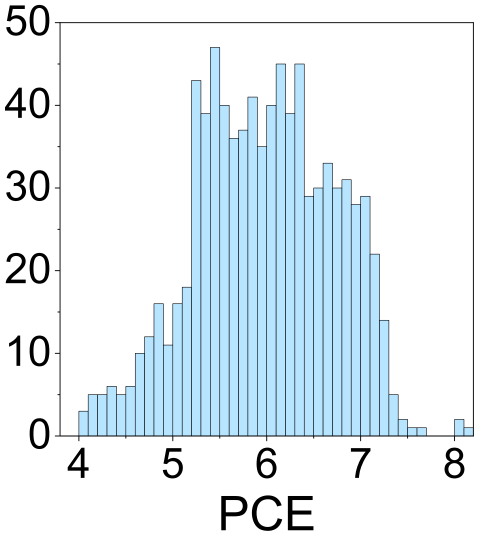


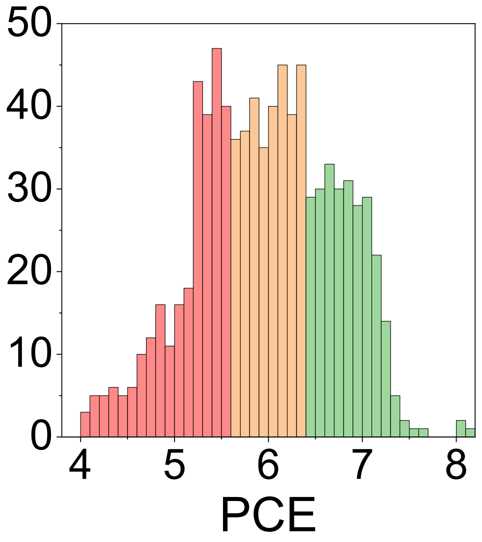
Selection of property to validate the research **(Data for labeling)**

Distillation of the quantifiable parameter (**labeling**)

Discretization in relevant groups (**Classification**)



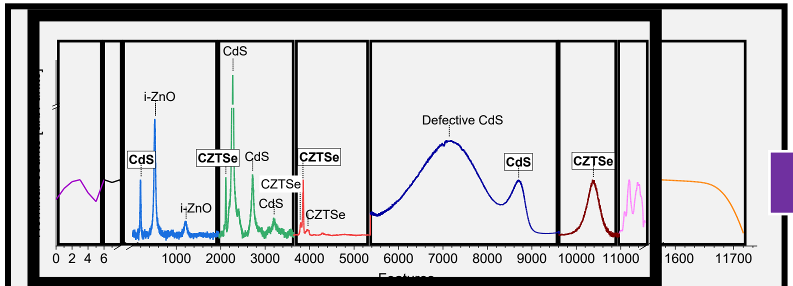


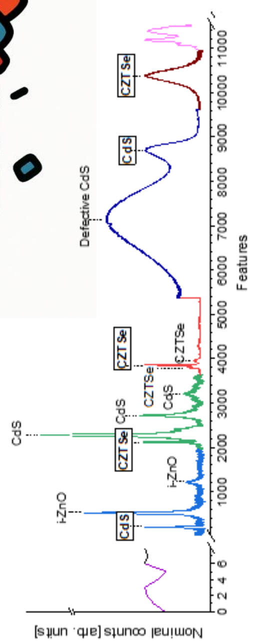


We select the PCE as relevant parameter for our research or control optimization

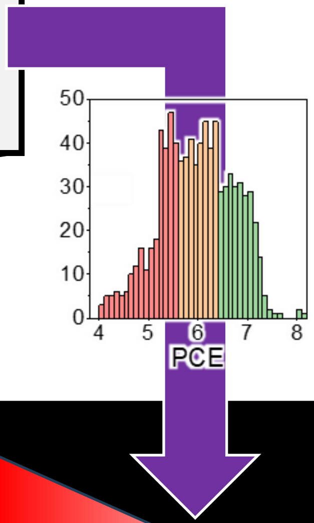
Class1: PCE >6.4%
Class2: 6.4%>PCE>5.6
Class3: 5.6>PCE

AI ASSISTED MODELING





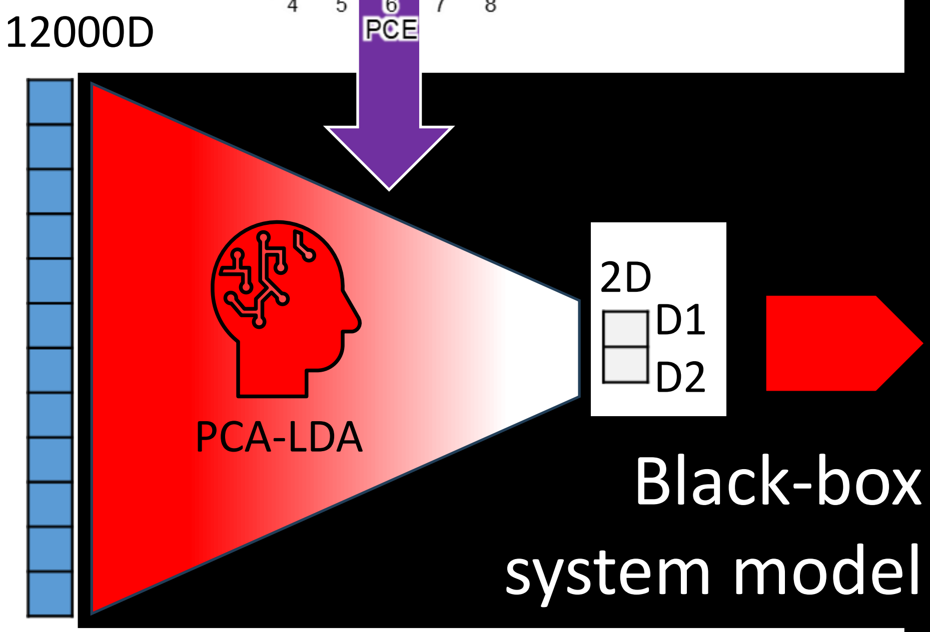
Discretization in relevant groups (**Classification**)



Class1: PCE >6.4%

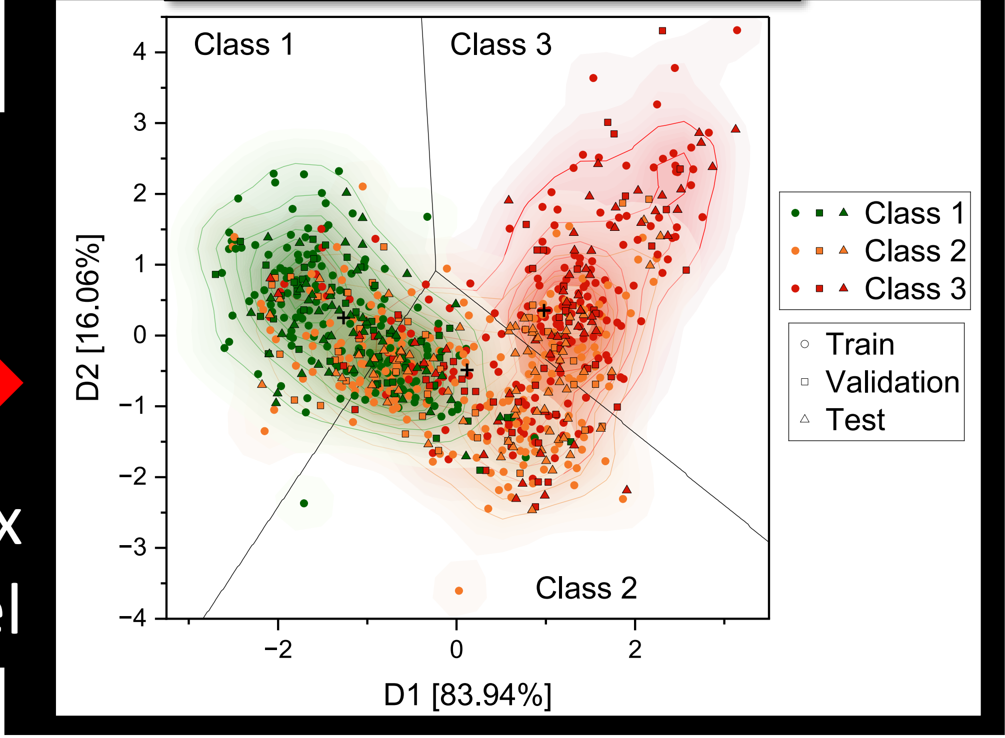
Class2: 6.4%>PCE>5.6

Class3: 5.6>PCE



The AI model reduces data dimensionality using a combination of input features that maximizes discrimination based on the classification criterion.

Latent space
Cross-validated Score: 56.55%

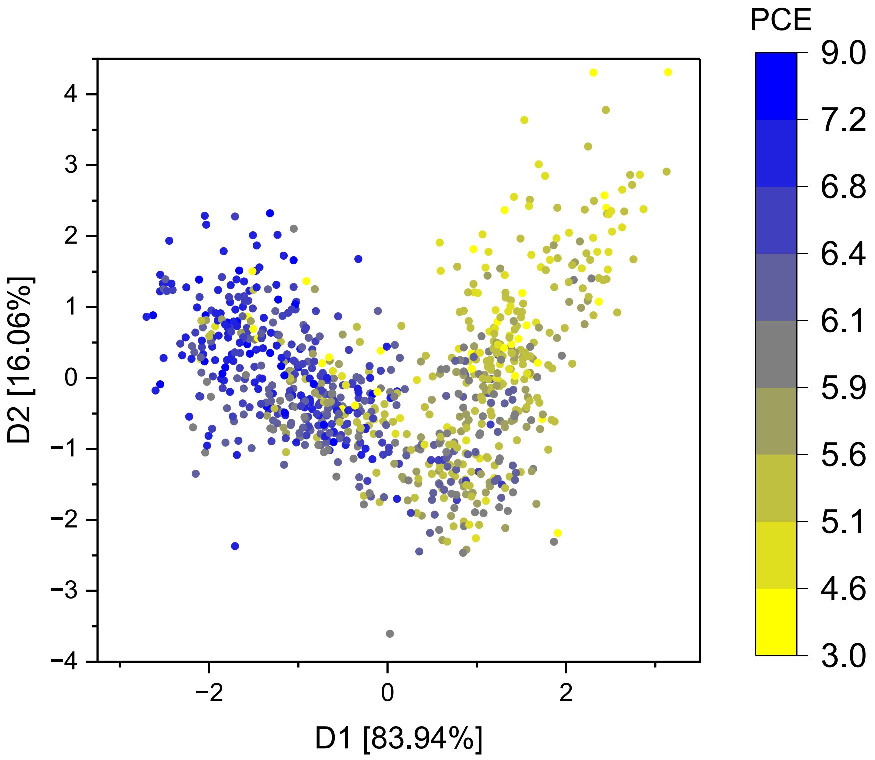


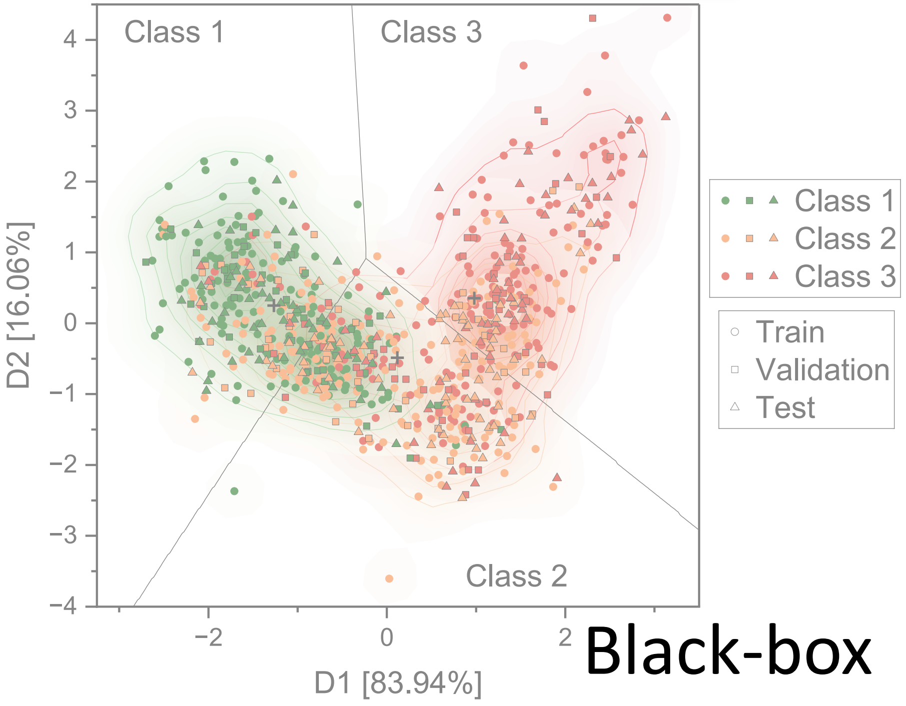
EXPLAINABILITY

Including the efficiency of each point as a color scale reveals a gradual ordering

The AI model reduces data dimensionality using a combination of input features that maximizes discrimination based on the classification criterion.

Latent space
Cross-validated Score: 56.55%



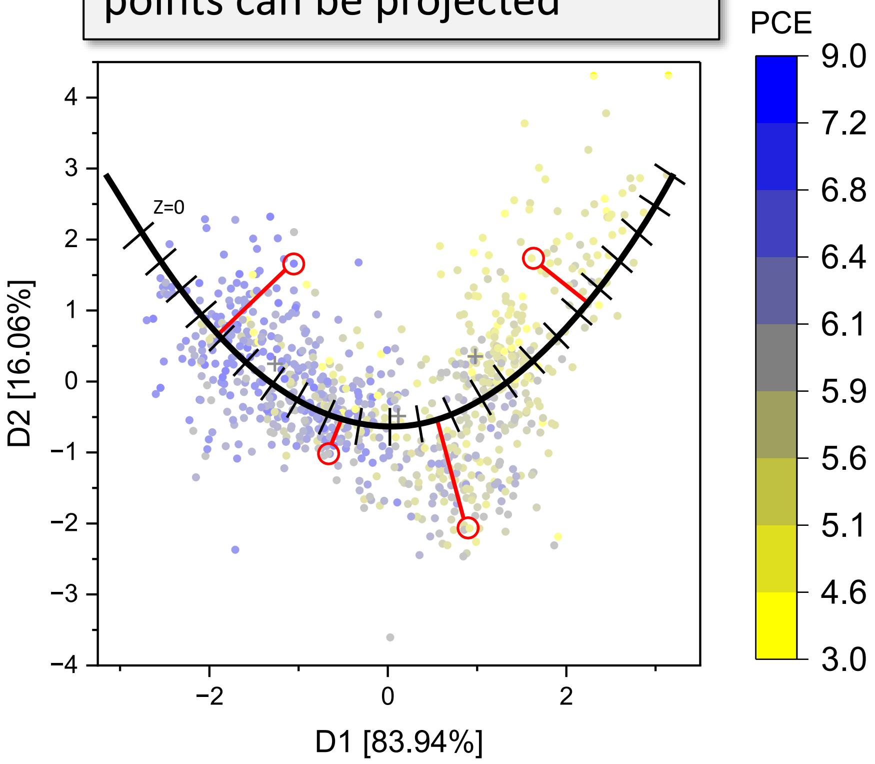


Black-box system model

EXPLAINABILITY

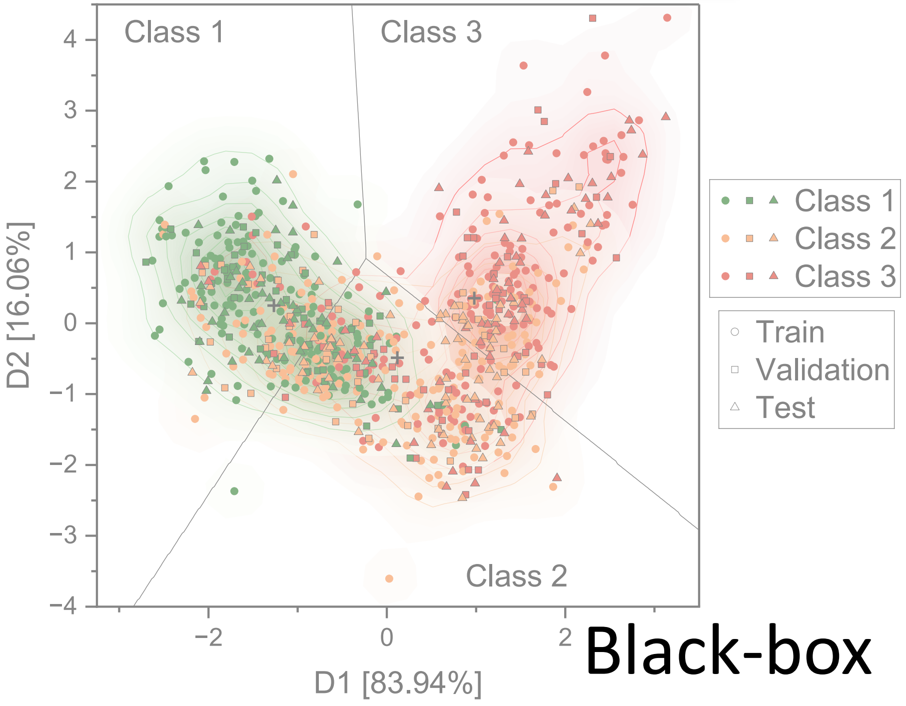
Including the efficiency of each point as a color scale reveals a gradual ordering

It can be parameterized through a Z-curve, dependent on (D1 and D2), onto which all points can be projected



The AI model reduces data dimensionality using a combination of input features that maximizes discrimination based on the classification criterion.

Latent space
Cross-validated Score: 56.55%



Black-box system model

EXPLAINABILITY

Which exhibits a quasi-linear relationship with the PCE
PCE depends on D1 and D2

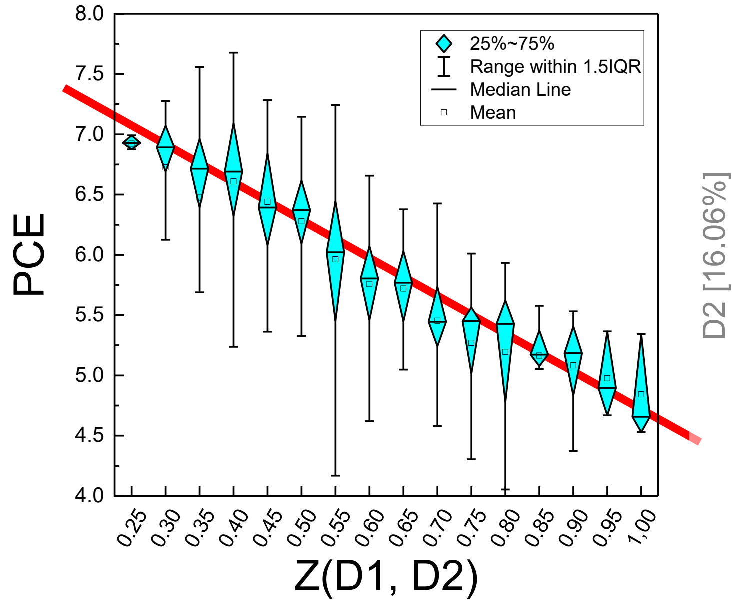
What is the physical meaning of these parameters D1 and D2?

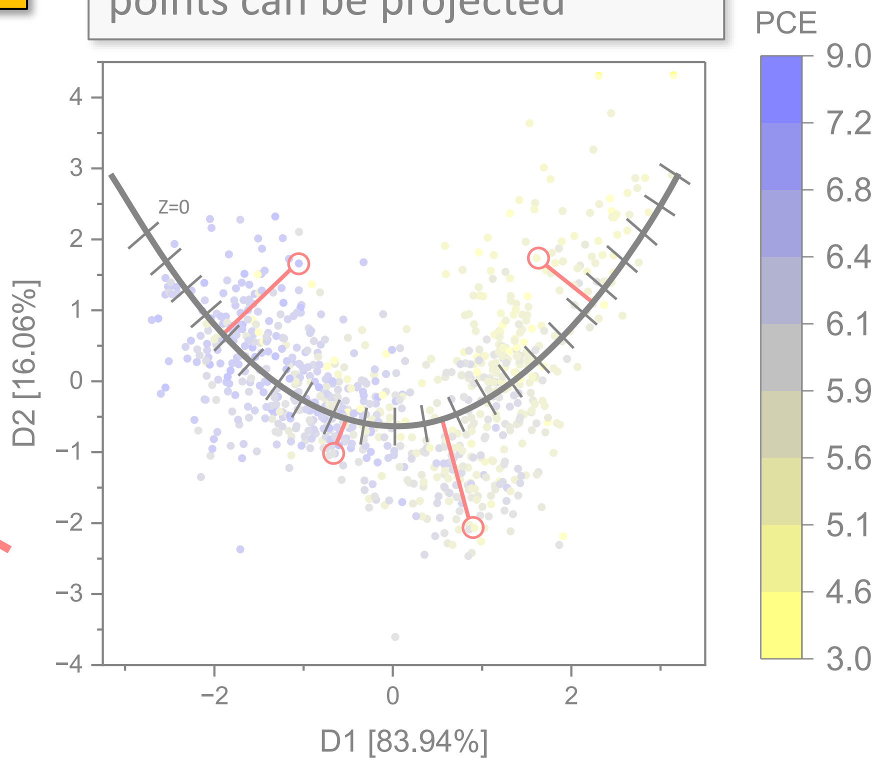
Including the efficiency of each point as a color scale reveals a gradual ordering

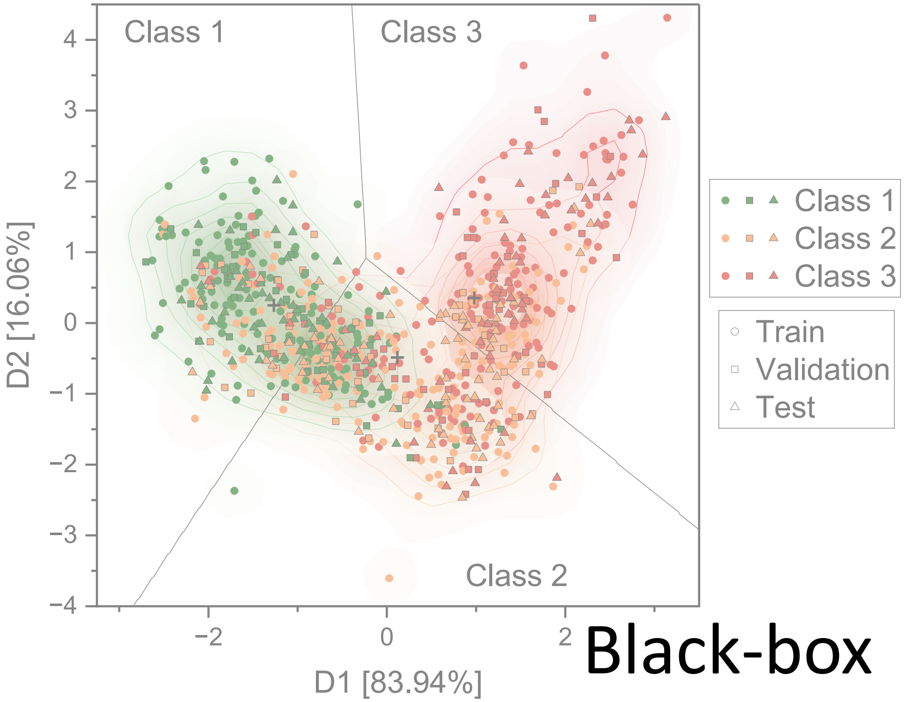
It can be parameterized through a Z-curve, dependent on (D1 and D2), onto which all points can be projected

The AI model reduces data dimensionality using a combination of input features that maximizes discrimination based on the classification criterion.

Latent space
Cross-validated Score: 56.55%



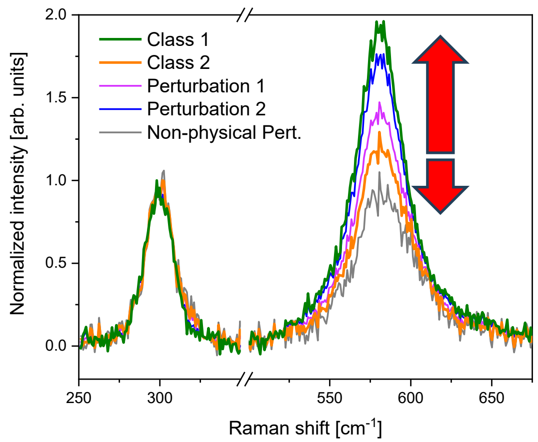




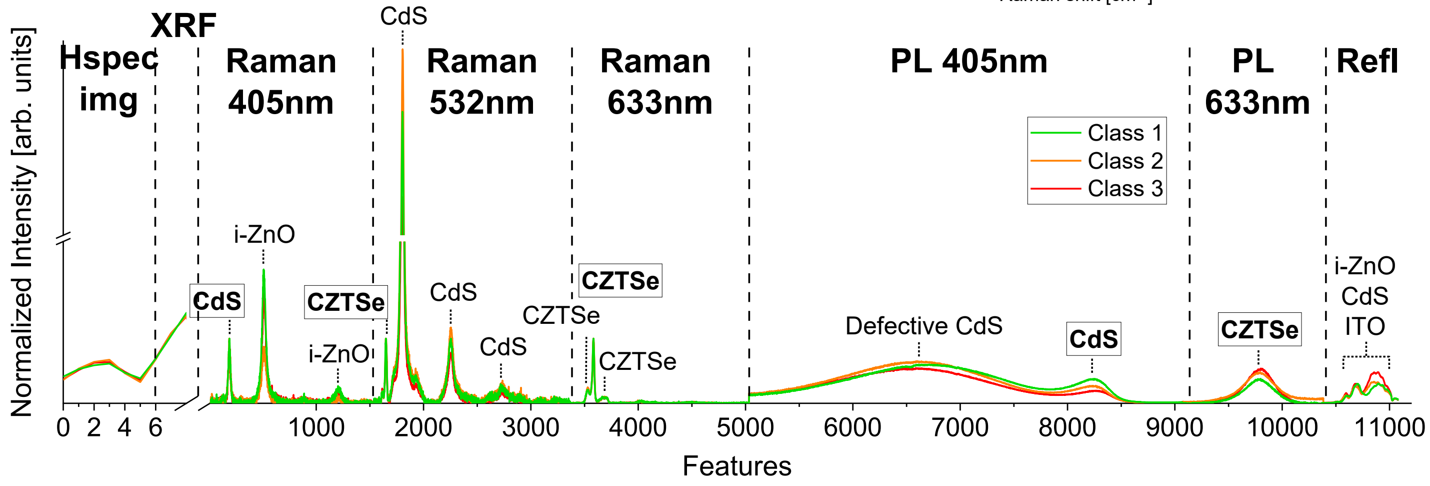
Black-box system model

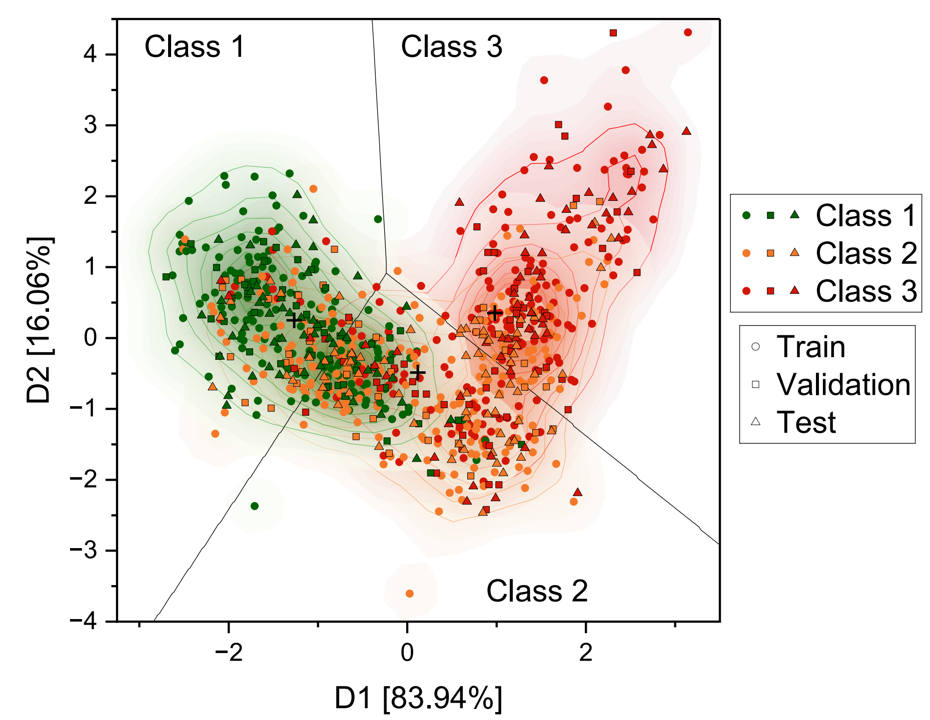
EXPLAINABILITY: What is the physical significance of D1 & D2 parameters?

We artificially perturb the data by modifying the intensity of different regions.



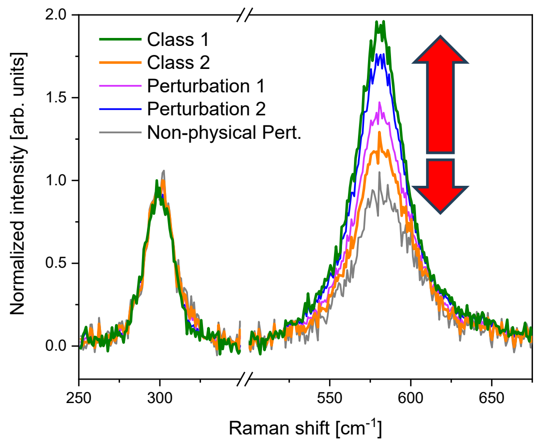
Black-box system model

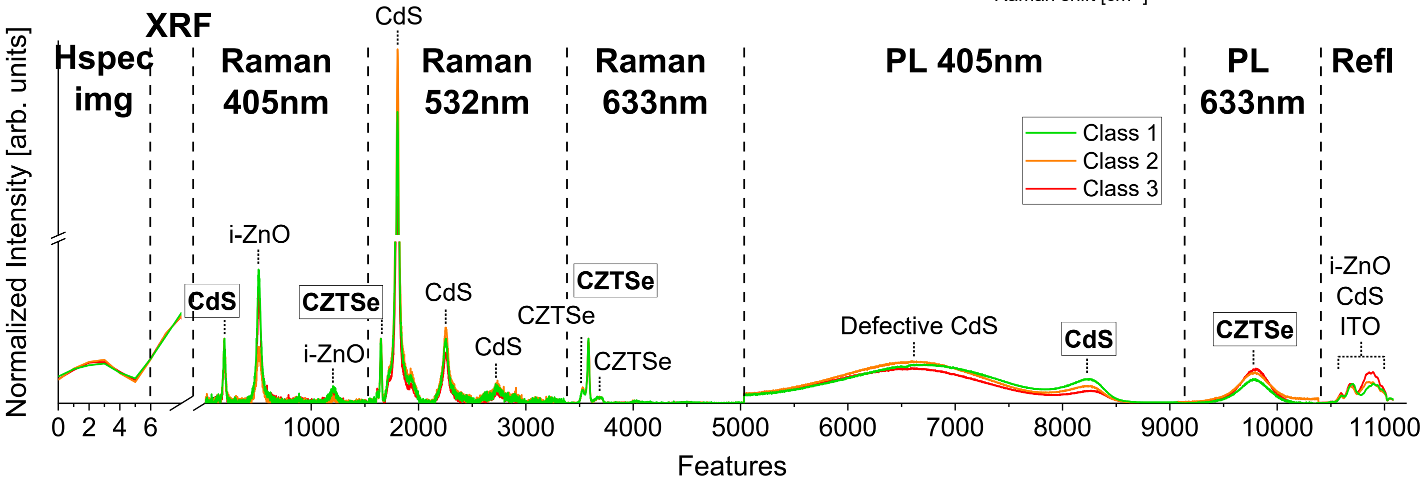


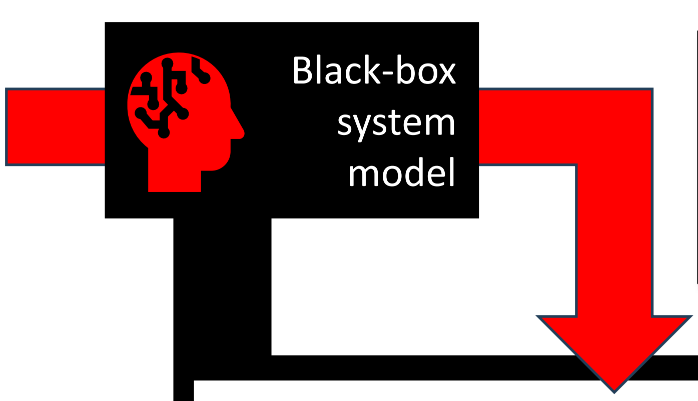


EXPLAINABILITY: What is the physical significance of D1 & D2 parameters?

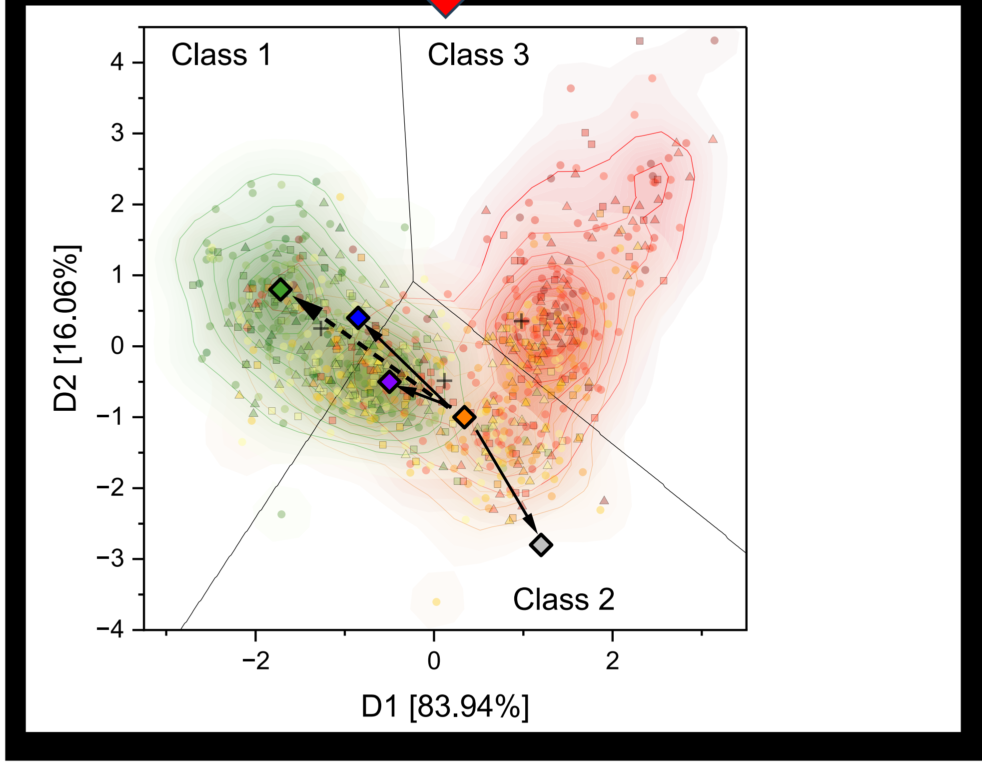
We artificially perturb the data by modifying the intensity of different regions.







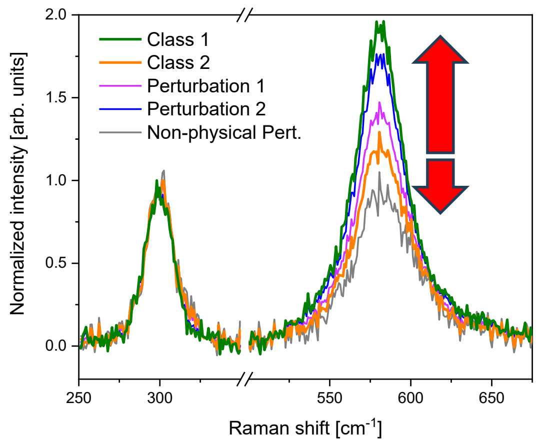
This generates displacements within the lattice space map



EXPLAINABILITY: What is the physical significance of D1 & D2 parameters?

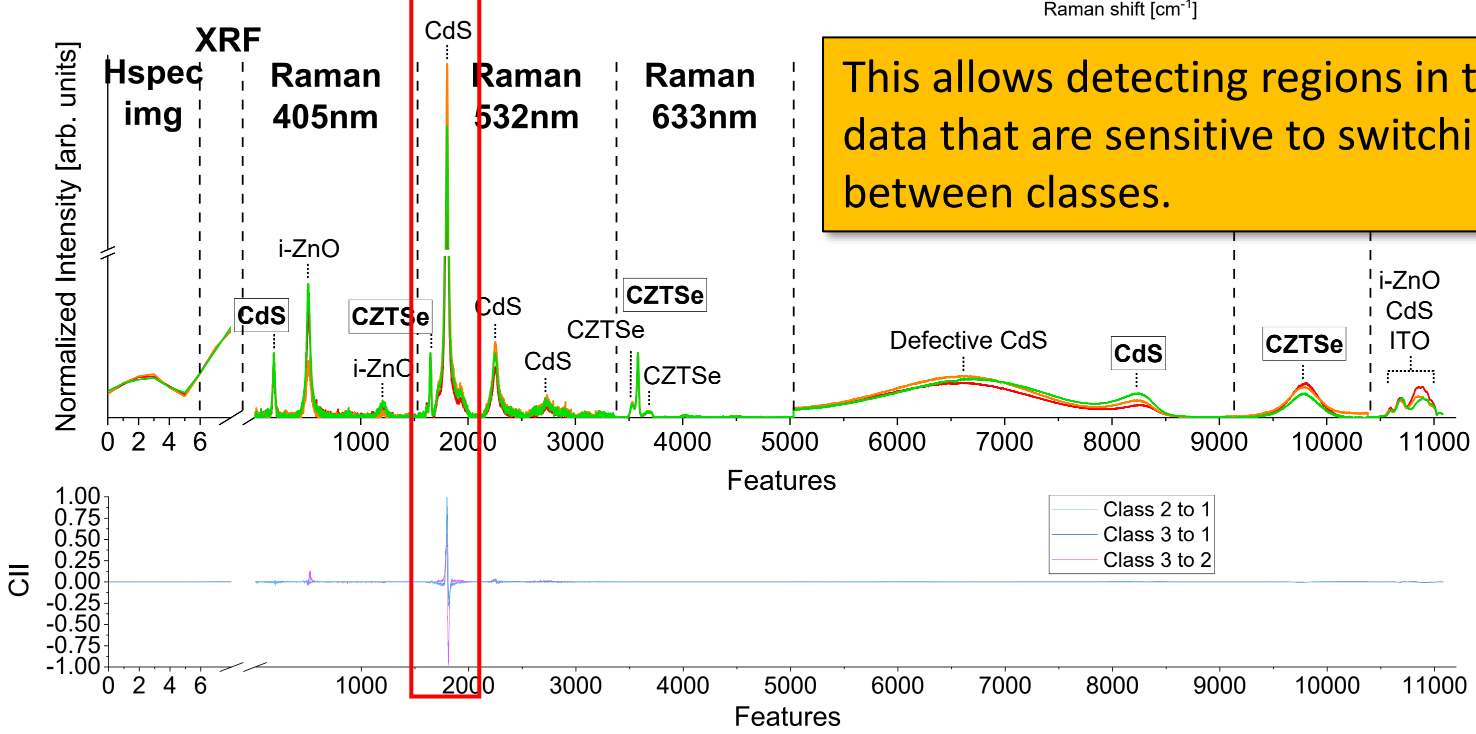
We artificially perturb the data by modifying the intensity of different regions.

D1 and D2 are controlled by CdS

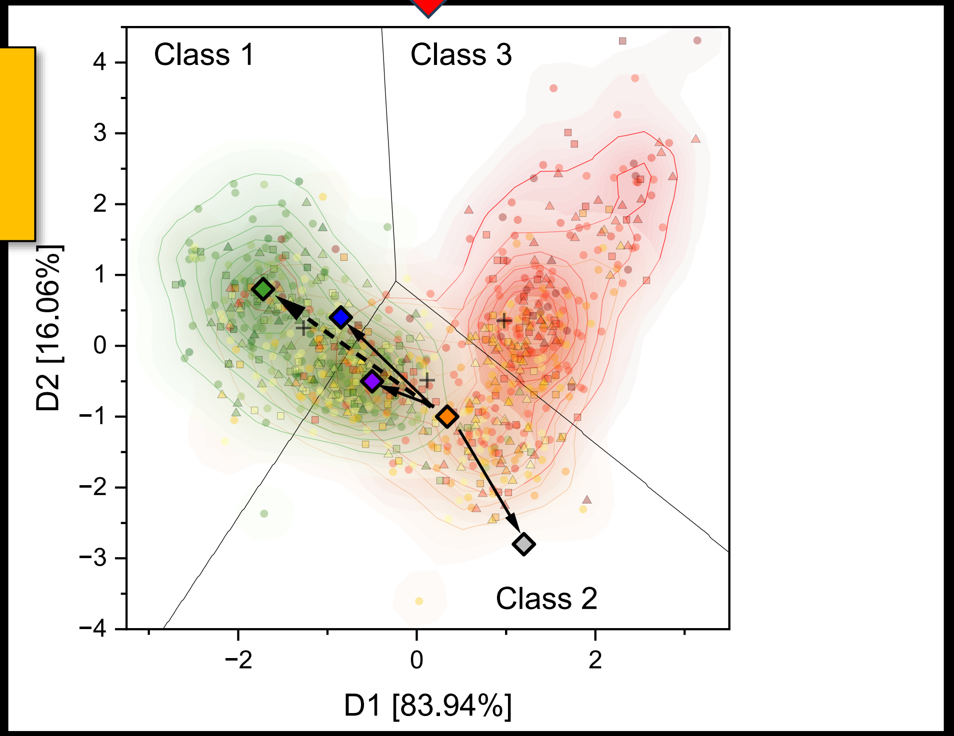


Black-box system model

This generates displacements within the lattice space map



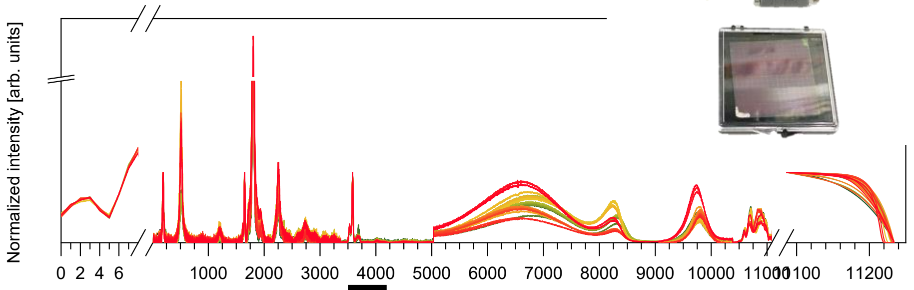
This allows detecting regions in the data that are sensitive to switching between classes.



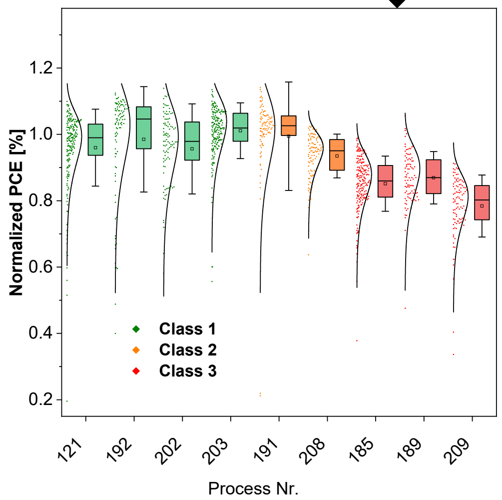
Application in process control: What are the origin of the process fluctuations?

Industrial process monitoring

Data set



Classification



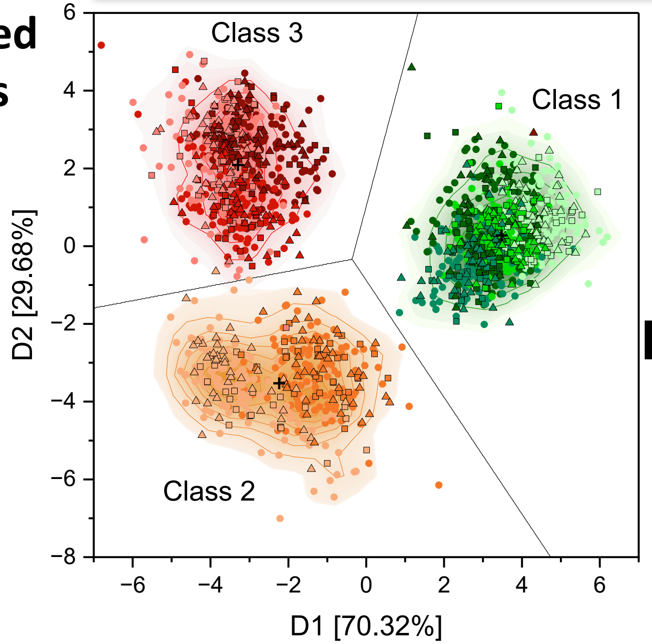
Class1: Voc ☑ Jsc ☑

Class2: Voc ☑ Jsc ☒

Class3: Voc ☒

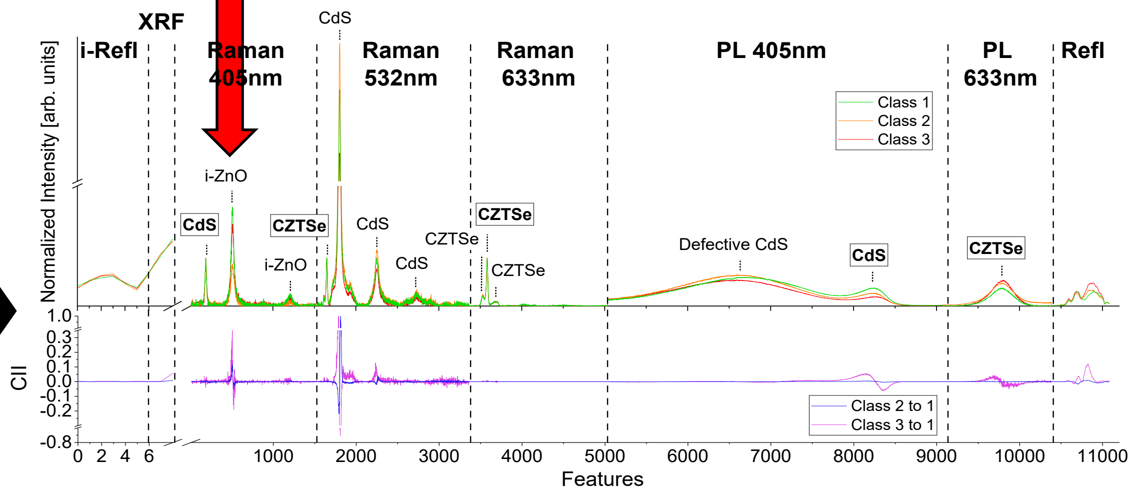
AI-assisted analysis

Cross-validated Score: 99.71 %



Origin of the process fluctuation related to:

1) CdS
2) i-ZnO



Explainability by perturbation method

CONCLUSIONS

In this work, we have demonstrated the potential of artificial intelligence in the field of scientific and technical research:

1) AI-assisted big data analysis opens the door to truly multidimensional data analysis in an accelerated way.

2) A correct selection of label parameters and classification classes allows addressing different research and industry questions.

3) The use of AI self-explanation strategies enhances the interpretability of AI models by enabling the extraction of a physical model of the system

4) Combined with the automation of holistic characterization, this drives research toward acceleration and the development of self-driving laboratories.

Future actions

5) Automate design of experiment (DoE).
6) Automate the synthesis of materials and devices.
7) Collaborations and test in other technologies

Methodology	Steps	Approximated duration
Traditional	Manual characterization of the >1750 cells	> 1 month
	Manual data pre-processing (baseline subtraction…)	2-3 weeks
	Data Analysis, identification of relevant correlations	~ 1 year[a]
	Total	$\boxed{\textbf{> 1 year}}$
Accelerated	Multimodal automate characterization of the >1750 cells	2 days
	Automated data pre-processing (baseline subtraction…) and dataset generation	1 hour
	Modeling based in PC-LDA trainings	Minutes
	Sensitivity analysis	1 day
	Multivariate polynomial regression	1 day
	Total	$\boxed{\textbf{4-5 days}}$

More details

Jon Garí-Galíndez et.al.,
Explainable Artificial Intelligence Driven Methodology for Accelerated Research of Complex Systems: Case Study of Thin-Film Photovoltaic Kesterite-Based Technology
Adv. Energy Mater. 15 (2025) 2502420

Jon Garí-Galíndez, et al
Revealing the Impact of CZTSe/CdS Interface Fluctuations on PV Device Performance through Big Data Analysis Assisted by Machine Learning Methods
Small Methods (2025) 2400661/9.

Enric Grau-Luque, et al
Accelerating the Development of Thin Film Photovoltaic Technologies: An Artificial Intelligence Assisted Methodology Using Spectroscopic and Optoelectronic Techniques,
Small Methods (2024) 2301573/17.

Thank you for your attention!!

Customizable AI-based in-line process monitoring platform for achieving zero-defect manufacturing in the PV industry (Platform-ZERO)

January 2023 – December 2026

http://www.platform-zero-project.eu/

(Soon this presentation will be accessible on this webpage.")

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Health and Digital Executive Agency (HADEA). Neither the European Union nor the granting authority can be held responsible for them.

Contacts for collaboration, PhD or postdoctoral opportunities:

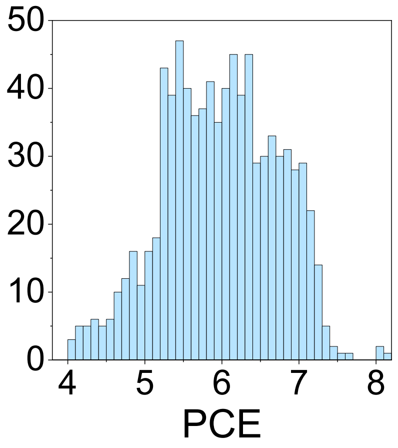
Victor Izquierdo vizquierdo@irec.cat

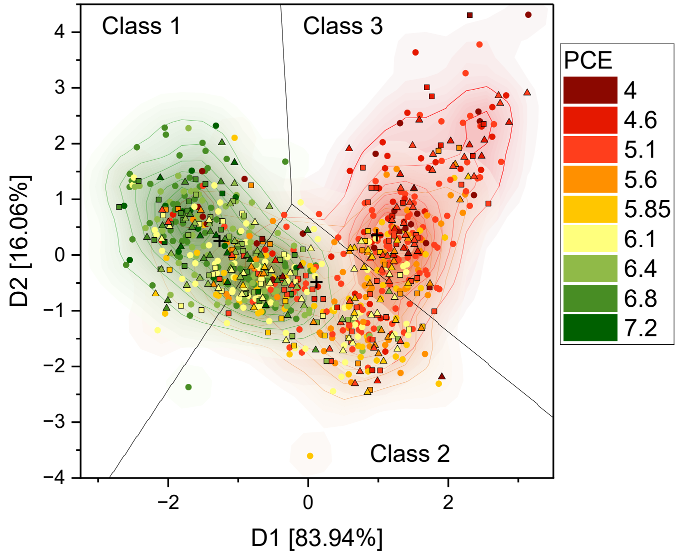
Pedro Vidal pvidal@irec.cat

Maxim Guc mguc@irec.cat

Research vs process monitoring

Research case





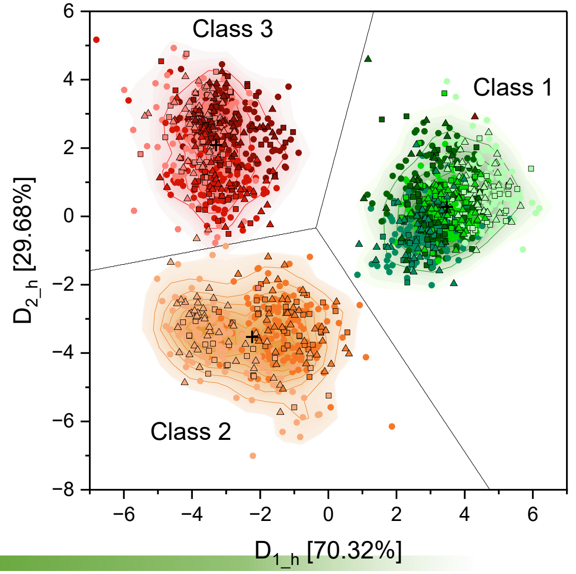
Latent space
Cross-validated Score: 56.55%

PCE controlled by
1) CdS thickness

Continues question➔ continues "classification" ➔ lower score

Process monitoring case

Class1: Voc ☑ Jsc ☑
Class2: Voc ☑ Jsc ☒
Class3: Voc ☒

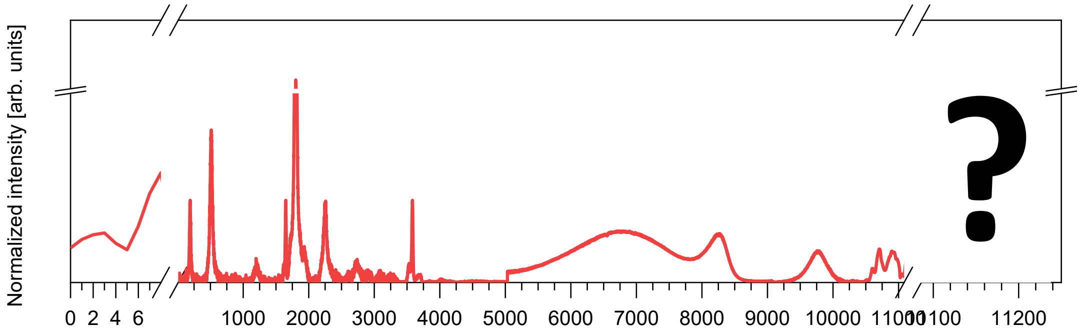


Latent space
Cross-validated Score: 99.71%

Fluctuation controlled by:
1) CdS variations
2) i-ZnO variations

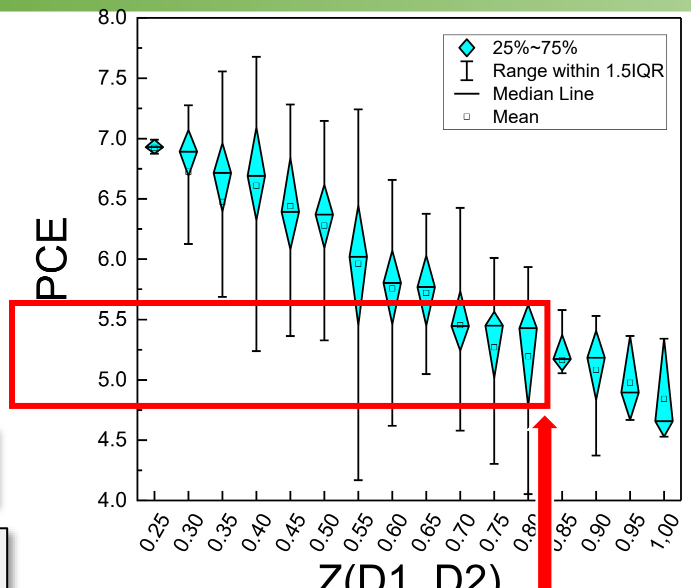
Discrete question ➔ discrete "classification" ➔ higher score

AI MODEL APPLICATION

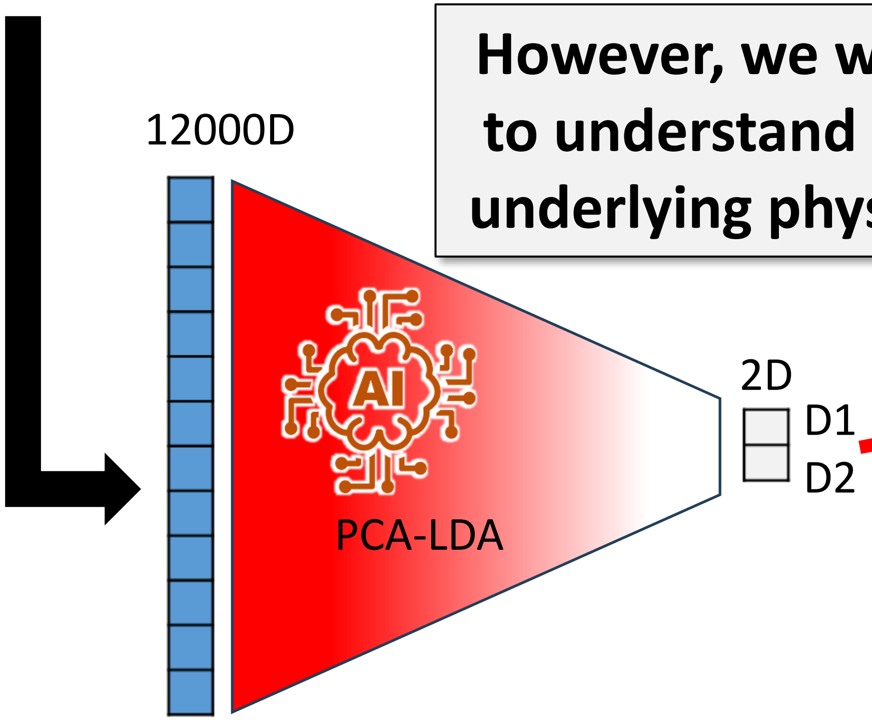


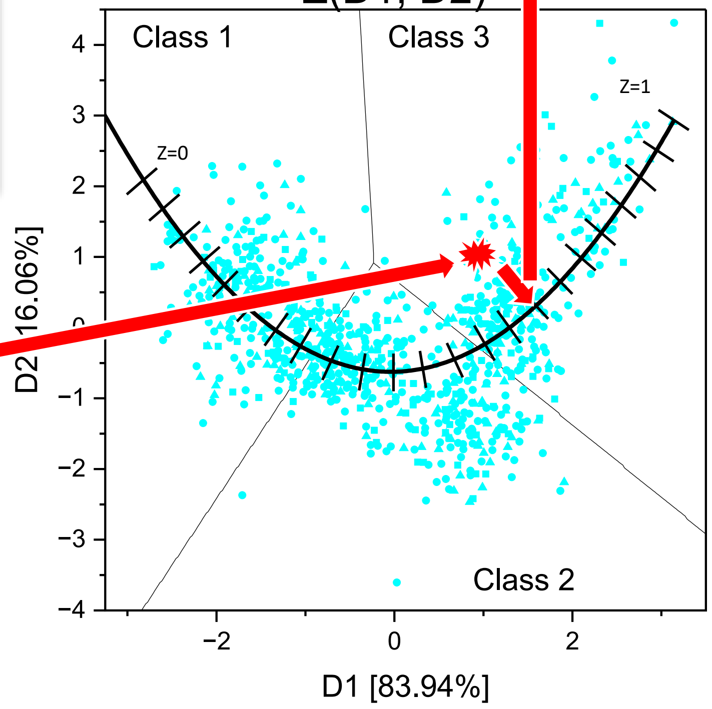
5.2 ± 0.5 %

PREDICTION!!!!!

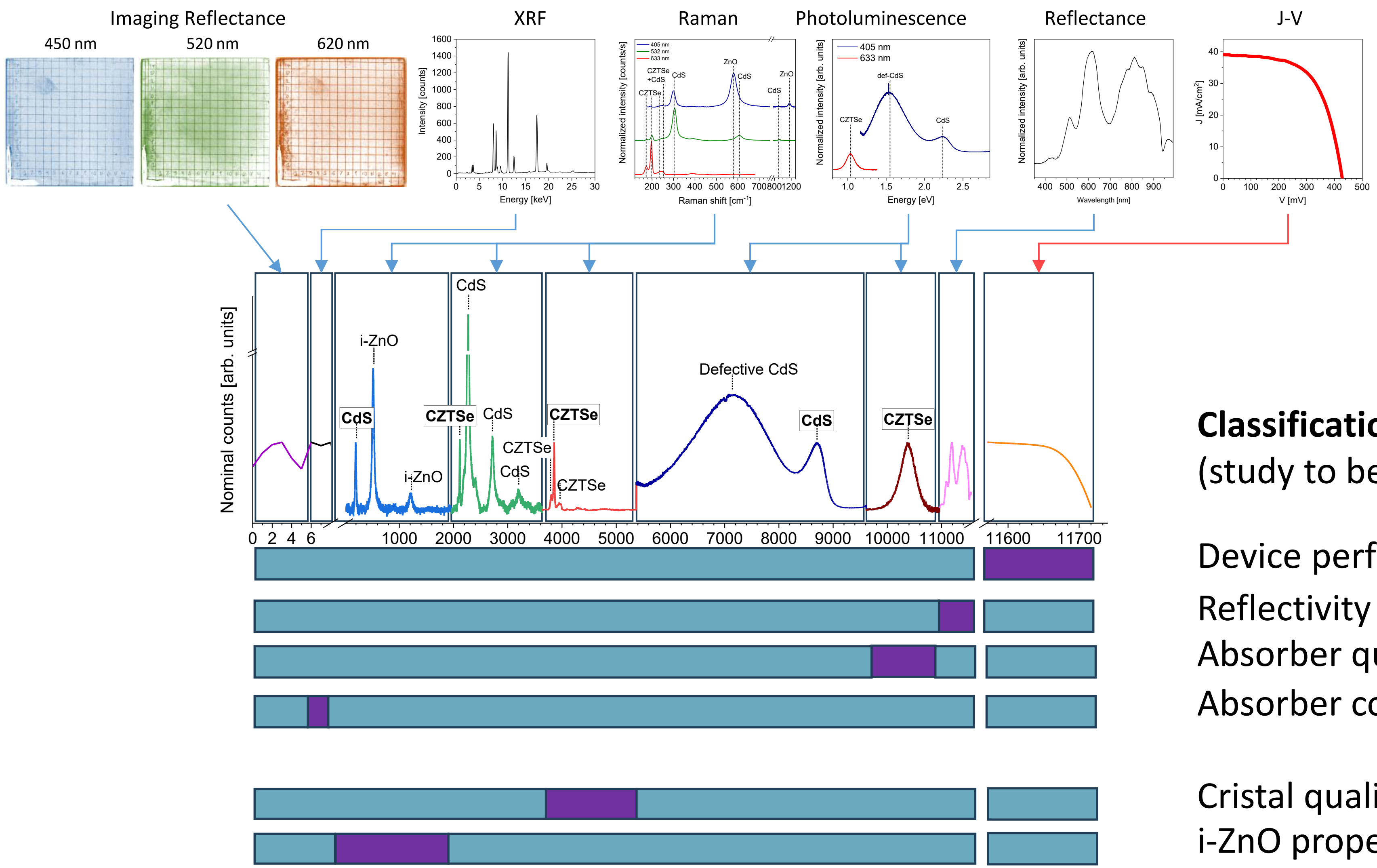


However, we want to understand the underlying physics.





Data classification



Data for LABELING

FEATURES for AI ingestion

Classification criterion
(study to be carried out)

Device performance

Reflectivity

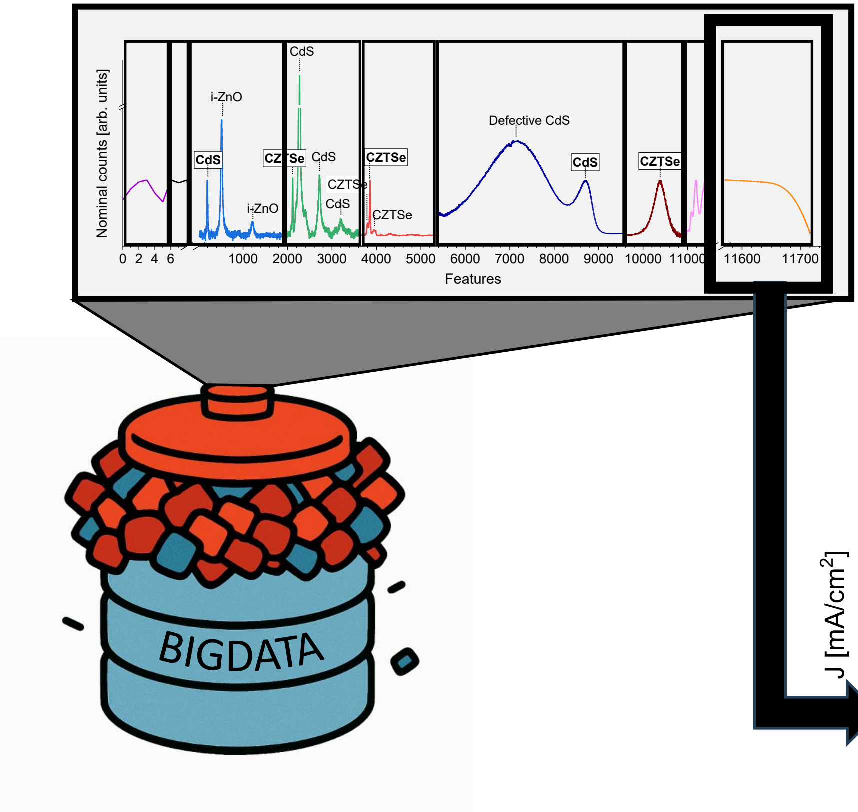
Absorber quantum yield

Absorber composition

Cristal quality of the absorber

i-ZnO properties

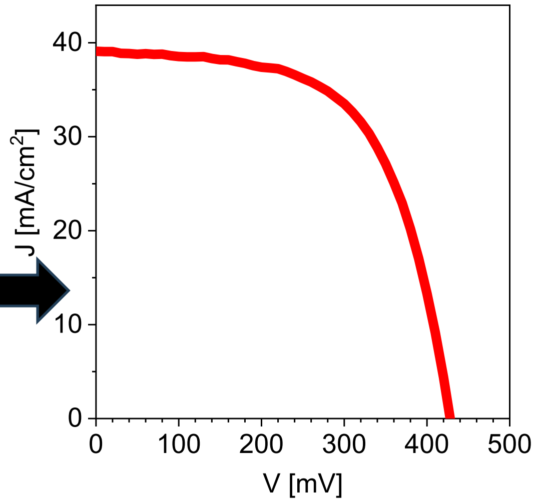
DATA CLASSIFICATION

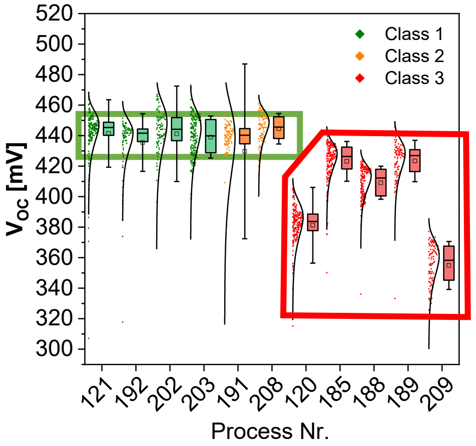


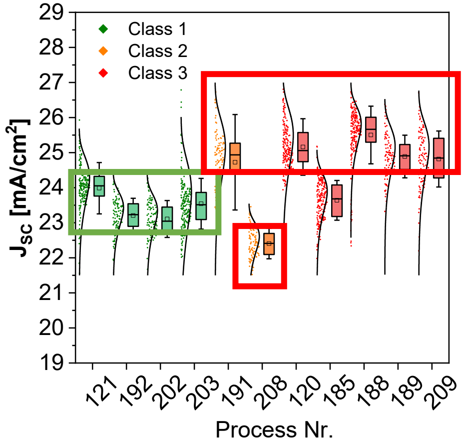
Selection of property to validate the research
(Data for labeling)

Distillation of the quantifiable parameter
(labeling)

Discretization in relevant groups
(Classification)



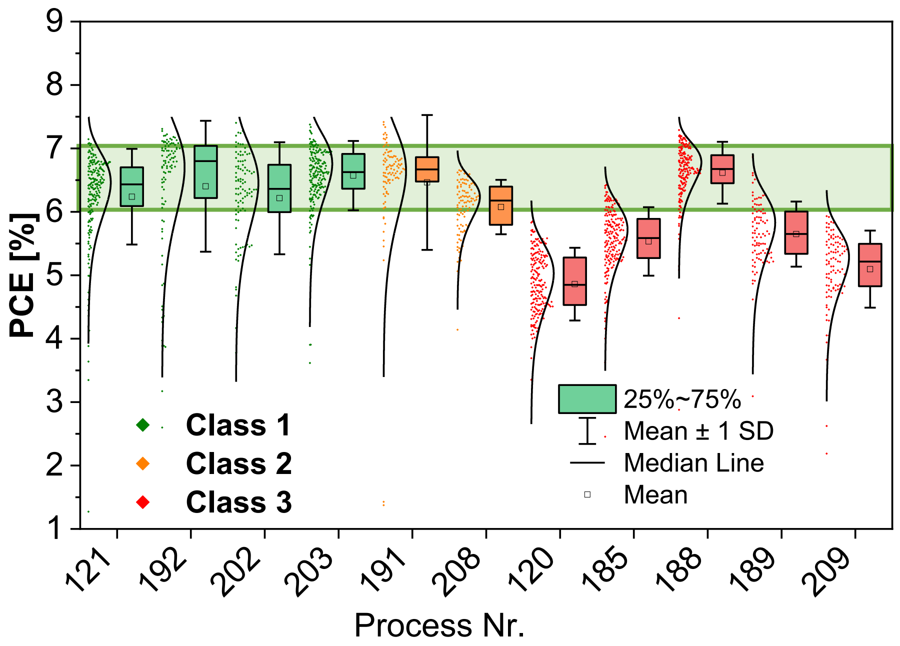


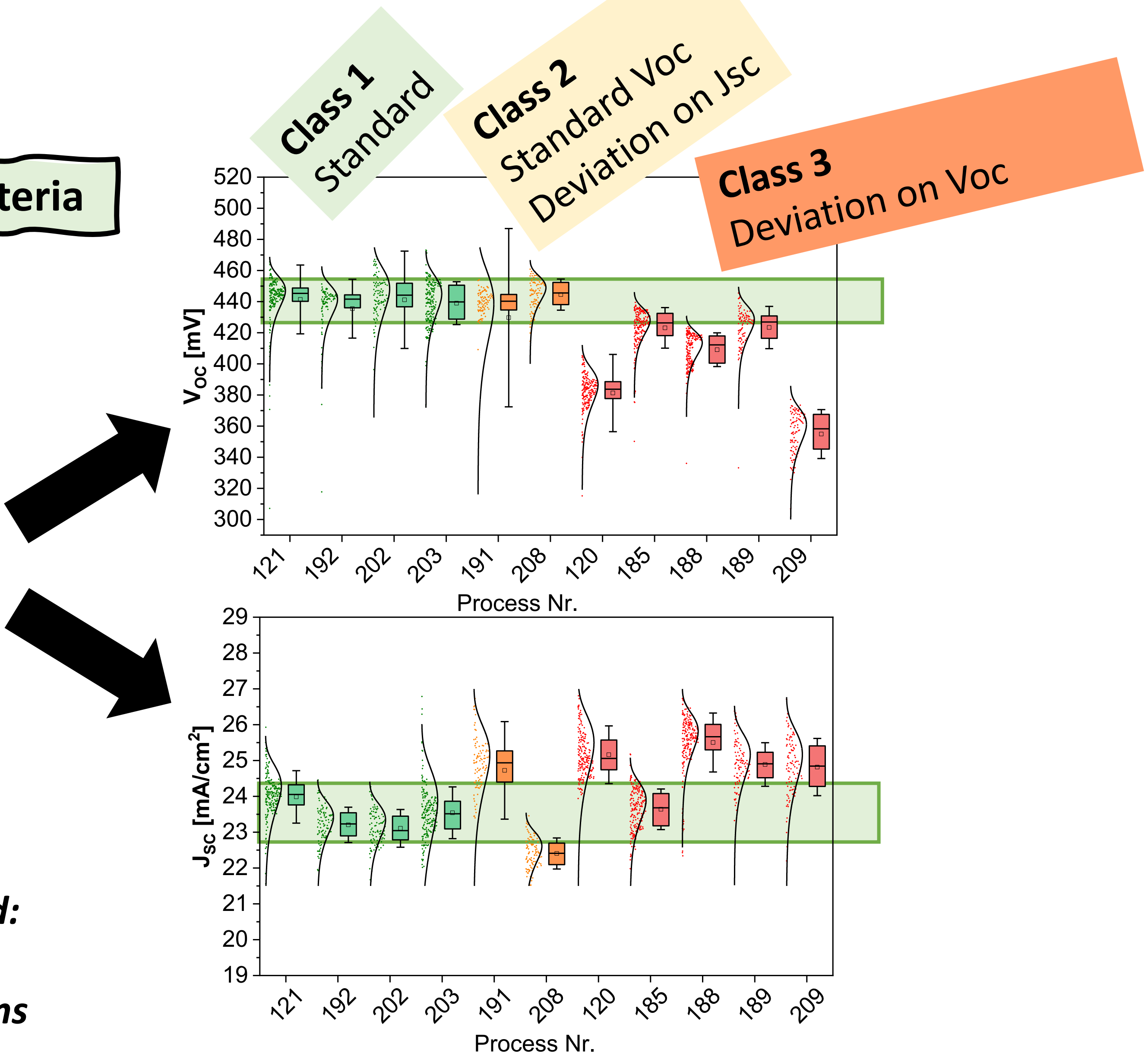


Class	Voc (mV)	Jsc (ma/cm²)
1	>440	>22.5-24.5
2	>440	<22.5 >24.5
3	<440	--

Labeling (classification criteria)

Definition of classification criteria

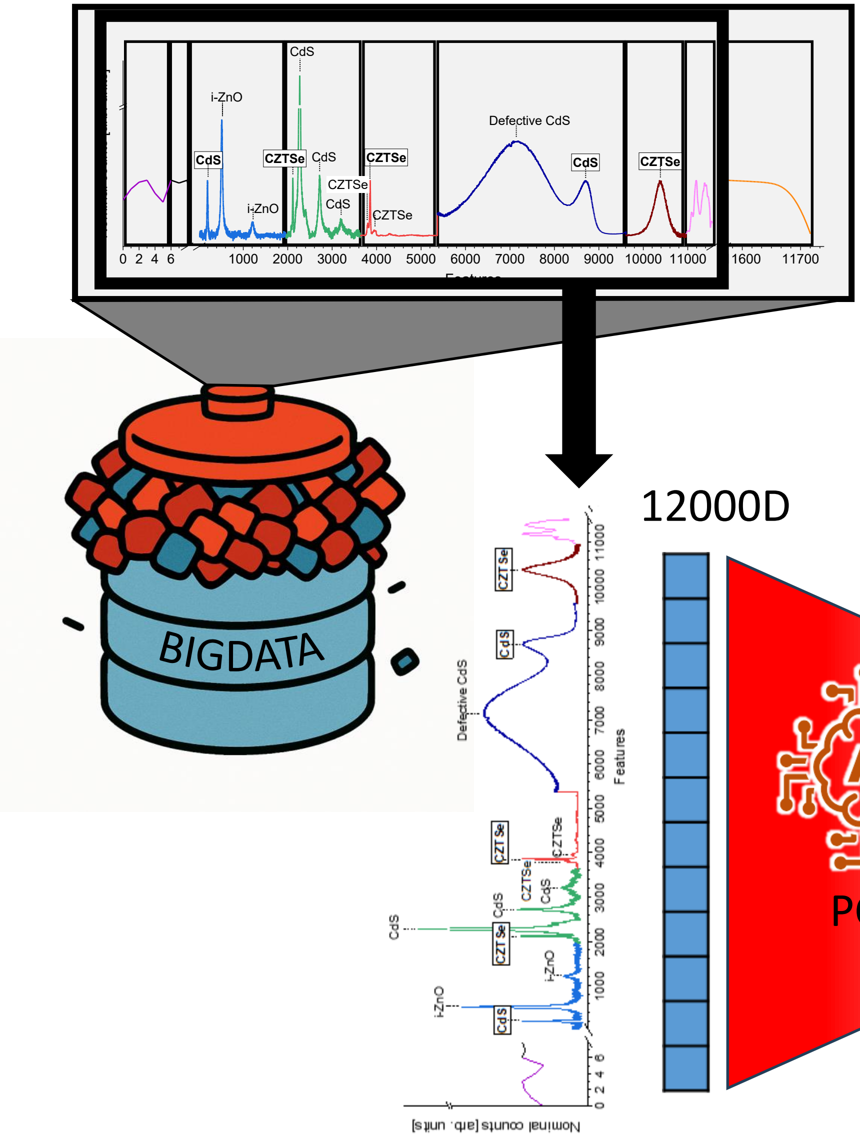




11 samples (≈2000 cells) have been selected:

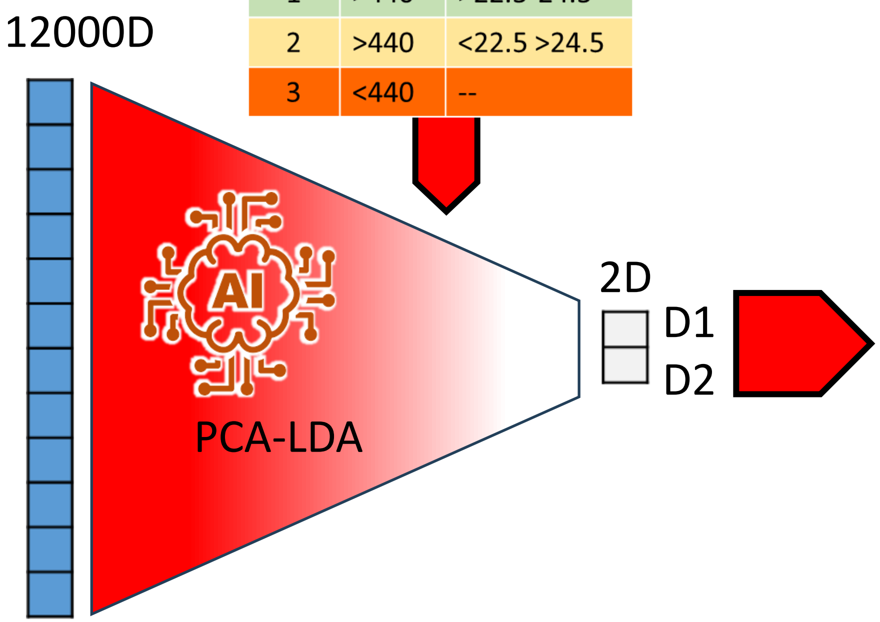
- ***4 with standard J_{sc} and V_{oc} values***
- ***7 with non-controlled process fluctuations***

(process control) What are the origin of the process fluctuations?

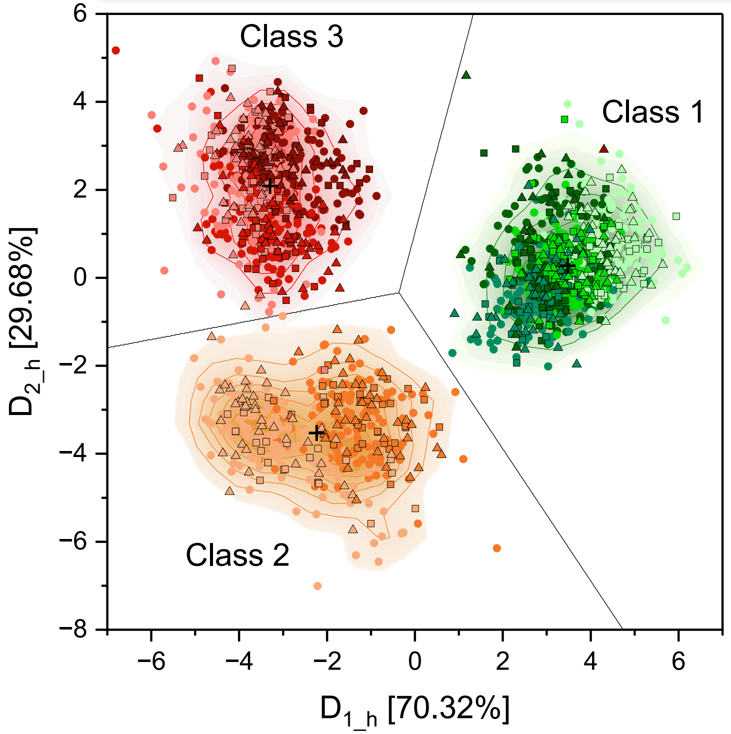


Discretization in relevant groups (**Classification**)

Class	Voc (mV)	Jsc (ma/cm²)
1	>440	>22.5-24.5
2	>440	<22.5 >24.5
3	<440	--

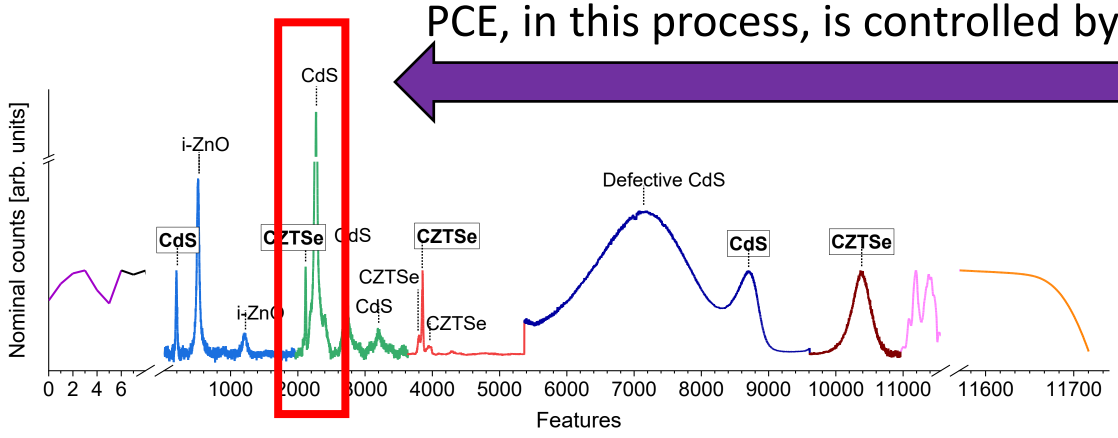


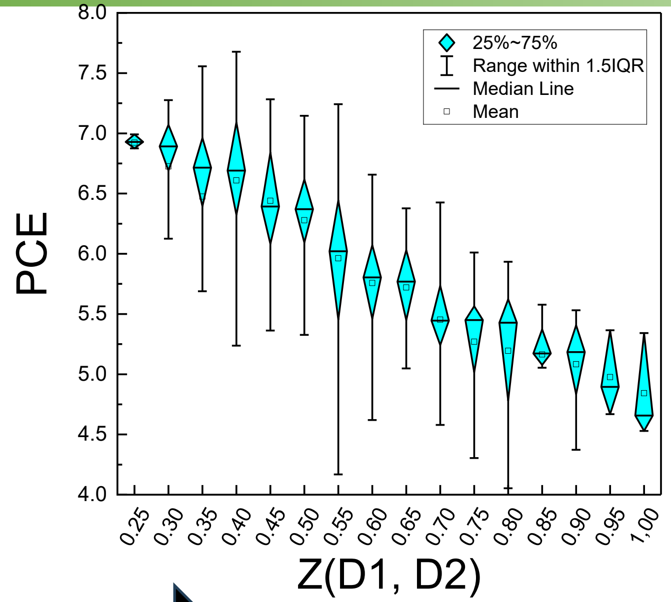
Latent space
Cross-validated Score: 99.71%



AI MODEL APPLICATION: EXPLICABLE AI (XAI)

PCE, in this process, is controlled by CdS thickness





$$Z = \mathcal{F}(D1, D2) = \mathcal{F}(\mathcal{G}_1(Features), \mathcal{G}_2(features))$$

Polinomial Expansión

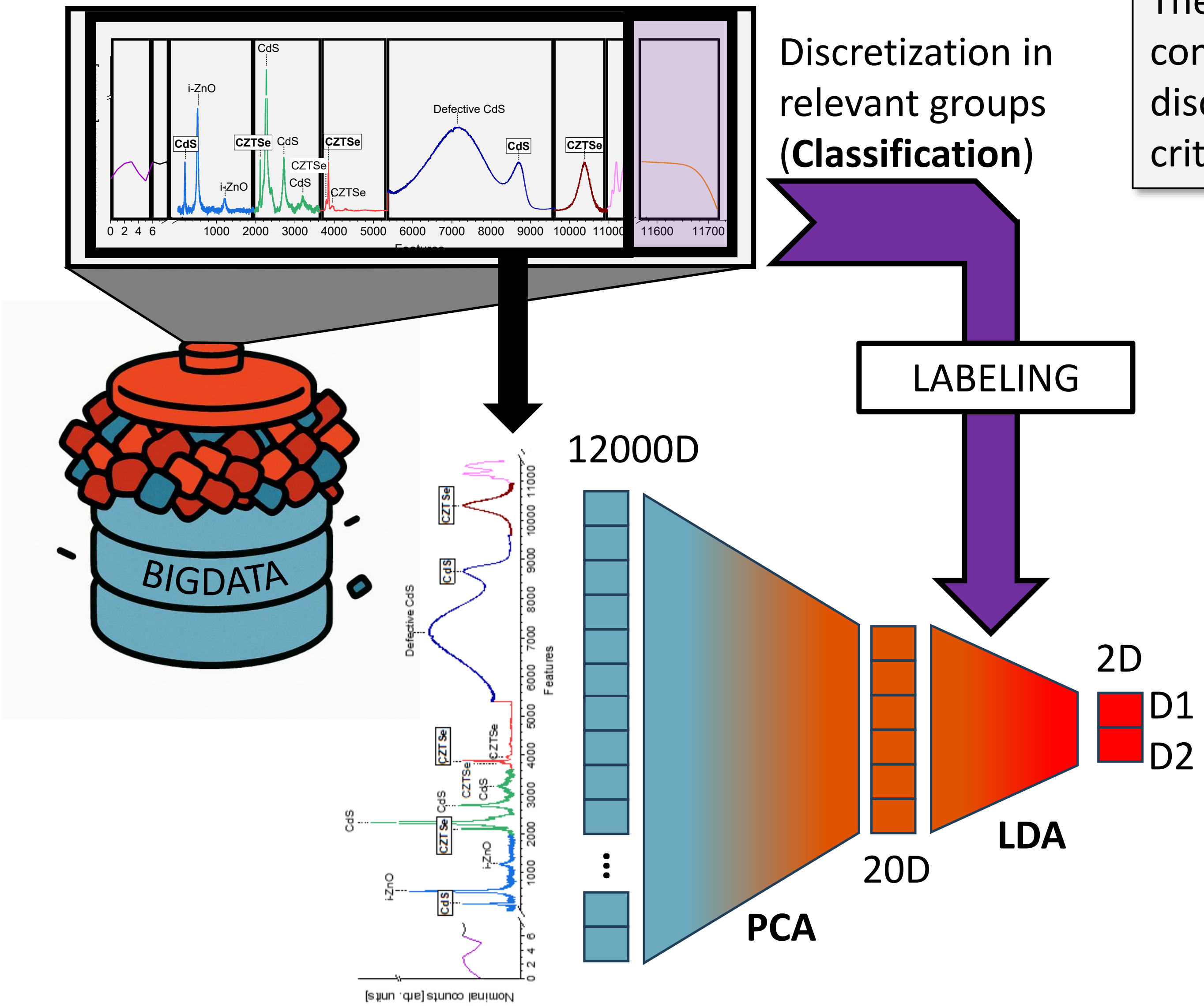
$$Z \backsimeq a + a_{11} \cdot F_1 + a_{12} \cdot F_2 + a_{13} \cdot F_3 + \cdots + a_{21} \cdot F_1^2 + a_{22} \cdot F_2^2 + \cdots + b_{12} \cdot F_1\, F_2 + b_{13} \cdot F_1\, F_3 + \cdots + O(f^3)$$

In the training process, we know the input vector, we can deconstruct the AI model and identify which parts of the input are relevant to the defined target, thereby providing scientific insight.

Revealing the Impact of CZTSe/CdS Interface Fluctuations on PV Device Performance through Big Data Analysis Assisted by Machine Learning Methods (SMALL METHODS 20,2025, 2400661)

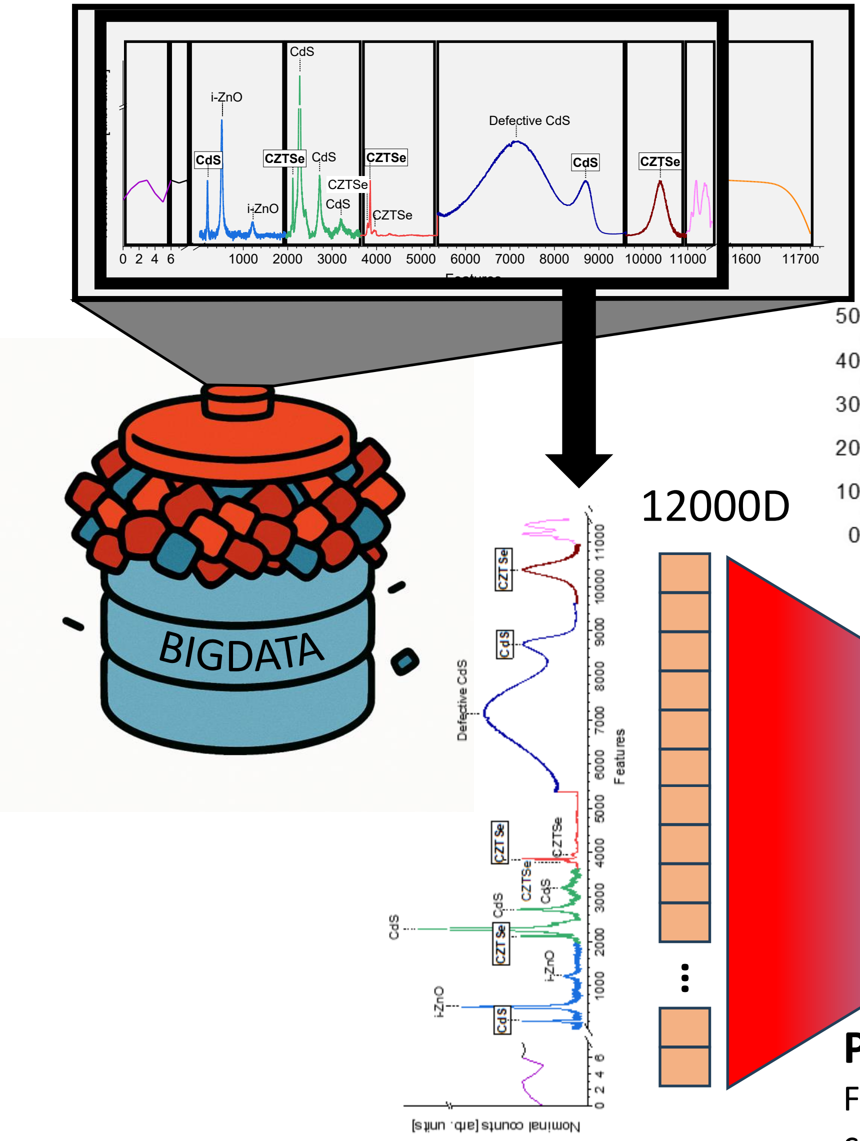
AI MODEL TRAINING

The AI model reduces data dimensionality using a combination of input features that maximizes discrimination based on the classification criterion.

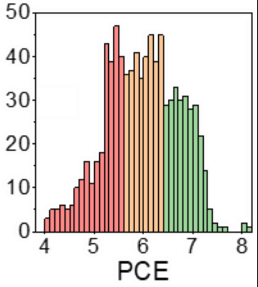


AI MODEL TRAINING

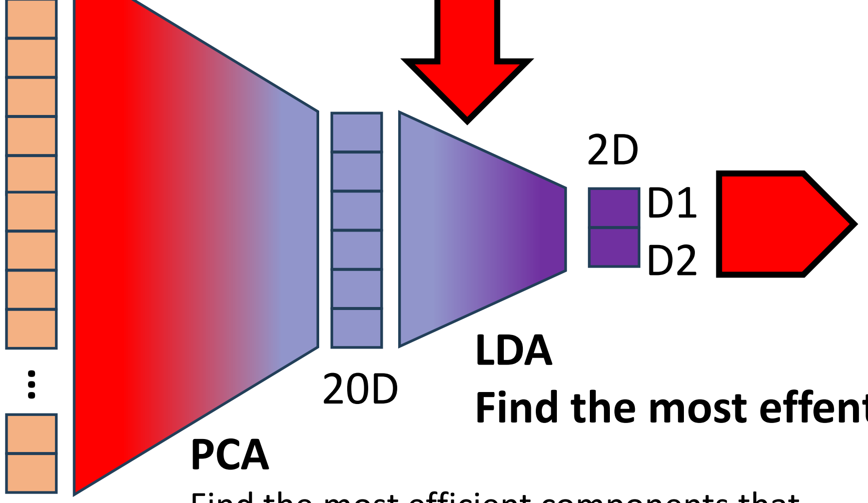
The AI model reduces data dimensionality using a combination of input features that maximizes discrimination based on the classification criterion.



Discretization in relevant groups (**Classification**)



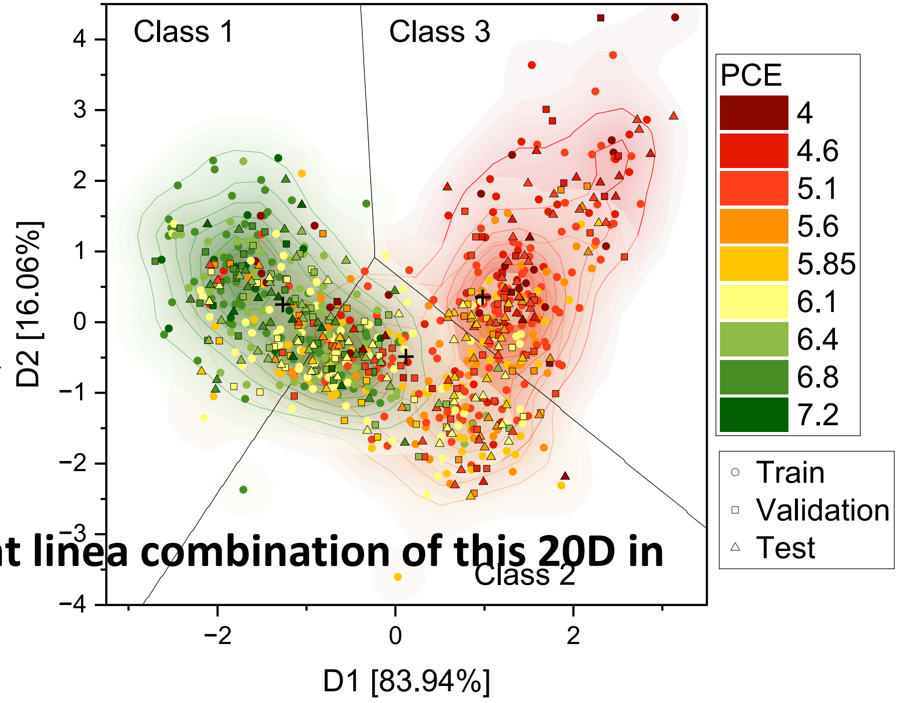
Class1: PCE >6.4%
Class2: 6.4%>PCE>5.6
Class3: 5.6>PCE



LDA
Find the most effent linea combination of this 20D in

PCA
Find the most efficient components that calcification to separate in n (20D) classes

Latent space
Cross-validated Score: 56.55%



Methodology: workflow for the application of AI-assisted research

Descriptors (data signal)

Characterization techniques → Data acquisition Automatization

Structured database

Data conditioning → Data fusion → Database splitting

Labels (Target)

Relevant device parameters → Creation of classification criterion

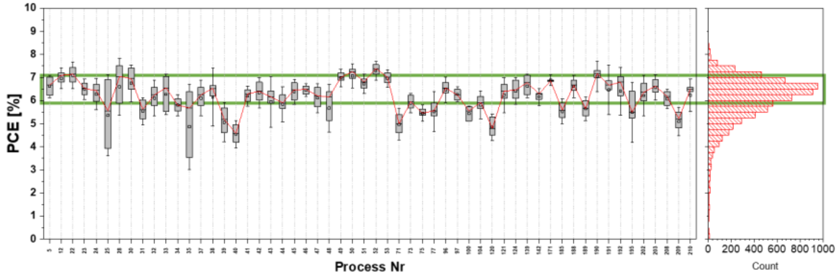
ML Training data 70%

ML Model Testing data 30%

Which is the technological limitation?

Enric Grau-Luque, et al
Accelerating the Development of Thin Film Photovoltaic Technologies: An Artificial Intelligence Assisted Methodology Using Spectroscopic and Optoelectronic Techniques, Small Methods (2024) 2301573/17.

PV Technology limitation



AI semi-supervised classification algorithm

PCA Algorithm
Unsupervised dimension reduction

LDA Algorithm
Supervised dimension reduction

≈10000D → 20D → 2D

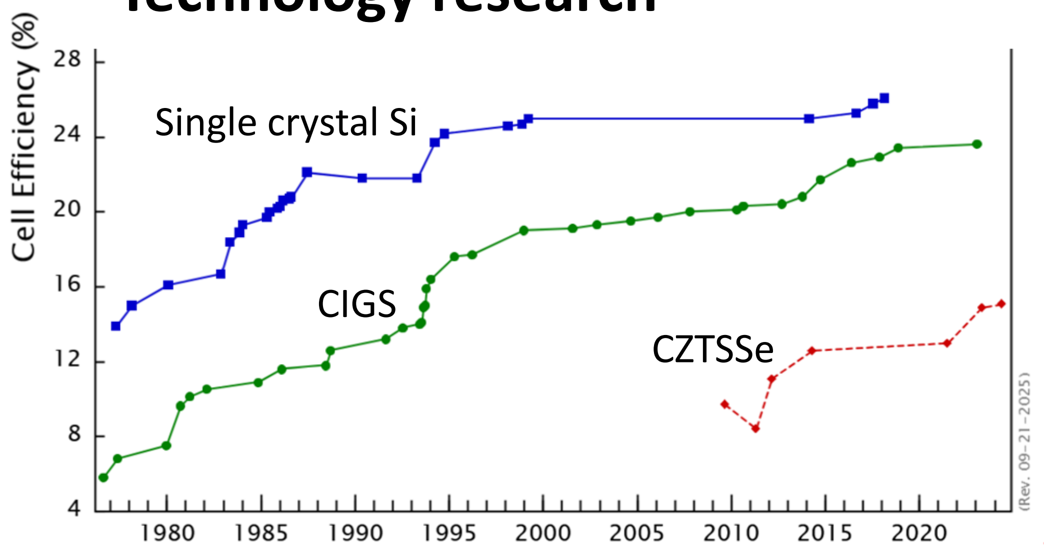
Technology Optimization strategies

Research and knowledge generation

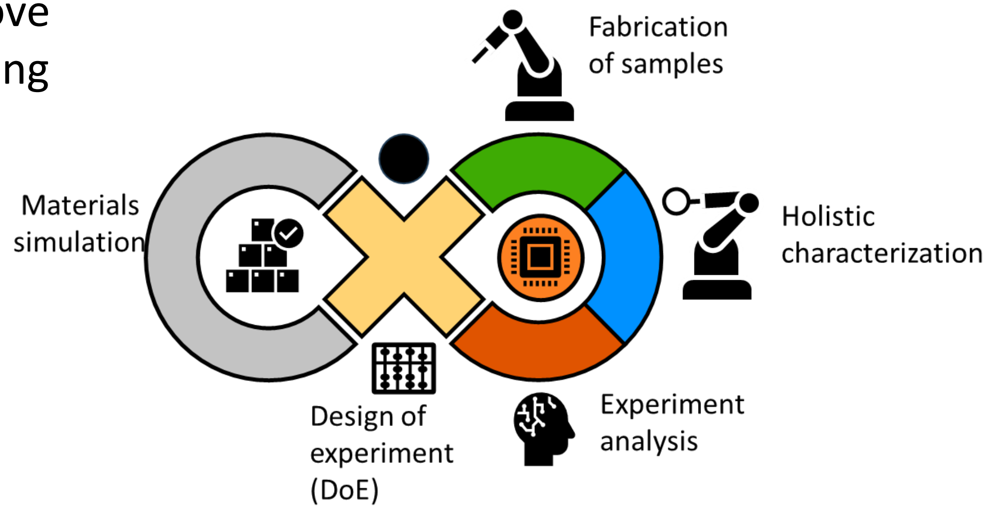
research in relevant properties / process steps ← Identification of significant features ← descriptors validation

Motivation: System complexity

Technology research



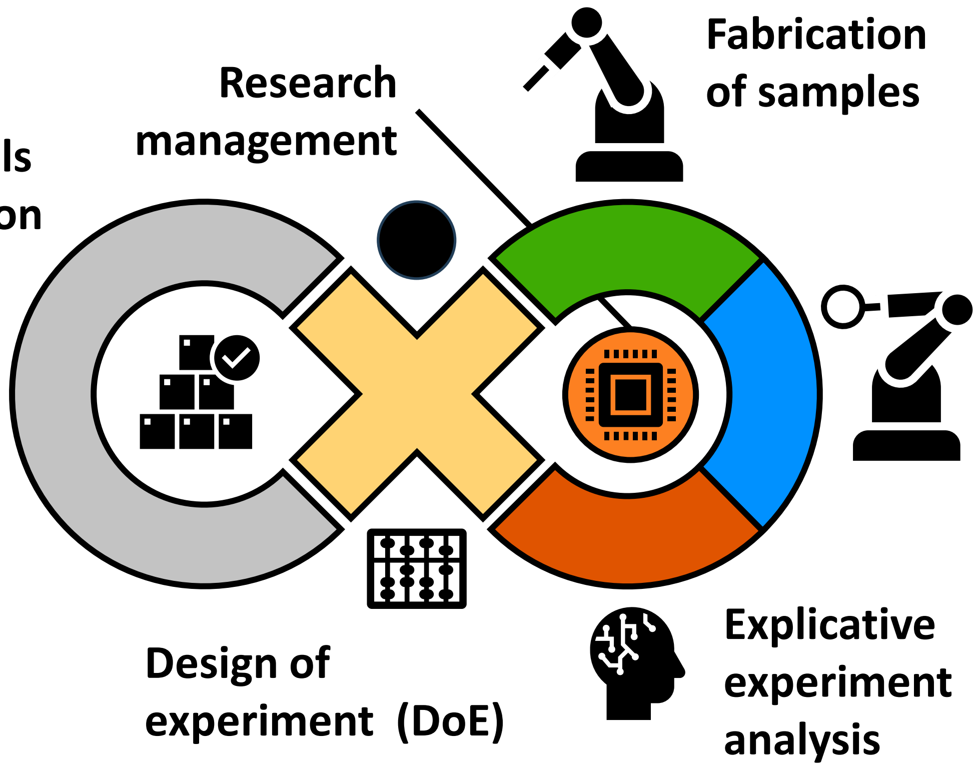
Motivation to move towards self-driving laboratories



Handicaps in materials research

- **Multidimensional complexity:** large number of variables (synthesis, structure, properties) complicates correlation between parameters.
- **Extended time scale:** development of a new material often takes 10–20 years before reaching commercial application.
- **High resource demand:** significant costs in equipment, energy, and consumables.
- **Limited reproducibility:** results strongly depend on researcher expertise and specific experimental conditions.
- **Massive data generation:** difficult integration and analysis of heterogeneous information (optical, electrical, structural, etc.).
- **Others…**

Motivation: Self-driving laboratory



Fabrication of samples
- Robotization of process
- Synthesis of relevant set: Different conditions synthesis & process steps

Holistic characterization
- Robotized ion of the process
- Combination of different techniques
- Optimized methodologies

TALK 2AO.2.2

Explicative experiment analysis
- Heterogeneous data fusion
- Multiangle dimensional analysis
- Acceleration of Interpretability

TALK 2AO.2.4

Relevance of self-driving laboratories

- **Automation** of synthesis and characterization to accelerate results.
- **Integration of automatic DoE** to design and optimize experiments in real time.
- **Higher reproducibility** and reduction of human bias.
- **Broader exploration** of material spaces compared to traditional methods.
- **Continuous optimization** enabled by explainable AI algorithms.
- **Reduced costs and shorter timelines** for developing innovative materials.

This concept can be applied in an industrial context as an optimization platform or as a tool for process monitoring and quality control.