Funded by the European Union

Computer Vision Insights from the Development of MicrostructureDB

24th of October 2024

Rina Jaeken*, Daria Tomecka*, Sepideh Khalatabad**, Leo Kestens**, Stefaan Cottenier**, Michaël Sluydts*

The AID4GREENEST project has received funding from Horizon Europe under Grant Agreement nº 101091912

Contact: catharina.jaeken@epotentia.com

AID4GREENEST

Project Partners

Spain

- IMDEA Materials Institute
- Reinosa Forgings and Castings
- Spanish Association of Standardisation

Belgium

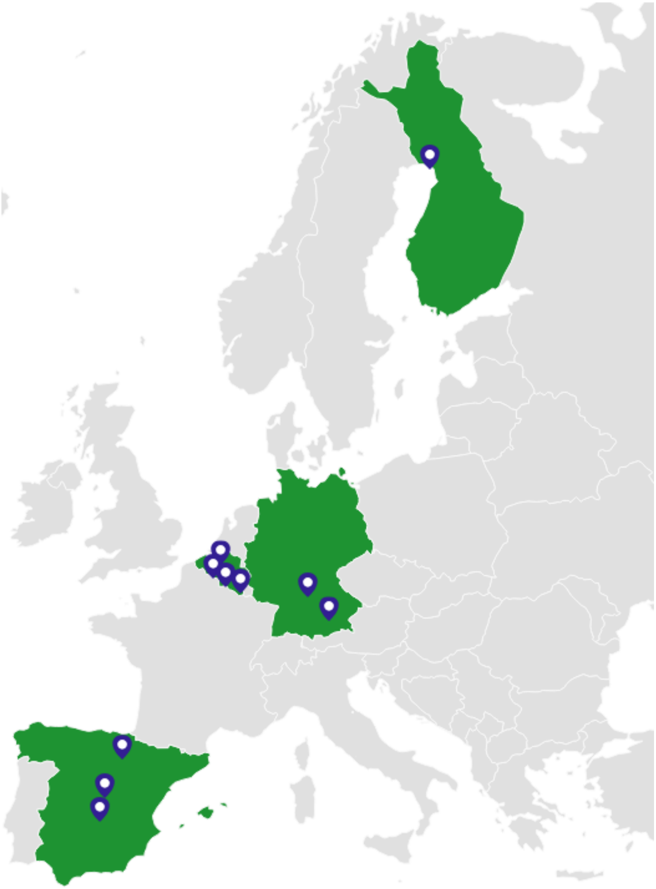
- Ghent University
- University of Liege
- OCAS NV
- ePotentia

Germany

- Eura AG
- Fraunhofer Society

Finland

- University of Oulu



AI Powered Characterisation and Modelling for Green Steel Technology

What is our challenge?

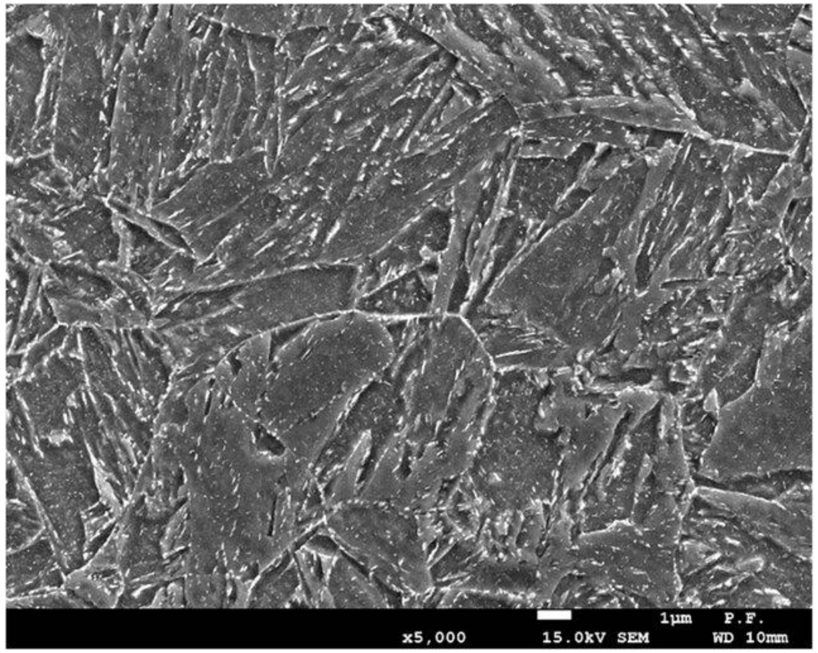
Advanced steels are becoming more and more complex

Creating, characterizing, and testing new steels costs a lot of time, money, and resources

Can AI help us accelerate steel characterization?

Two experimental methods:

- EBSD for high quality SE images to establish process – structure – property links
- Creep testing for understanding reliability of steels under load



Source: DOI://10.3390/ma13030747

We wish to accelerate both methods using a combination of advanced experimental techniques and AI ePotentia

Data•Science•Cloud

www.epotentia.com

ePotentia in AID4GREENEST

MicrostructureDB

OCAS business case

AI-enhanced characterization

- EBSD information from SE
- Determine processing from structure (regression)
- Structure from processing (generative AI)

ePotentia business case

AI-enhanced data management

- Sorting unlabeled data using AI representations
- Detect anomalies
- Fill in missing data with AI predictions
- Create captions and explain predictions
- Help standardization

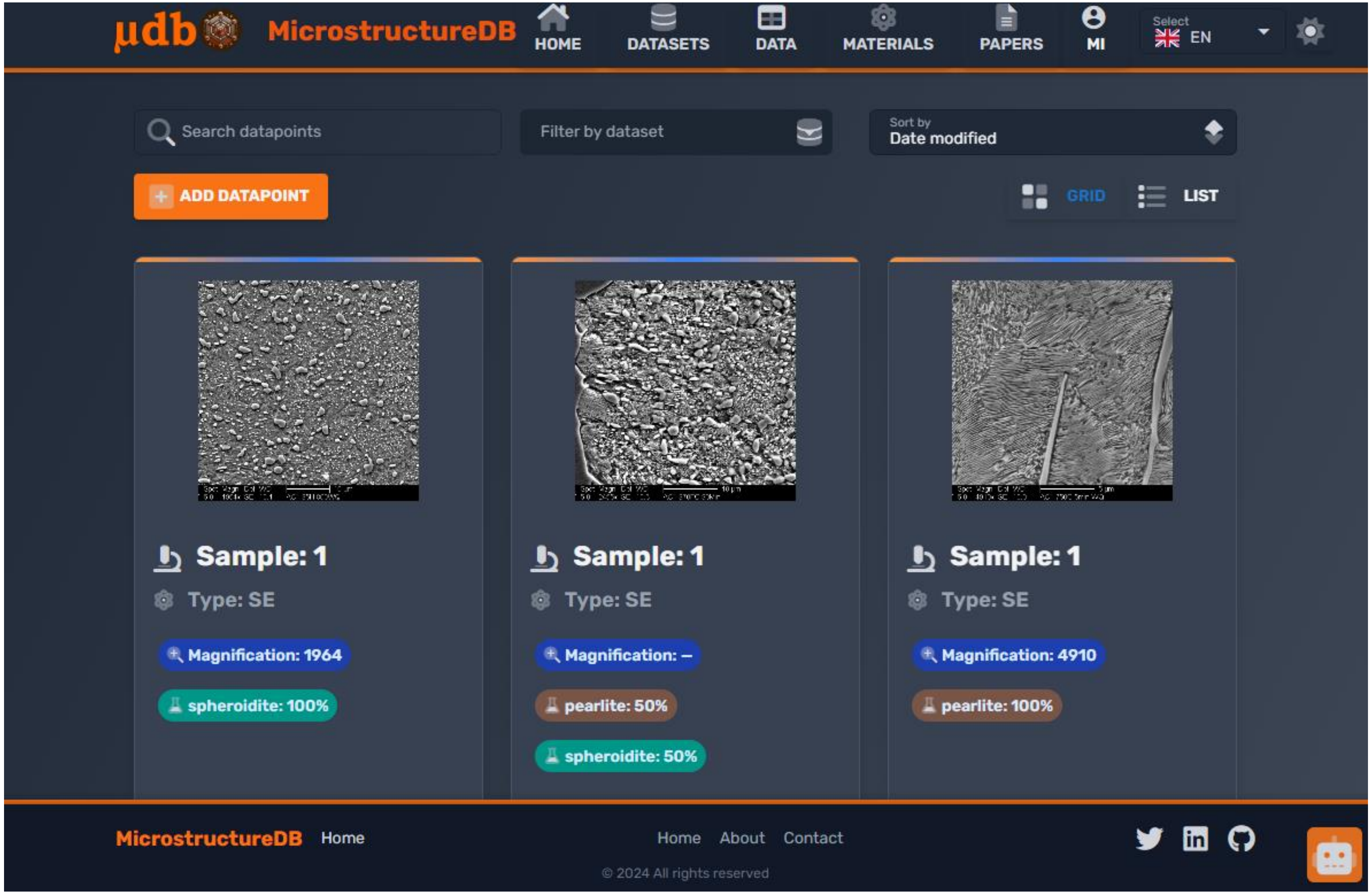
MicrostructureDB is meant both as a tool and a future product in both the cloud and the edge

MicrostructureDB

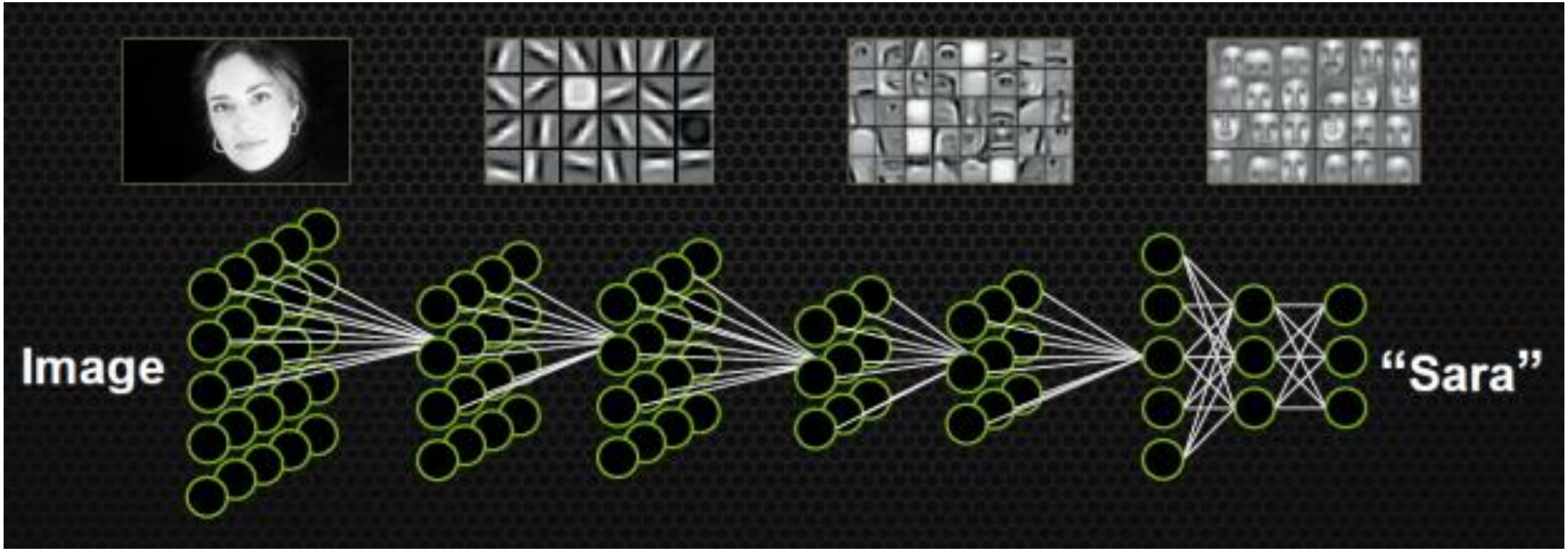
- Open repository of microstructural data
- AI-enhanced lab management system
- Different privacy levels:
  open – model training only – private
- Image-based search, classification, automated annotation and anomaly detection
- Credit-based system: those who provide data are rewarded with compute credits
- Integration with CHADA/MODA

https://microstructuredb.com

MicrostructureDB: work in progress

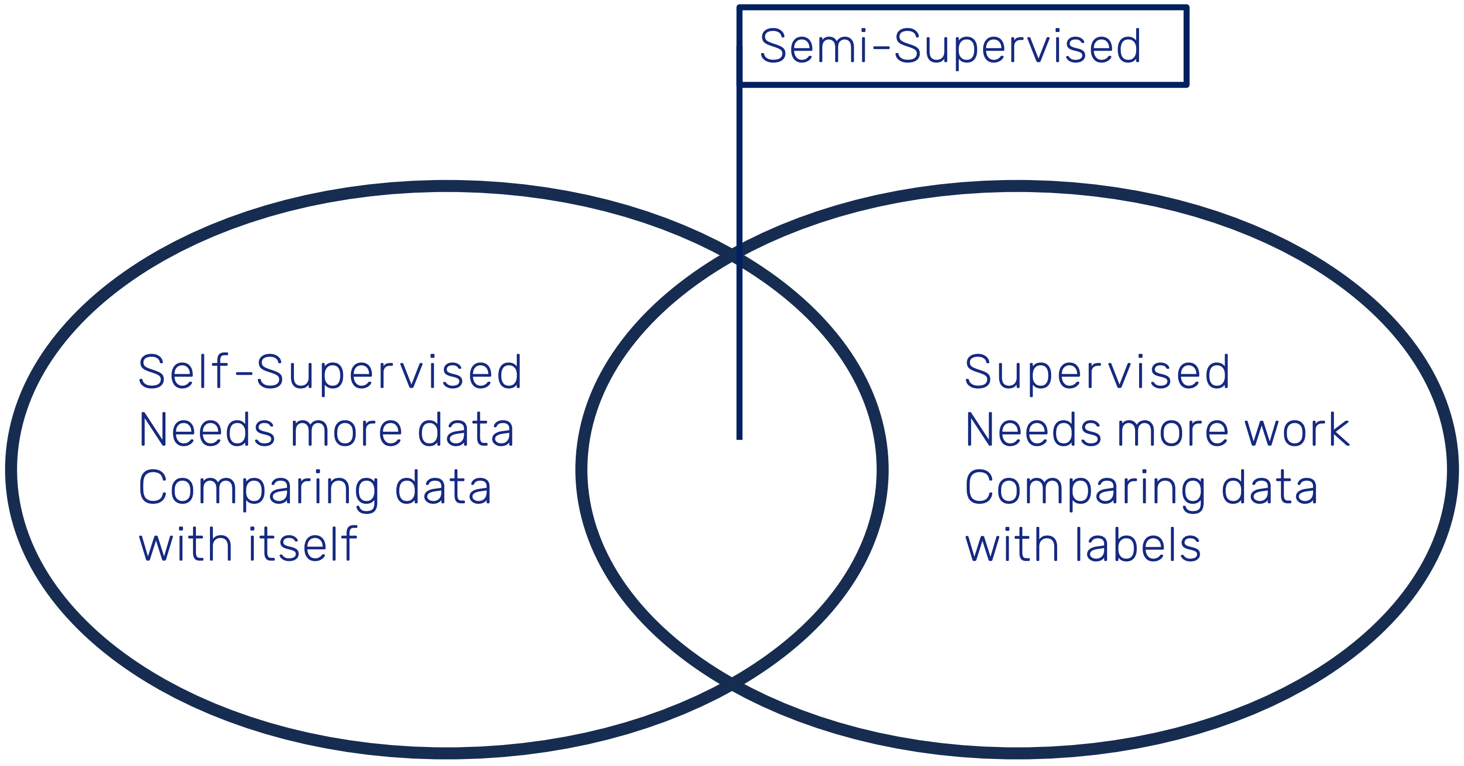


Deep learning



Given enough good data deep learning will organize your data for you

DL paradigms Overview



Representation Learning:

The process of automatically discovering the representations or features needed for a specific task from raw data.

Supervised learning: UHCSDB

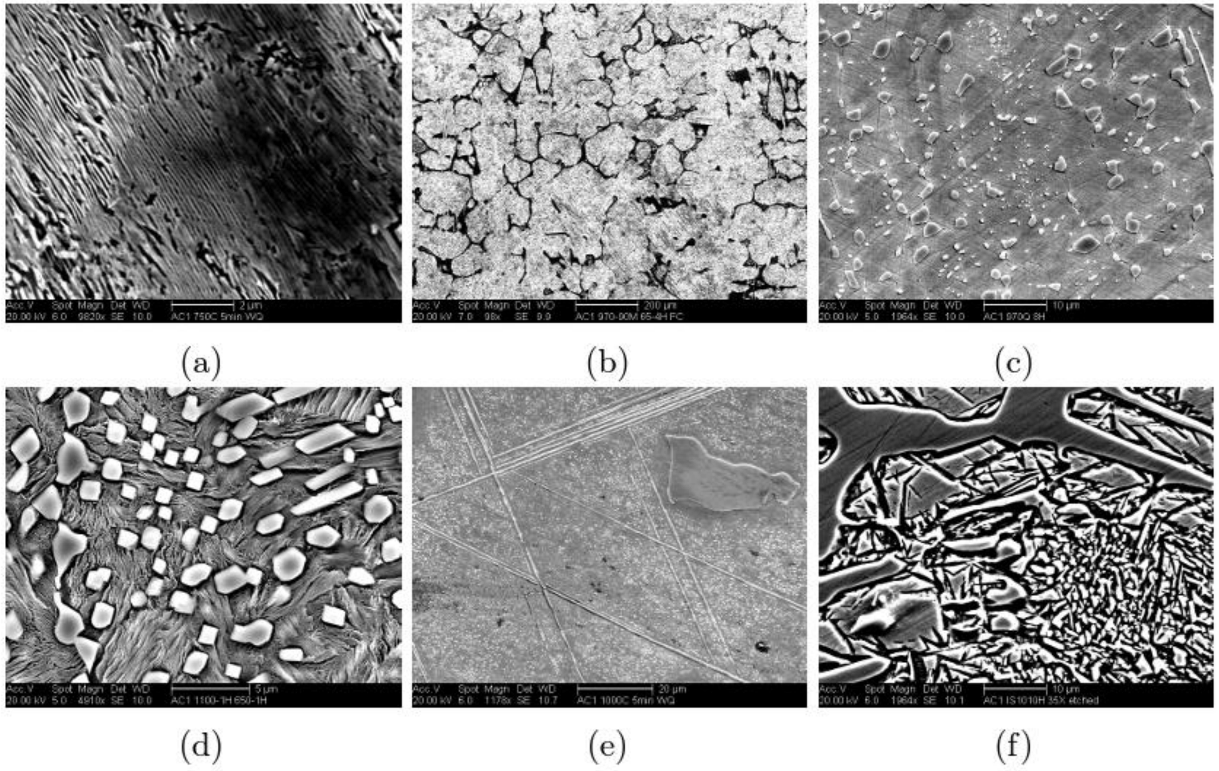


Figure 1: Primary microstructure constituents in the UHCS dataset: (a) pearlite, (b) proeutectoid cementite network microstructure, (c) spheroidized cementite, (d) pearlite containing spheroidized cementite, (e) Widmanstätten cementite, and (f) martensite and/or bainite.

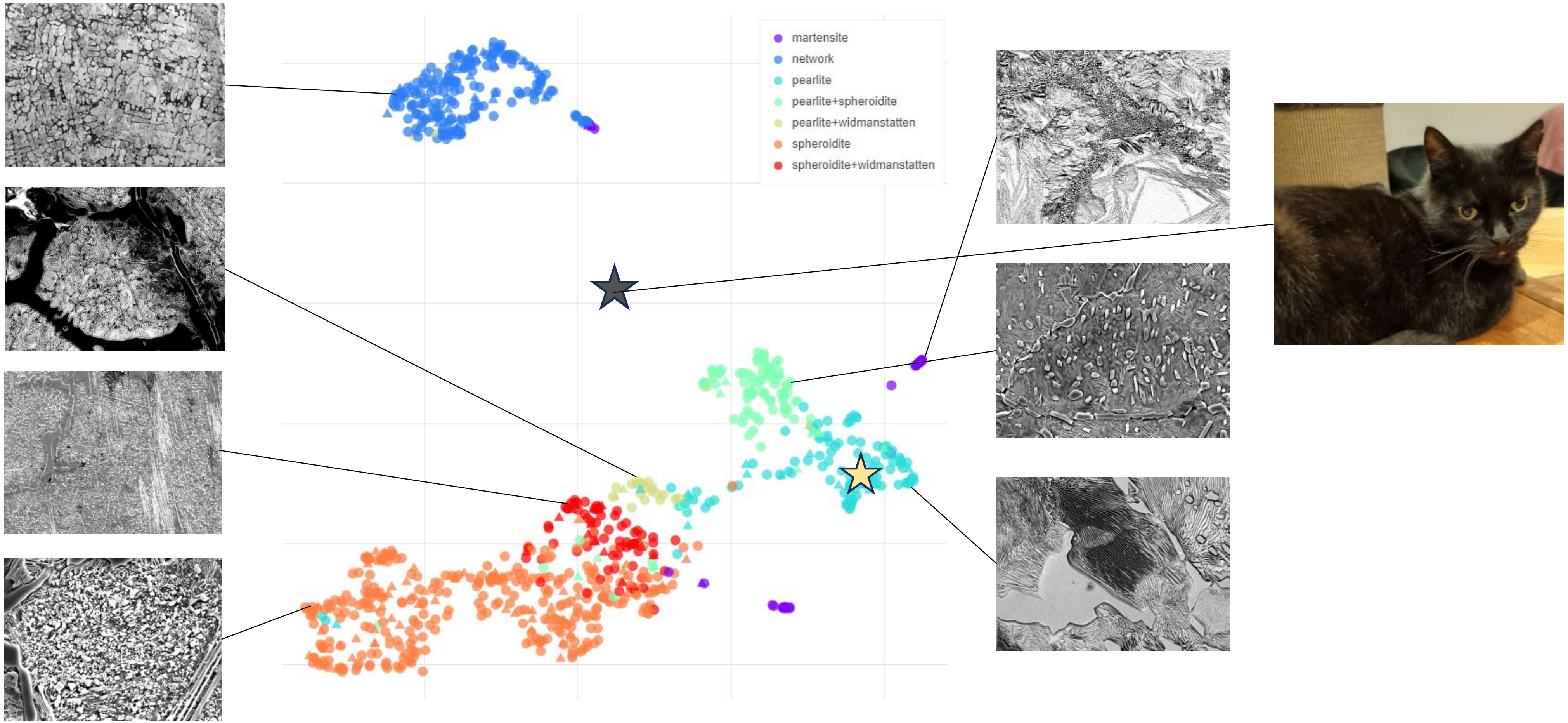
UltraHigh Carbon Steel Micrograph DataBase

Microstructure dataset consisting of ultrahigh carbon steel (UHCS) micrographs taken over a range of length scales under systematically varied heat treatments.

Open source.

Ultrahigh Carbon Steel Micrographs
Hecht, Matthew D. and DeCost, Brian L. and Francis, Toby and Holm, Elizabeth A. and Picard, Yoosuf N. and Webler, Bryan A.
https://hdl.handle.net/11256/940

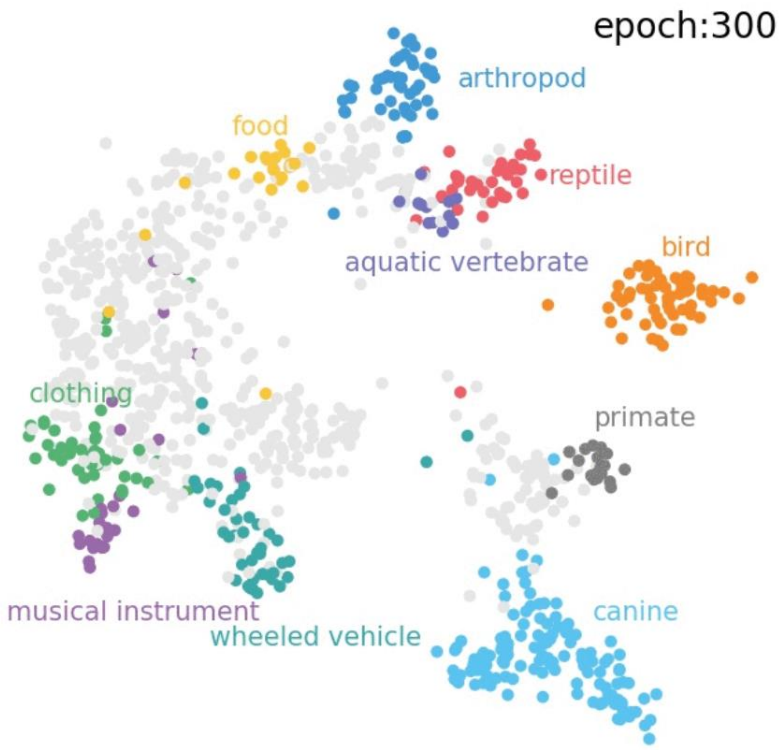
Creating a map of metals: UHCSDB



The structure of the map can be used to automatically label new data and find anomalies

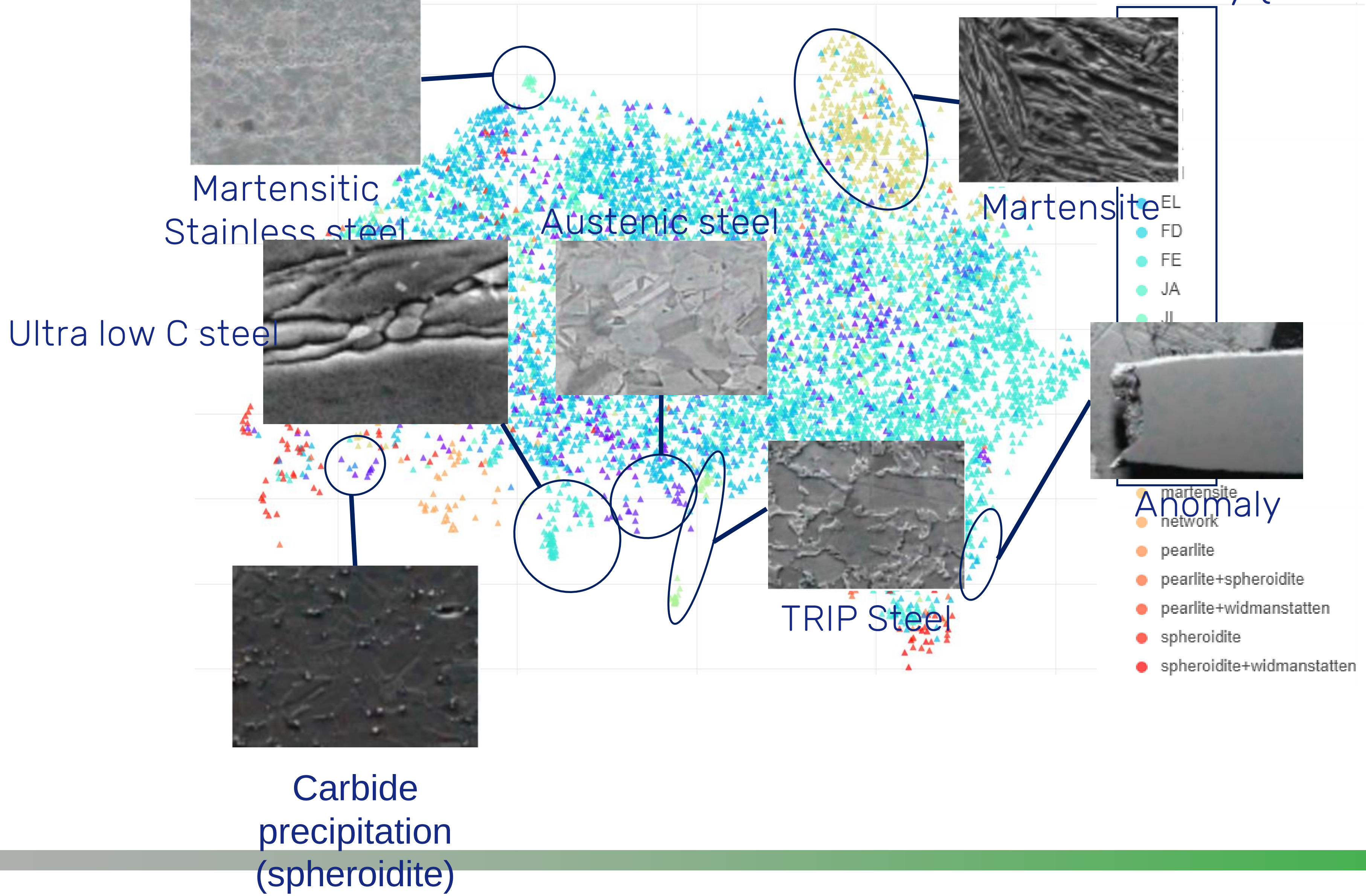
Self-supervised learning

Example: Automatic nearest neighbor classification of animals



Organizing a real dataset

We combine UGent internal data + UHCSDB to maximize diversity (+-10k SEM images).



From real to synthetic data

Real data

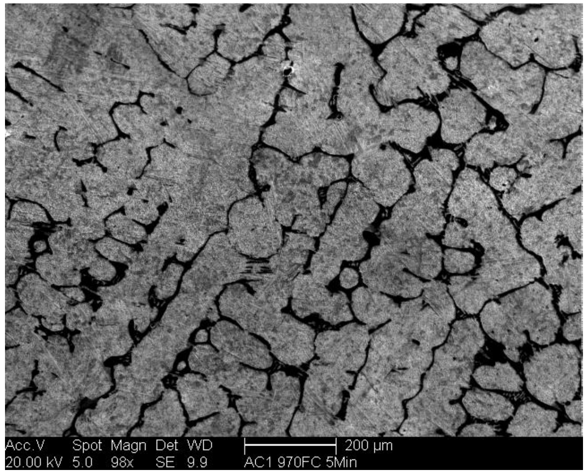
- Sparse, limited availability
- Expensive to create
- Biased towards certain parts of material space and image space

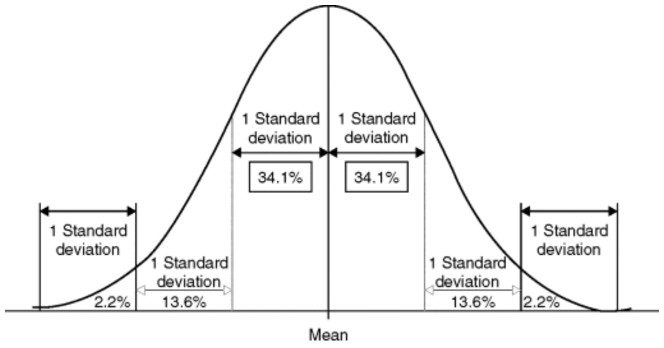
Synthetic data

- Useful for predictions
- Can “fill in the gaps” and expand dataset towards regions that are difficult to explore
- Can create inputs for other models
- We can study the synthetic data to understand what the model deems important

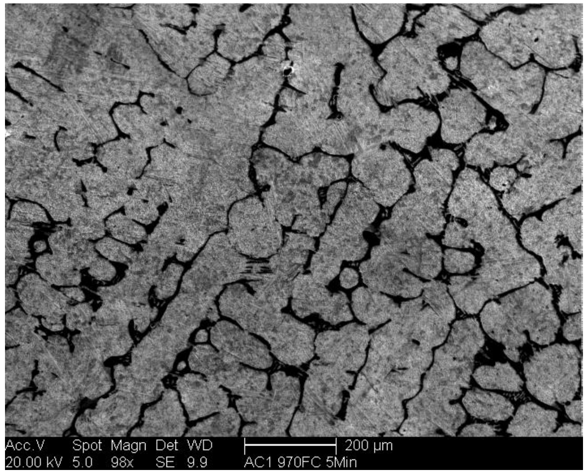
Training a synthetic data model

Training Image





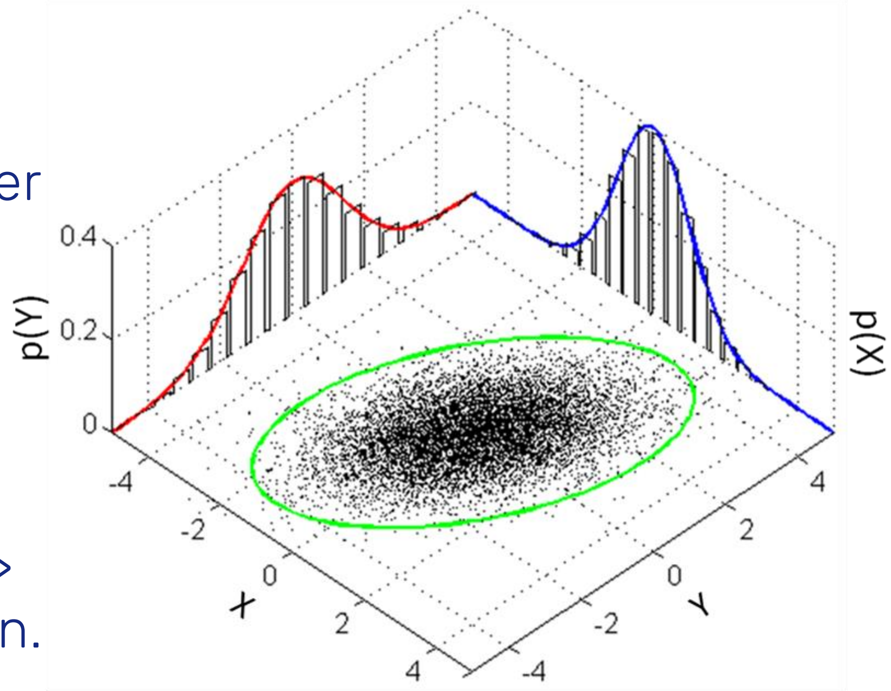
Generated image



Generative models fit a distribution which we sample from to generate new images, captions help us navigate this space

Synthetic data: property space

- Multivariate distribution with strong correlations
- By filling in the gaps we want to go from the sample closer to a full population, but we must keep in mind what population we want to approximate
- Certain correlations indicate physical reality (annealing temperature <-> microstructure)
- Other correlations indicate human interest (grain size <-> magnification). This is where materials expertise comes in.

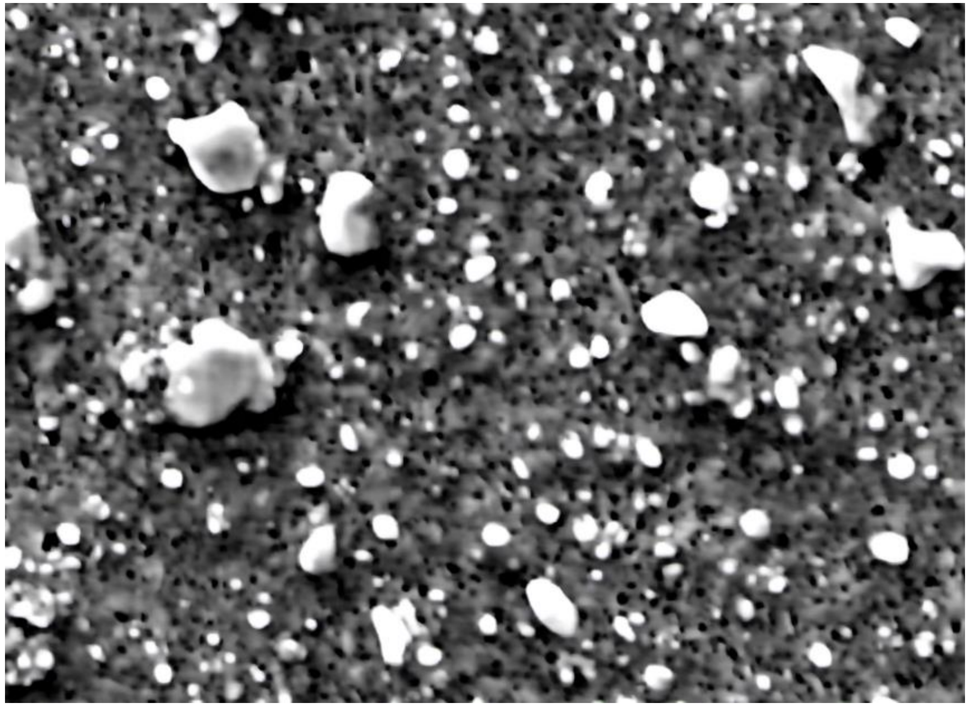


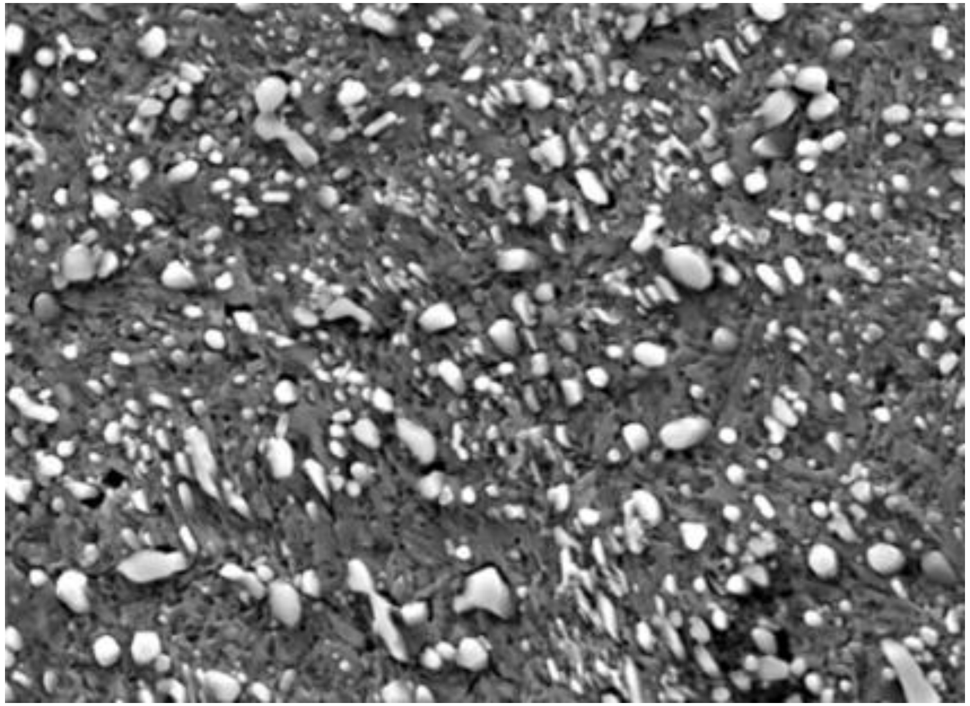
Quiz!

Interactive:

Can you tell the difference between real and synthetic microstructure images?

Synthetic data vs real





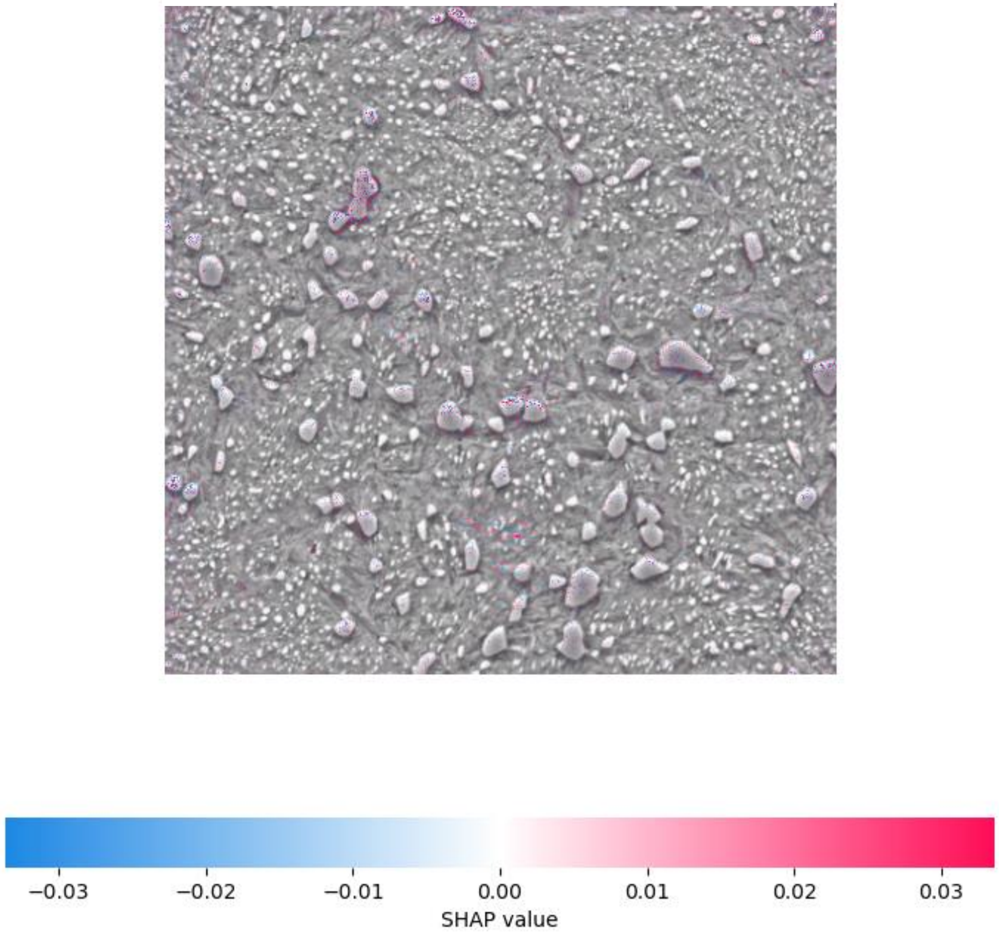
Learning from synthetic data

Training an artificial expert

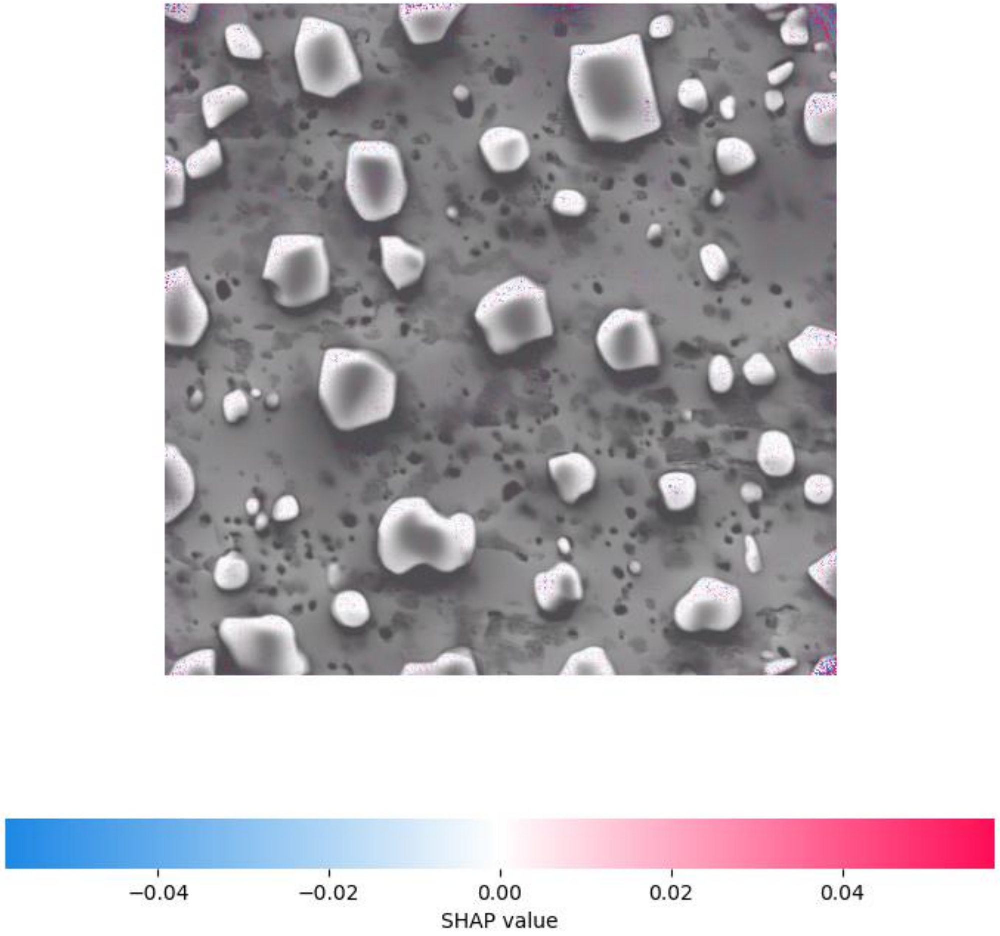
- Teaching AI to tell the difference between real and synthetic data
- Using explainable AI to see on what criteria the AI expert may consider the synthetic image unrealistic
- The goal is not to create good-looking images, but to create scientifically relevant data
- This can reveal hidden microstructural knowledge that our generator might have missed!

Explainable AI

Real



Synthetic



Conclusions

- AI offers a pathway to strengthen the links between composition, processing, microstructure, characterization, and properties of steel.
- A centralized knowledge platform will help not only humans, but also machines to navigate the space of steel microstructures and properties
- Representation learning allows us to sort and organize large datasets from various sources
- Generative AI allows to expand these datasets to make them more suitable for training further models and deepening understanding of steel
- Advanced modeling and AI tools can reduce time, effort, and material spent in R&D cycles towards new advanced steels
- We need expert knowledge and data: https://microstructuredb.com

European Commission

Thank you!

FEMS 2025
EUROMAT

18th European Congress and Exhibition on Advanced Materials and Processes

Granada,
14 – 18 September 2025

We will be at EUROMAT!

D9 – Artificial intelligence, modelling and data science in advanced alloy and process design

https://euromat2025.com/