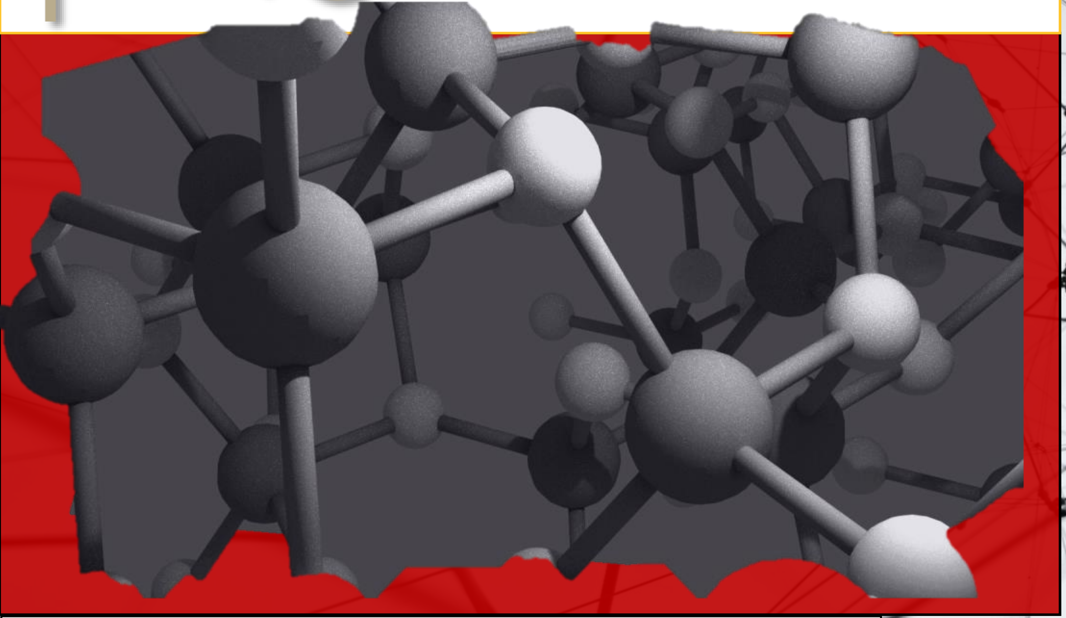
R-NANO Research Unit of Advanced, Composite, Nano-Materials & Nanotechnology

National Technical University of Athens
School of Chemical Engineering
Materials Science and Engineering Department

EuroNanoForum 2019



Applying machine learning to process and characterisation data of nanomaterials: A means for prediction

Elias P. Koumoulos
Costas A. Charitidis

IRES Innovation in Research & Engineering Solutions

RNANO-About us…

The **National Technical University (NTUA)** is the **oldest** and **most prestigious** educational institution of Greece in the field of technology, founded in 1836.

Campuses

Patision Complex

Averof building is one of the most important and elegant buildings of the Athenian Neoclassical period located in the centre of Athens.

Zografou Campus

The main campus is located in the Zografou area of Athens, housing all the schools of NTUA except architecture. The main campus spreads over an area of about **190 acres**, 6 km from the centre of Athens. It includes buildings of 65 acres with fully equipped lecture theatres, laboratories, libraries, gyms, a central library, a computer centre and a medical centre.

R-NANO

Extensive experience in Design, Production and Characterization of Advanced-, Composite- and Nano- Materials.

Head: Professor Costas Charitidis

People:

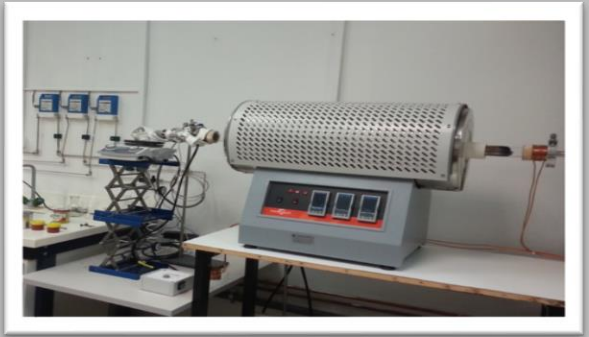
- 6 Professors
- 9 Post Doctoral Researchers
- 6 Researchers
- 10 PhD & MSc students

Clusters

- The European **Materials Characterisation** Cluster
- The European **NanoSafety** Cluster
- The European **Materials Modelling** Council
- The European **Carbon Fibre & Advanced Composites** Cluster
- The European **Pilot Production** Network

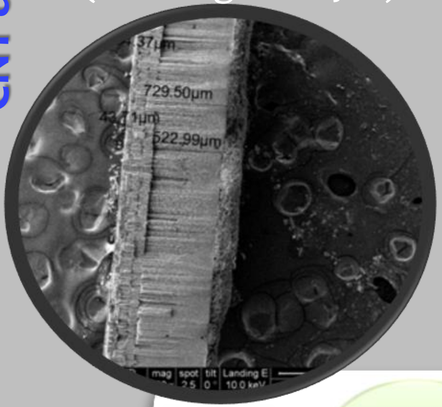
RNANO-About us…

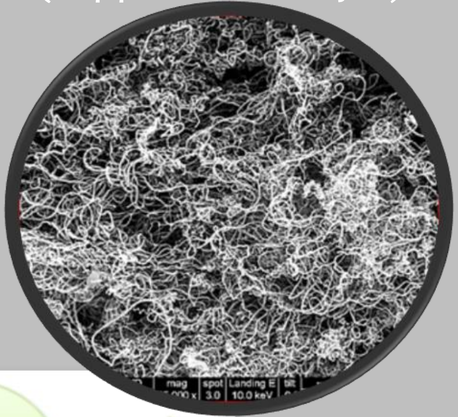
Thermal Chemical Vapor Deposition Method :
CNT & CNF



Vertically Aligned-CNTs
(floating catalyst)

Entangled CNTs
(supported catalyst)





Semi-Industrial scale Carbon Fibre Production

- Polymer blending
- Pilot melt spinning line – Single/Twin Screw extruder- Winding System - Continuous stabilization line – oxidative treatment on precursor fiber – **Continuous carbonization line (lab scale)**

Twin screw extruder

Winding system

Continuous stabilization line

Typical multifilament line for melt spinning

Post-processing and modification of CFs: CVD/EPD : CNT/CNF deposition on CFs

Relevant Projects: FIBRALSPEC, MODCOMP, SMARTFAN

Surface analysis and mechanical behaviour mapping of vertically aligned cnt forest array through nanoindentation EP Koumoulos, CA Charitidis Applied Surface Science 396, 681-687
Carbon nanotube/polymer nanocomposites: A study on mechanical integrity through nanoindentation, Elias P. Koumoulos Pravin Jagdale Ioannis A. Kartsonakis Mauro Giorcelli Alberto Tagliaferro Constantinos A. Charitidis, Polymer Composites, 36, 8, 2015 Pages 1432-1446
Assessing the Critical Multifunctionality Threshold for Optimal Electrical, Thermal, and Nanomechanical Properties of Carbon Nanotubes/Epoxy Nanocomposites for Aerospace Applications. Trompeta, A.-F.A.; Koumoulos, E.P.; Stavropoulos, S.G.; Velmachos, T.G.; Psarras, G.C.; Charitidis, C.A. Aerospace 2019, 6, 7.

IRES-About us…

IRES is an independent **consulting firm**, established in **2015**, that provides **specialized services** of:

- ✓ innovation management
- ✓ technology transfer
- ✓ support of European and International research projects
- ✓ enhancement and support for new and established Businesses.

Main Activities..

1. Life Cycle Assessment (LCA)

LCA enables the transformation of a scientific assessment into a **decision making tool** by taking into consideration the whole **life-cycle of products** and performing a holistic evaluation of social, environmental and economic aspects **based on EU standards and regulations.** The Combination of LCA with **Life Cycle Cost (LCC)** facilitates the extraction of financial information.

2. Data Management (DM) and Machine Learning (ML)

Data management is an administrative process that includes **acquiring, validating, storing, protecting, and processing, required data** to ensure the **accessibility, reliability**, and **timeliness** of the data for its users. A good DM Plan helps exploit Data, **learn patterns and understand its underlying structure** using supervised and unsupervised models in order to make predictions (**Machine Learning**).

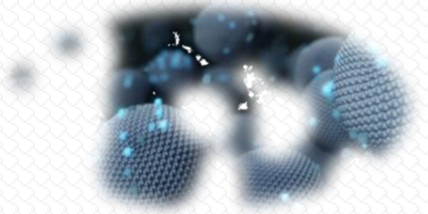
3. Risk Assessment (RA)

Risks Assessment deals with the **identification, evaluation and management of possible hazards** that may arise during the lifetime of a project.

4. Safety Recommendations (SR)

Safe-by-Design (SbD) methods enhance the **safety of final products**, emphasizing on the decision-making process regarding their design.

On the Machine Learning



Artificial Intelligence:

- Programs capable of sensing, reasoning, acting, adapting

Machine Learning:

- subsection of artificial intelligence
- computational methods using prior knowledge and data in order to improve performance or make accurate predictions.
- Various algorithms are used to go through Data and learn patterns, rules or underlying structures from it.
- These algorithms are then used in order to make predictions or determinations

Deep Learning:

- subsection of Machine Learning
- Models are called Neural networks, which are a set of algorithms, modeled loosely after the human brain, that are designed to recognize patterns
- Huge growth and research since 2005

Machine Learning is the science of getting computers to learn and improve their learning over time in autonomous fashion, by feeding them data and information

…using algorithms to parse data, learn from it, and then make a determination or prediction about something in the world

InstanceID	Feature1	Feature2	Feature3	Label
1	1.444	3.323	9.2323	25
2	3.1212	5.232	8.2342	29
3	2.32	5.322	9.111	30
4	2.099	4.99	9.022	28

Nothing rather than statistical methods…

Learn from features what the label is…

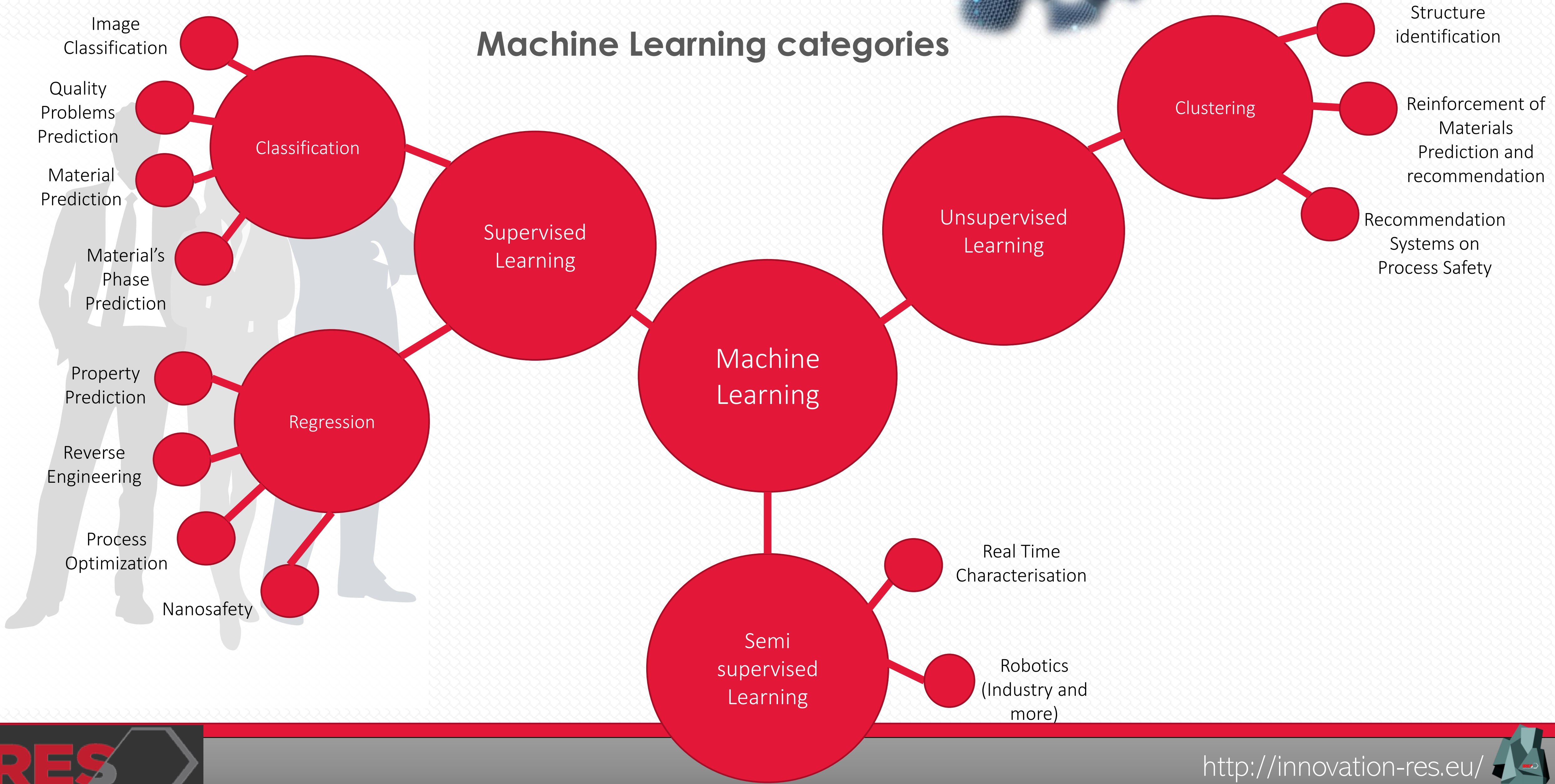
… in order to predict the label of a new unlabeled instance:

InstanceID	Feature1	Feature2	Feature3	Label
5	1.555	4.9922	8.001	?

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT press.

On the Machine Learning

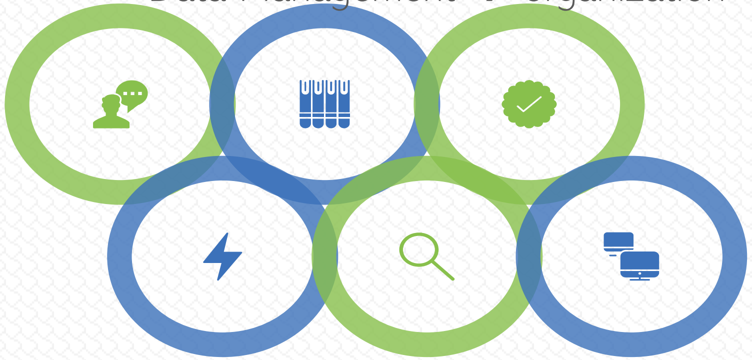
Machine Learning categories



The prerequisites-Data Management Plan

Data management is an administrative process that includes acquiring, validating, storing, protecting, and processing, required data to ensure the accessibility, reliability, and timeliness of the data for its users. A good DM Plan helps exploit Data, learn patterns and understand its underlying structure using supervised and unsupervised models in order to make predictions **(Machine Learning)**.

Data Management → organization



- ✓ A powerful data management system also provides the structure for information that can easily be shared with others and stored for future reference and easy retrieval.
- ✓ Compare the results or conclusions. Speed and ease of decision making, take action faster.
- ✓ Avoid unnecessary overlap employees conducting the same research, analysis, or work that has already been completed.
- ✓ A data management process will greatly reduce the risk of losing vital information.
- ✓ Finally, a good data management plan will simplify the implementation of Machine Learning

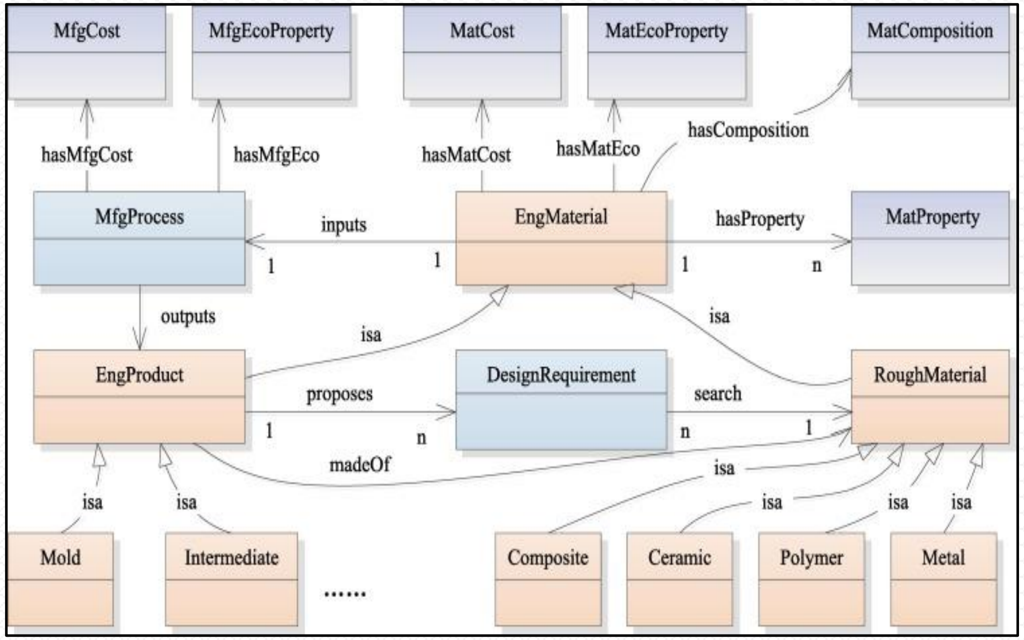
..more here: Innovative Data Management in advanced characterization: implications for materials design, N. Romanos, M. Kalogerini, E.P. Koumoulos, A.K. Morozinis, M. Sebastiani, C. Charitidis, Materials Today Communications, in press (2019)

..template here: Data Management Plan template for H2020 projects, Elias P. Koumoulos; Marco Sebastiani; Nikolaos Romanos; Maritini Kalogerini; Costas Charitidis, Zenodo 2019 https://doi.org/10.5281/zenodo.2635768

Available for download: DMP template for H2020 projects.pdf

The prerequisites-Ontology

- Knowledge representation of a specific domain separated in classes populated by variables
- Connections between the different classes describing their relation
- Allows for the execution of logical calculations
- Formal language, understandable by both humans and machines
- Shared Ontologies allow for the reuse of existing knowledge for the acquisition of new one



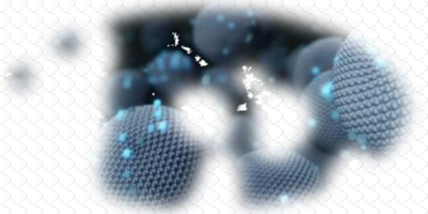
The structured storage of information in shared ontologies facilitates the Machine Learning process

..more here: Innovative Data Management in advanced characterization: implications for materials design, N. Romanos, M. Kalogerini, E.P. Koumoulos, A.K. Morozinis, M. Sebastiani, C. Charitidis, Materials Today Communications, in press (2019)

ML in characterisation

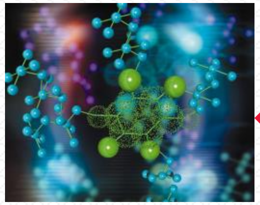
Fields and applications: ..only a few here

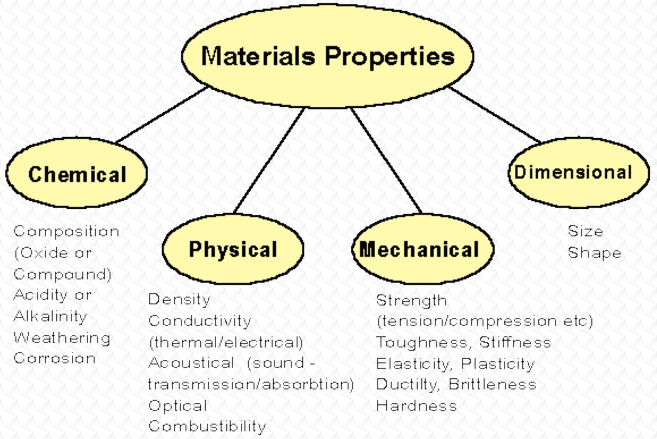
- Characterization
  - Material's property prediction
  - Structure prediction
  - Automated determination of phase diagrams using high-throughput combinatorial experiments
  - Image Classification (SEM, Optical Microscope)
  - Reinforcement type prediction
- 3d printing
  - Prediction of quality problems in 3d printed materials
  - Advanced and smart materials modelling and discovery
- Hazard and Nanosafety
  - Optimization of safety uptakes,measures
  - Recommendation systems on Process' Safety and Economical Impact
- Synthesis of Nanomaterials
  - Optimization of synthesis parameters for reverse engineering
  - Advanced and smart materials modelling and discovery
  - Text Mining from scientific papers



Trained Machine Learning Algorithm

Material Structure





Predict material properties of a known structure → Regression Model → Reduced Characterization Costs

Predict the structure with desired material properties ← Reverse Engineering ← Reduce the number of the required experiments

Materials Informatics Platforms

i-Mat project

https://imat-asp-project.com

Step 1: Collection of industrial and academia research papers

Step 2: Extract data on chemical compositions with Text Mining and Natural Language Processing

Step 3: Train Machine Learning Models with the extracted Data

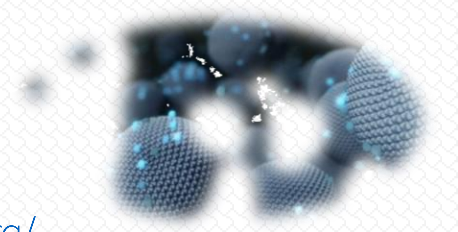
Step 4: Identify the best chemical compositions for a given application

- The i-Mat project tries to connect Industry and Academia, via giving process optimization solutions to specific industrial applications.
- Natural Language Processing and Machine Learning are combined to learn on chemical composition Data the several process parameters
- The trained models are then used to optimize one's process and recommend optimized parameters for a desired output property or composition.

Approach reduces:

- ✓ time requirements
- ✓ computational efforts
- ✓ risks of failure
- ✓ requested investment

Materials Informatics Platforms



Intellegens https://www.intellegens.co.uk/

- → **Brand Name:** Intellegens' Alchemite™
- → **Example Application:** artificial intelligence engine used **to design new alloy** for 3D printing project.
- → **Innovation: Elimination of need for expensive experiments** and **saving of millions** in the identification of new materials

Basic principle:

- Develop a framework which can train and predict models from incomplete data.
- The technology can be used to link large, easy to acquire, databases with small, hard to acquire datasets.
- Generated models can be used to design, predict and identify errors.

Proven applications with the following type of problems.

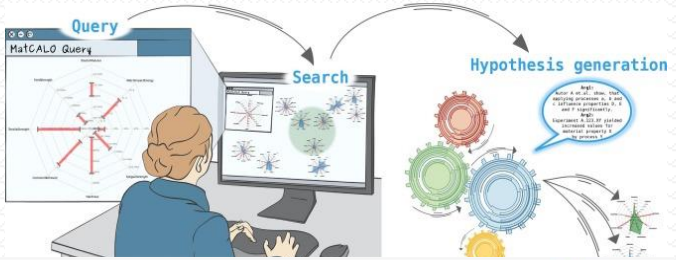
•**estimation of values** previously only accessible by expensive, empirical, experimentation
•ability to **estimate the endpoints** in complex, multistage, multi-ingredient **processes**
•**qualification of estimates** by robust and meaningful quality metrics indicative of uncertainty
•ability to **identify and correct outlier data** and to **suggest empirical experiments** that will improve overall uncertainty of the model
•**computationally efficient and scalable** from small matrices to **big data**
•**large amounts of incomplete anonymised, numerical data**
•numerical **data combined** with models or graph functions

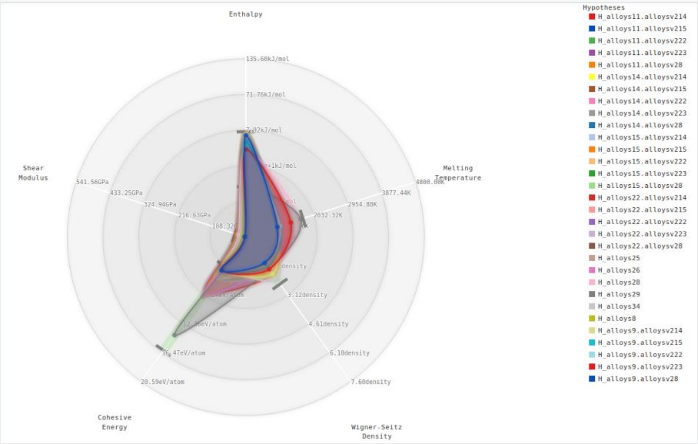
- ✓ Powerful tool for handling missing values, identifying correlations and outliers with minimum cost
- ✓ Robust and scalable AI tool for Data Preprocessing
- ✓ Reducing cost and risk by accurately modelling expensive experimental datapoints from minimal data

MatCALO http://matcalo.open-ease.org/

→ an intelligent, cognitive assistant system that supports material scientists in developing novel materials.

→ MatCALO combines modern machine learning techniques with machine-interpretable semantic knowledge in order to model representations of relationships between materials, processes and properties and allow reasoning about them.





- The system maintains a large database of experimental data which will be used to generate a set of candidate answers.
- In a later state the system will additionally use semantic knowledge to complement the trained model in refining its results, presenting alternative answers and generating final hypotheses.
- This implies the requirement of finding a suitable and machine-understandable representation of semantic knowledge.

→ The user can then visually compare the candidates with respect to their proximity to the requirement profile entered

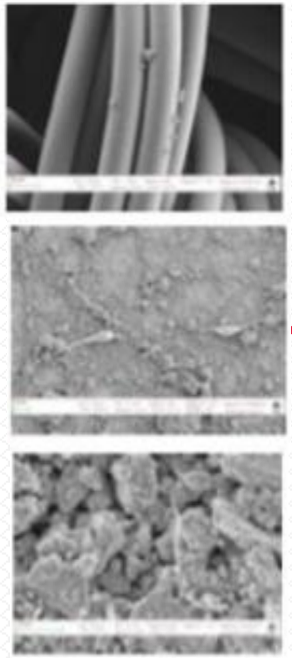
→ Many more functionalities: Data Analysis, Machine Learning, suggested process chains for the entered query

Materials Informatics works

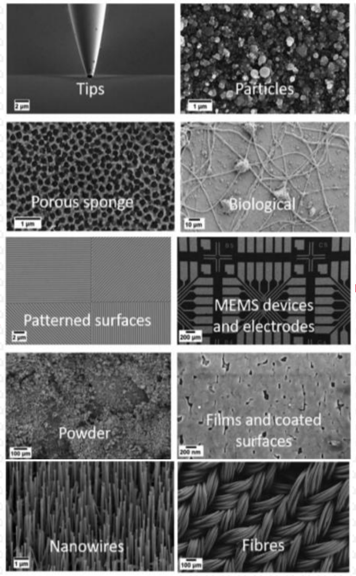
SEM Image Classification with Deep Learning*

- ✓ Part of the NFFA-EUROPE** project
- ✓ Supervised Deep Learning: Labeled Images
- ✓ Classification Scanning Electron Microscope (SEM) images at the nanoscale
- ✓ Use also of a Data Management infrastructure for storage and preservation***

Collection of >18k SEM Images

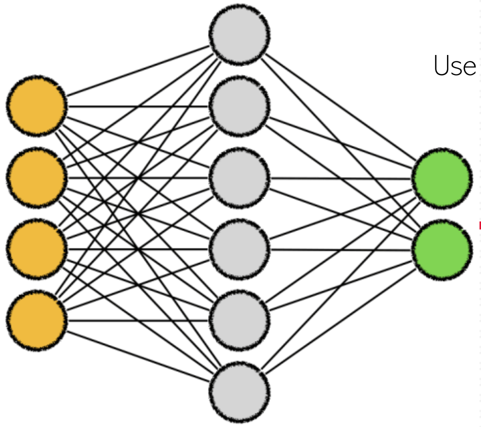


10 classes

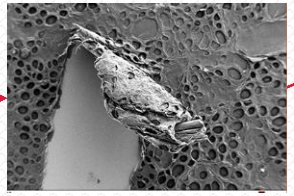


Manual Image Annotation and Segmentation

Train a convolutional Neural Network



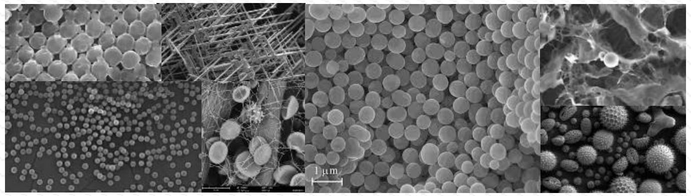
Use the Network to classify new Images



Biological

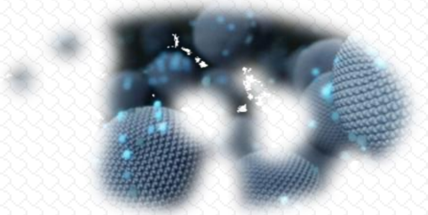
Category	N images
Porous Sponge	171
Patterned surface	3,310
Particles	3,412
Films and Coated Surface	308
Powder	895
Tips	1,561
Nanowires	3,656
Biological	953
MEMS devices and electrodes	4,158
Fibres	153
TOTAL	18,577

The Dataset used



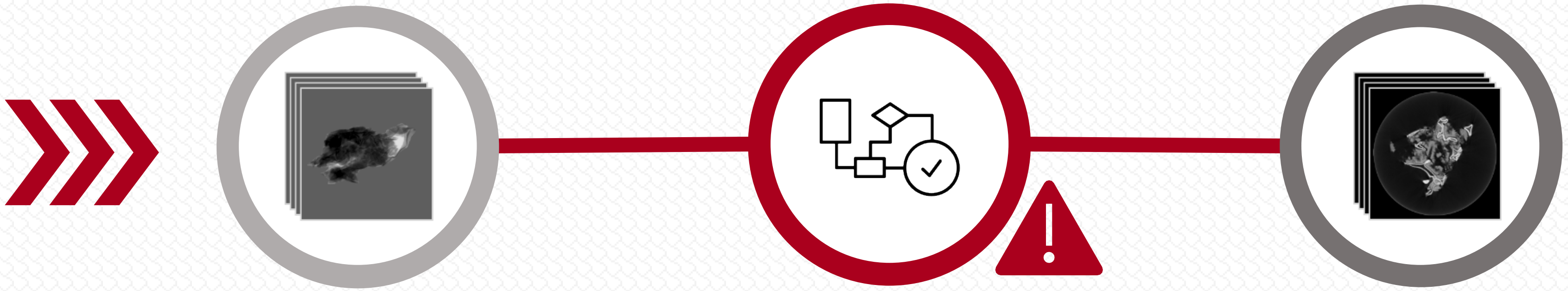
*Modarres, Mohammad Hadi, et al. "Neural network for nanoscience scanning electron microscope image recognition." *Scientific reports* 7.1 (2017): 13282.
**https://www.nffa.eu/
***https://www.nffa.eu/news/newsletters/issue-1/news/idrp/

Materials Informatics works



Deep Learning In High-Resolution XCT Image Analysis of Materials

Emre Topal[1,2], Markus Löffler[2], Ehrenfried Zschech[1,2]
[1]Fraunhofer Institute for Ceramic Technologies and Systems, [2]Technische Universität Dresden, Dresden Center for Nanoanalysis



Artificial Intelligence/Machine Learning application for HR-XCT:

- Autonomous detection of the motion from the radiographs → efficient compensation of the motion during the reconstruction
- Reduction of artifacts from missing data → high-quality reconstruction for incomplete data set

➔ High quality data

➔ Reduced time-to-data

Motion compensation

*without motion compensation

*with motion compensation

Missing data recovery

*with acquired projections from angular coverage of 180°

*with corrected projections from angular coverage of 180°

*with acquired projections from angular coverage of 360°

E. Topal, M. Löffler, E. Zschech, Radiology: Artificial Intelligence (2019, under review)
Application to energy storage/conversion materials: E. Topal et al, BMC Materials (submitted)

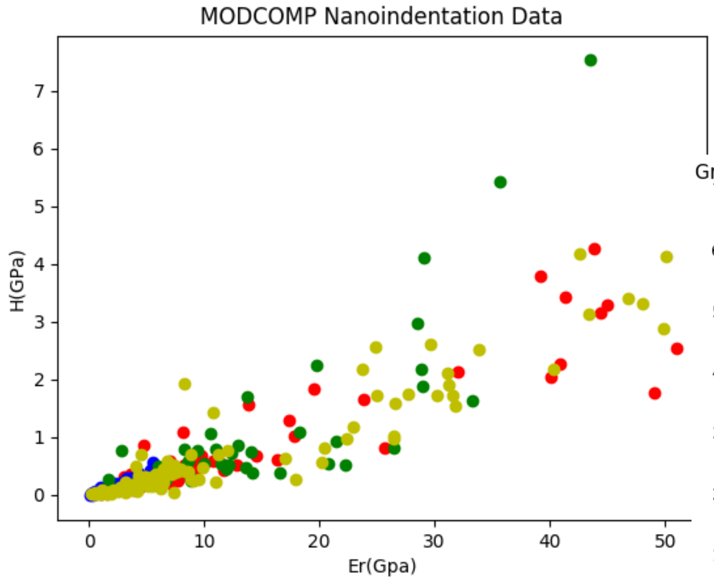
TECHNISCHE UNIVERSITÄT DRESDEN

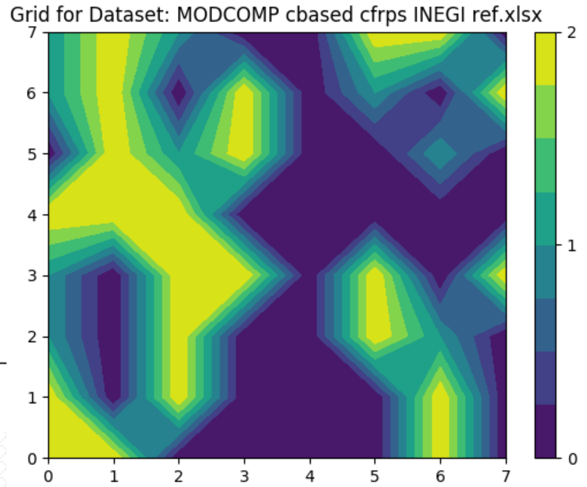
DRESDEN concept

Materials Informatics works

Combination of unsupervised and supervised Machine Learning for material reinforcement type prediction

- Nanoindentation Data hide patterns that may only be revealed with advanced Mining techniques
- Statistical Analysis can help identify hidden groupings of Data, which can be used as evaluation for later Machine Learning purposes
- Machine Learning offers opportunities from Data exploitation, such as material phase prediction and reinforcement type classification (supervised Learning for classification), property prediction (supervised Learning for regression) and re-construction of the phase topology (unsupervised Learning)
- Main Idea:
  - ✓ Find clusters of Data (in a pure statistical meaning)
  - ✓ Evaluate and correlate the clusters with the different material phases
  - ✓ Learn on Data with known label (**reinforcement type or property value**) and train a model
  - ✓ Use the model to predict/classify on previously unseen Data



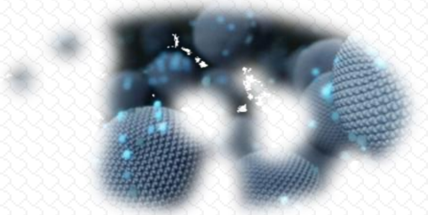


Useful for:

- Fast characterization evaluation
- Experiment and cost reduction
- Reverse Engineering and Materials Design

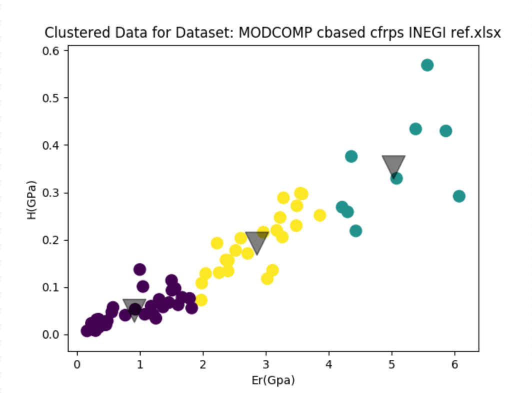
E. P. Koumoulos, C.A.Charitidis unpublished work (2019)

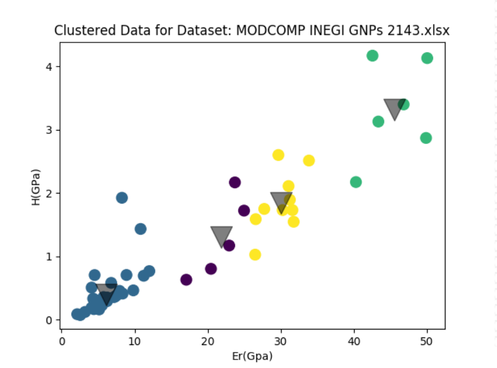
Materials Informatics works

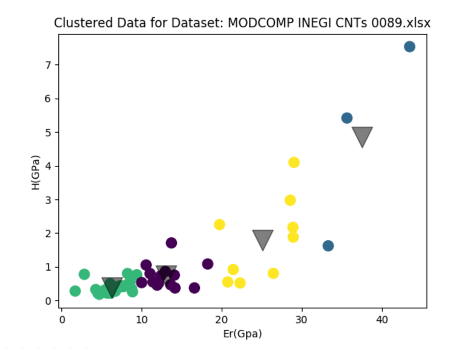


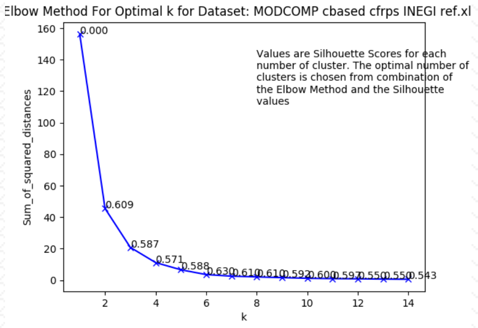
Combination of unsupervised and supervised Machine Learning for material reinforcement type prediction

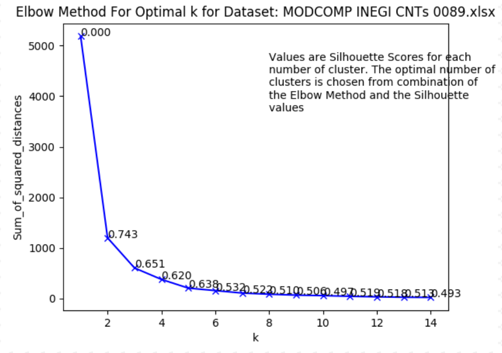
Use unsupervised Machine Learning (Clustering-Kmeans) and Statistical Analysis for creating groupings of Data (clusters) which represent (?) material phases

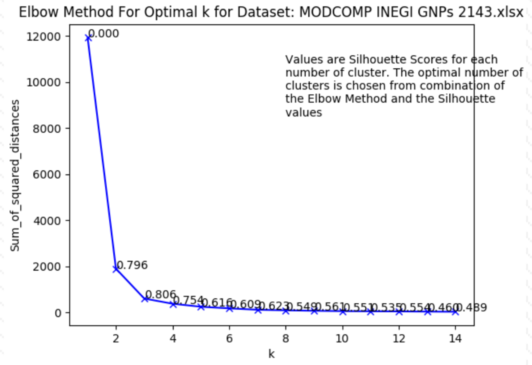












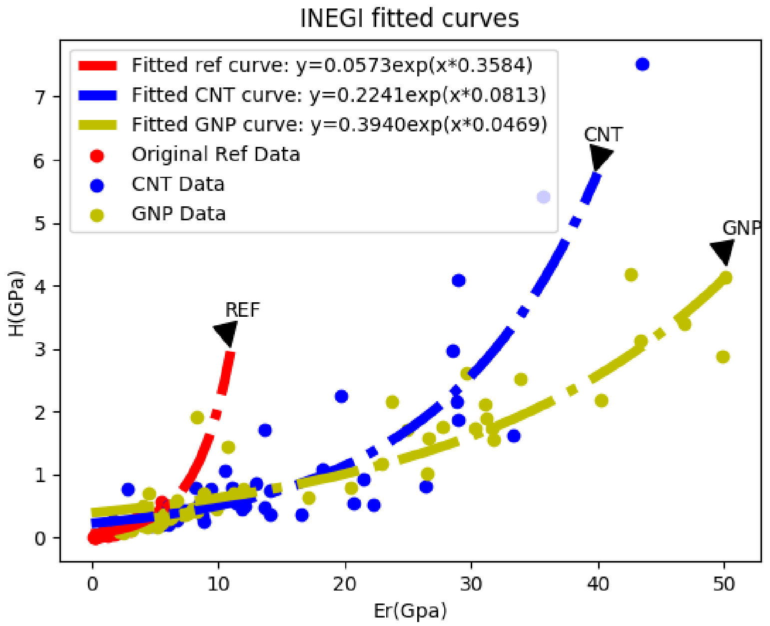
E. P. Koumoulos, C.A.Charitidis unpublished work (2019)

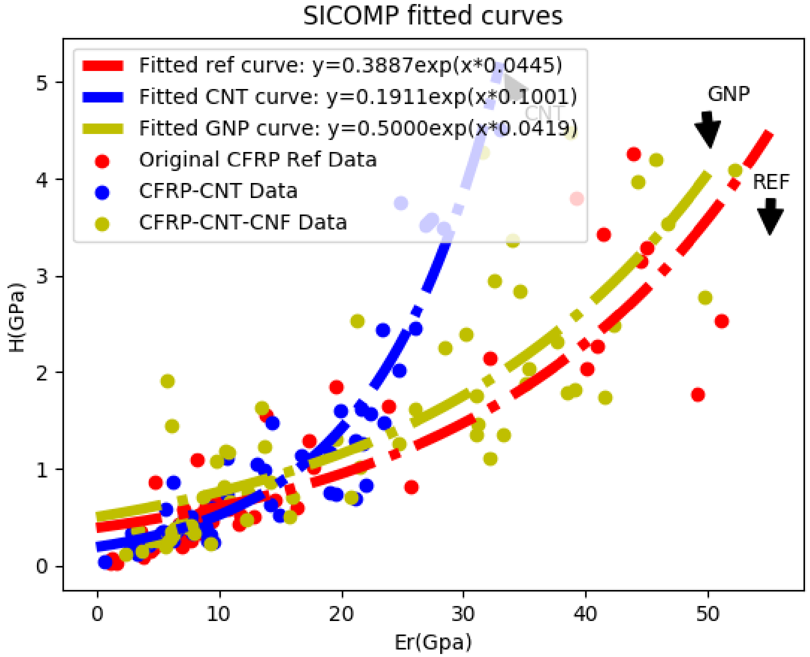
Objective: **Understand how Data is distributed across material phases**

Materials Informatics works

Combination of unsupervised and supervised Machine Learning for material reinforcement type prediction

- Fit log curves and find a good approximation of y=α*exp(β*x)
- Try to find correlation of reinforcement type (CNT,GNP etc.) with β/α ratio or similar
- Use the correlation in order to develop Machine Learning Models which will predict the type of reinforcement from Nanoindentation Data in unknown materials
- Use also the findings from previous work (unsupervised ML) in order to evaluate the displacement of each phase resulted from the reinforcement
- Train on lots of Data with known label (CNT, GNP etc.)
- Predict on Data with unknown label

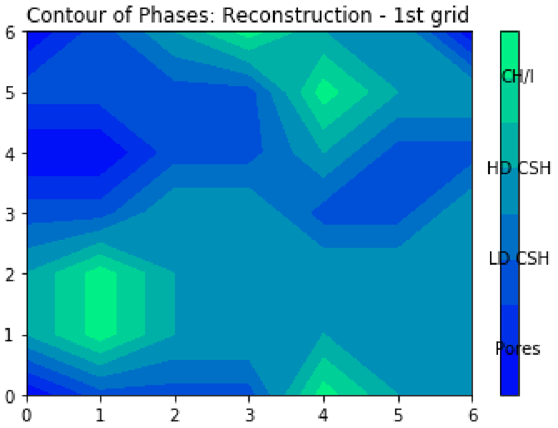


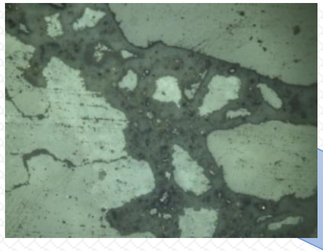


E. P. Koumoulos, C.A.Charitidis unpublished work(2019)

Materials Informatics works

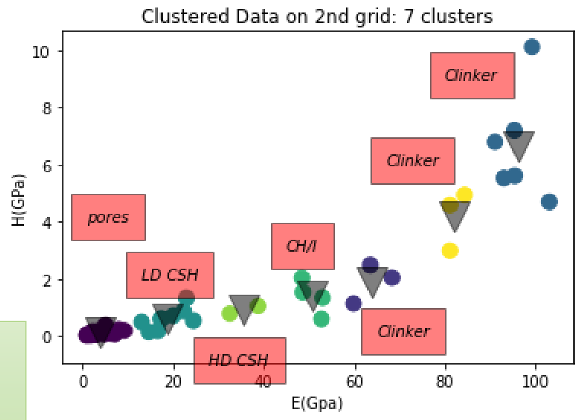
Constituents phase reconstruction through applied machine learning in nanoindentation mapping data of mortar surface

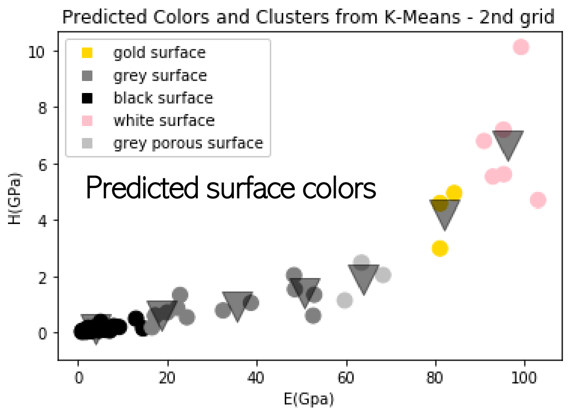




Reconstruction of constituent phases Image of mortar grid

Unsupervised (k-Means) and Supervised (Random Forests) Machine Learning

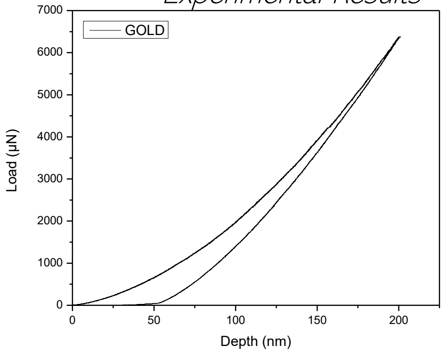


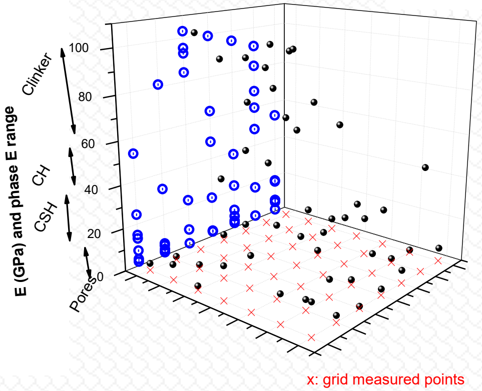


Evaluation of Experimental and Machine Learning Results

Nanoindentation in mortar grids

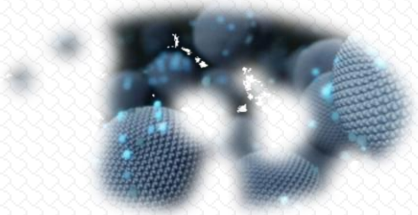
Experimental Results



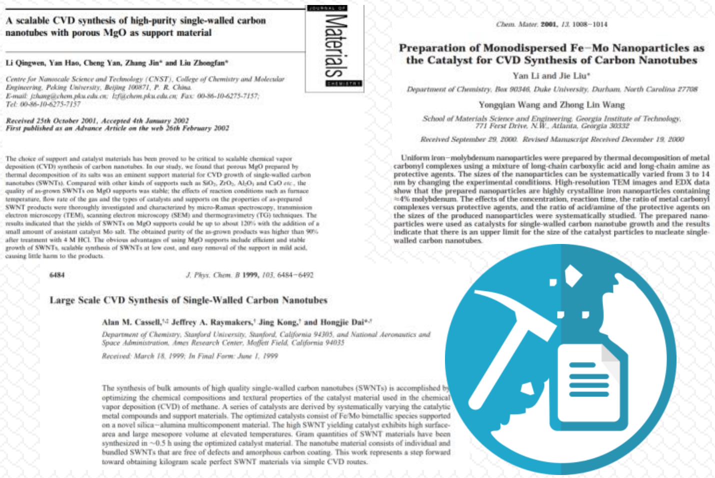


E. P. Koumoulos, K. Paraskevoudis, C.A.Charitidis, Constituents phase reconstruction through applied machine learning in nanoindentation mapping data of mortar surface, Journal of Composites Science, under review (2019)

Materials Informatics works



Text Mining from scientific papers to extract Synthesis Information and use it for Property prediction



Apply Natural Language Processing in a Database of CVD Synthesis papers and extract critical parameter values

Enrich existing Data with extracted Information

Data from CVD experiments

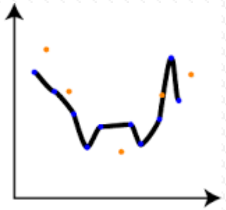
Label of interest to predict

	Precursor	Precursor Quantity	Catalyst	Catalyst Name	Catalyst Mass (g)	Catalyst Approach	Substrate	Inert Gas	Inert Gas Flow (mL/min)	Precursor Flow (mL/min)	Temperature (oC)	Reaction Time (min)	Product Form	Product Quantity (g)
1	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

Train a regression model on the obtained Dataset

Predicting a property (quantity here) from other parameters' values can help design a product from scratch in order to achieve a desired property (quantity here)

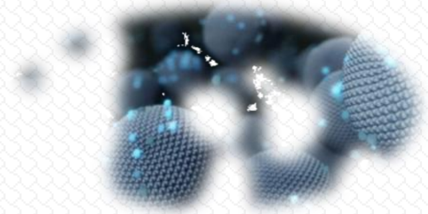
- ✓ Focus on experimental part and using regular expression find text which could possibly be experimental values (quantities, units etc.)
- ✓ Keep track of neighborhood of extracted field and also save it as a reference (the neighborhood gives information about the extracted value)
- ✓ Evaluate findings manually and accept or reject values



Predict product quantity (gr) from rest of parameters on new unlabeled Data

- ✓ *Experiments reduction*
- ✓ *Cost reduction*
- ✓ *Reverse Engineering and Smart Material Design*

Points to ~~solve~~ discuss



To consider:

- How much can machine learning Image outputs be trusted to represent data? Some ML algorithms have lead to the creation of hallucinations in the images when used to enhance image resolution.
- Really dependent on the choice and curation of the training data set
- Cases where a neural network was identifying patters in the ordering of the images rather than in the images themselves
- The “Black Box Problem”. The exact way a neural network is learning from a data set isn’t easily interpretable
- The design of neural networks requires a lot of hyperparameter tuning and therefore empirical knowledge
- Data Mining: who owns what? Trust level of data?
- Training of users to use principles and SOPs on top of Training of data analysts

Challenge:

- **Transfer learning**: train an ML algorithm to adopt to a new data set based on a similar known data set and some information regarding the relation between the two. E.g. classification of images coming from a new microscope, with different features than the one used for the original training data set.

R-NANO Research Unit of Advanced, Composite, Nano-Materials & Nanotechnology

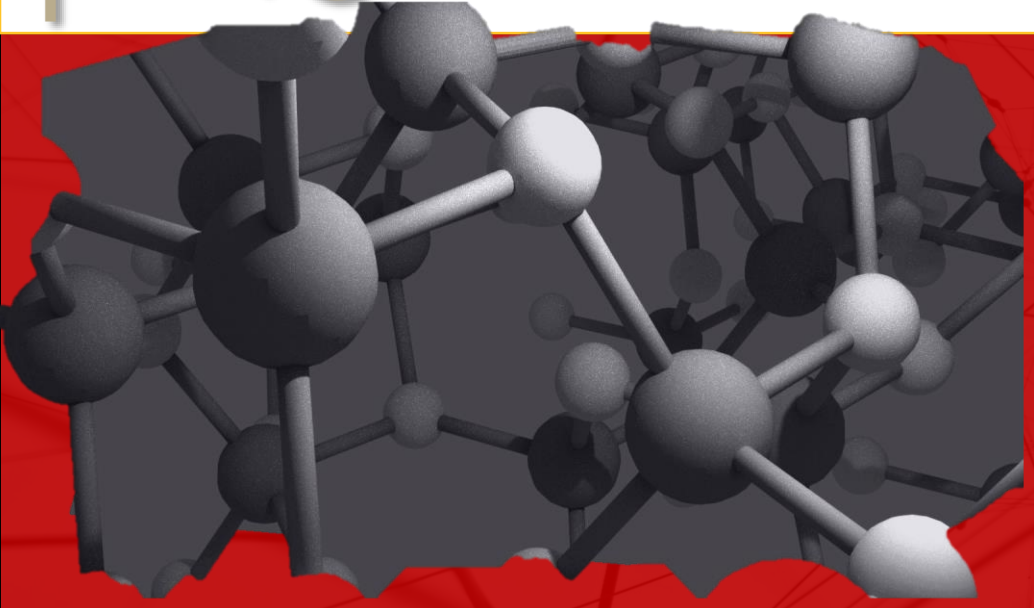
National Technical University of Athens
School of Chemical Engineering
Materials Science and Engineering Department

EuroNanoForum 2019

Applying machine learning to process and characterisation data of nanomaterials: A means for prediction

Elias P. Koumoulos
epk@innovation-res.eu

Costas A. Charitidis
charitidis@chemeng.ntua.gr



IRES Innovation in Research & Engineering Solutions