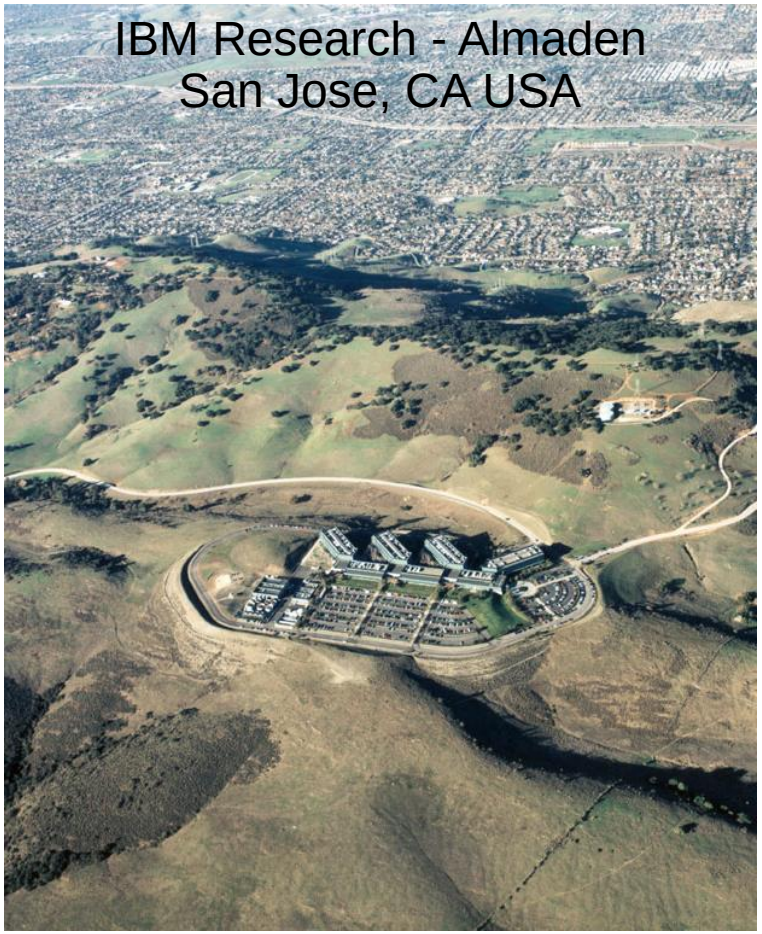
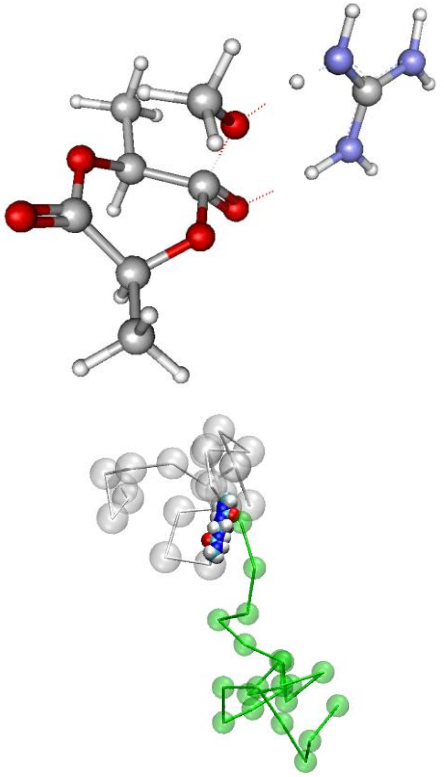
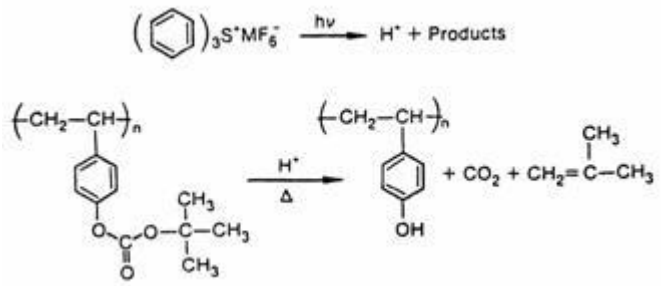
IBM Research - Almaden

Jed W. Pitera, Principal Research Staff Member, Senior Manager – Applications & Domain Science

AI and Machine Learning for Pharmaceutical & Materials Discovery







IBM's commitment to Data Responsibility

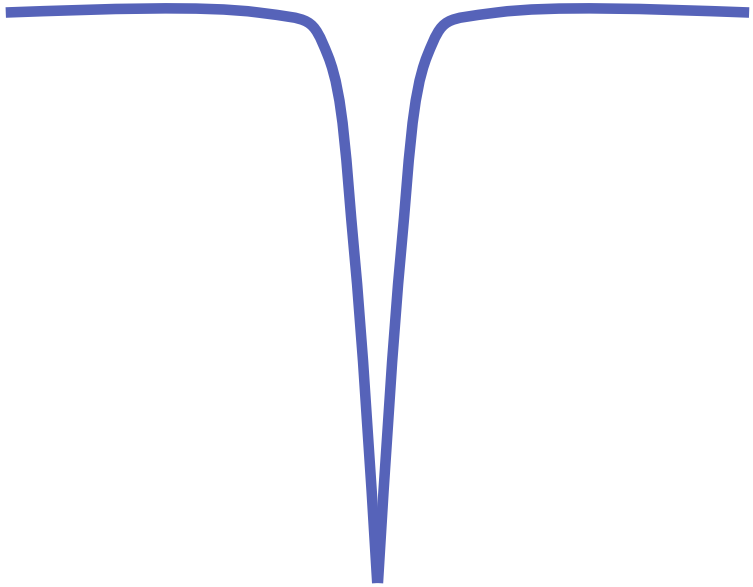
1. Data Ownership and Privacy -- clients own their data
2. Data Flows and Access – clients determine where data is stored & therefore legal jurisdiction of data
3. Data Security and Trust – full-strength encryption without backdoors; public-private cybersecurity partnerships
4. AI and Data – algorithms need to be transparent and explainable; bias must be understood & controlled
5. Data Skills and New Collar Jobs – future workforce needs to be ready to work with data and AI

https://www.weforum.org/agenda/2018/01/new-era-data-responsibility

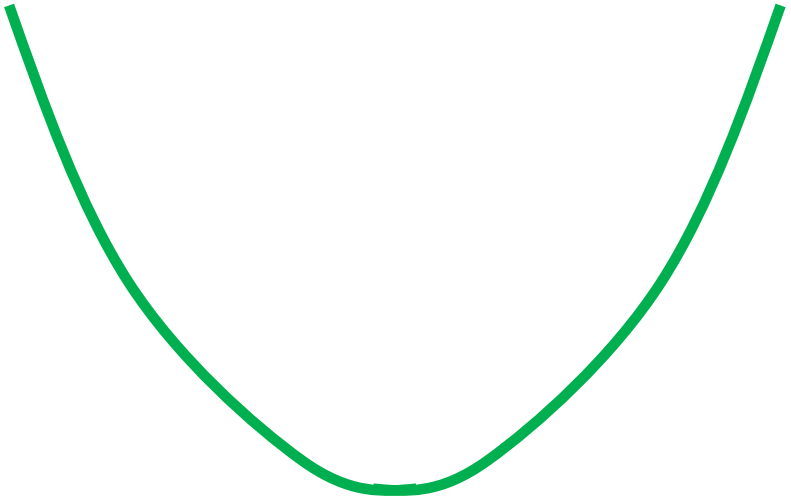
https://www.ibm.com/blogs/policy/dataresponsibility-at-ibm/

What sport are you playing?

Response surfaces/objective functions



- Small part of problem space is valuable
- No useful gradients
- Hard search problem
- Ex: drug discovery

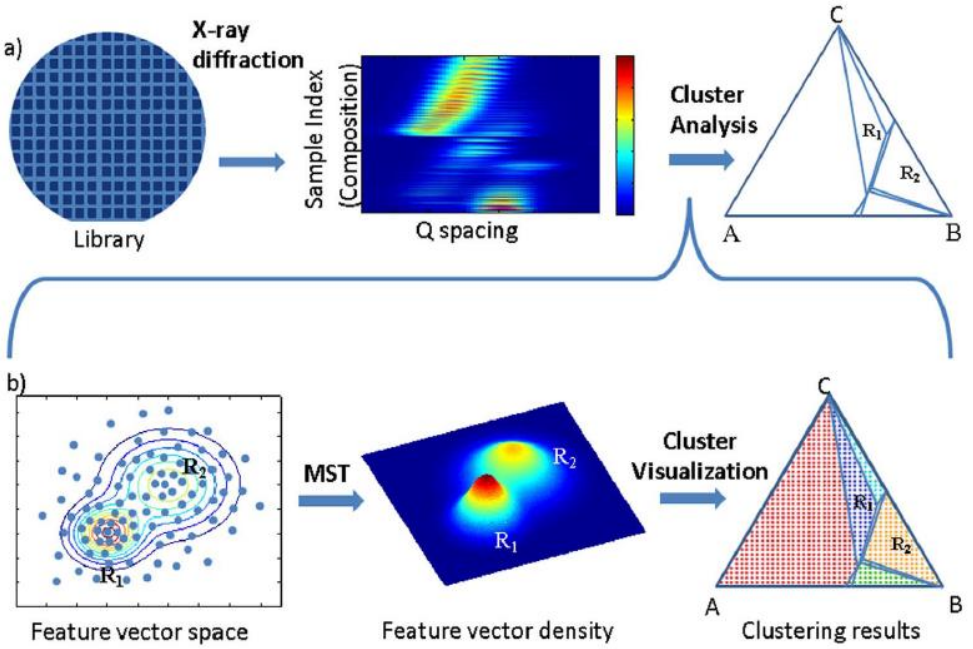


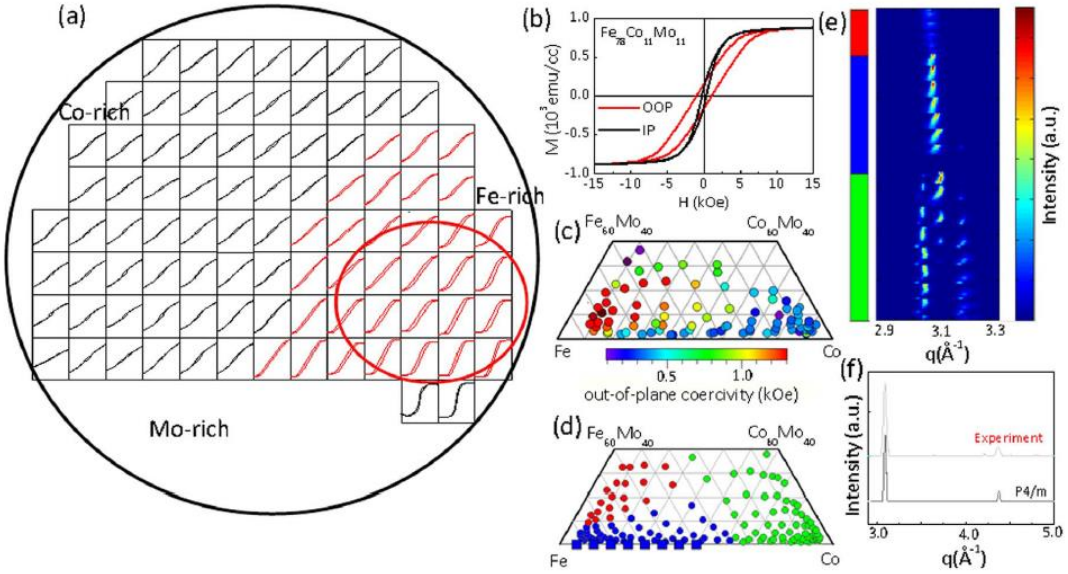
- “Close” counts
- Useful gradients
- Optimization problem
- Ex: alloy design

ML for experimental interpretation of X-ray materials data

Kusne, et al Scientific Reports 4, 6367 15 September 2014

doi: 10.1038/srep06367

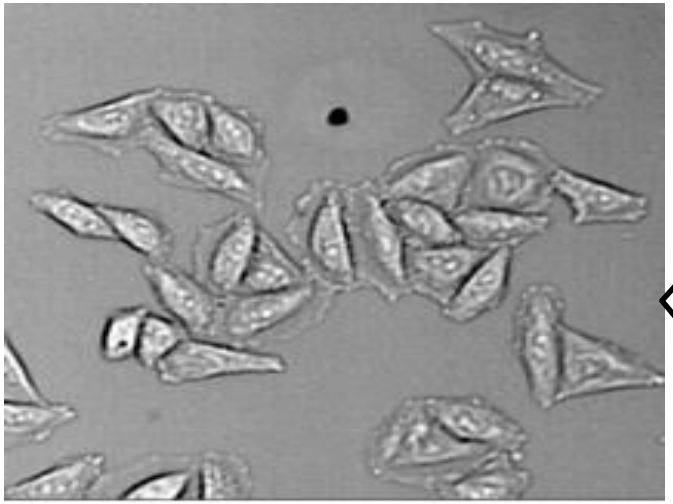




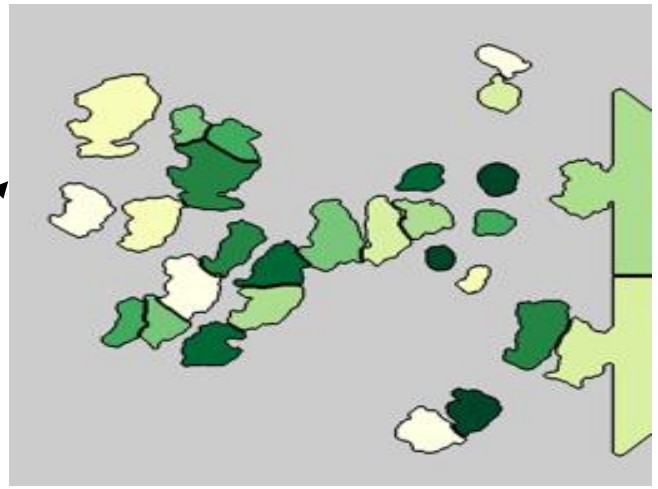
DNN for cell image segmentation with active learning

Chowdhury, Biswas, Bianco, bioRXiv preprint; doi 10.1101/211060

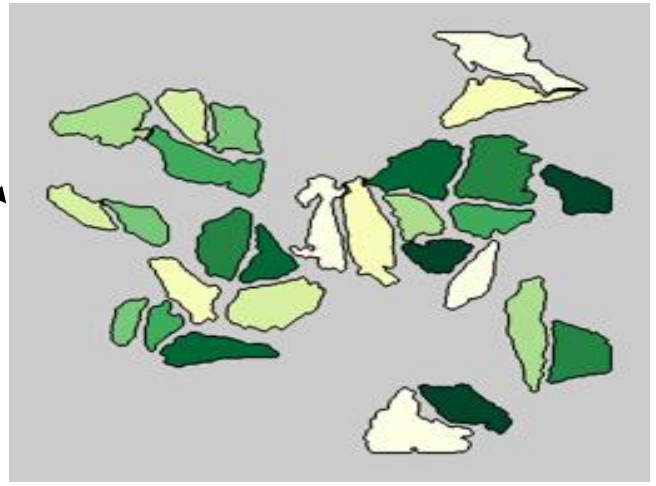
Original



Images needed for training = 1/6 of best method available



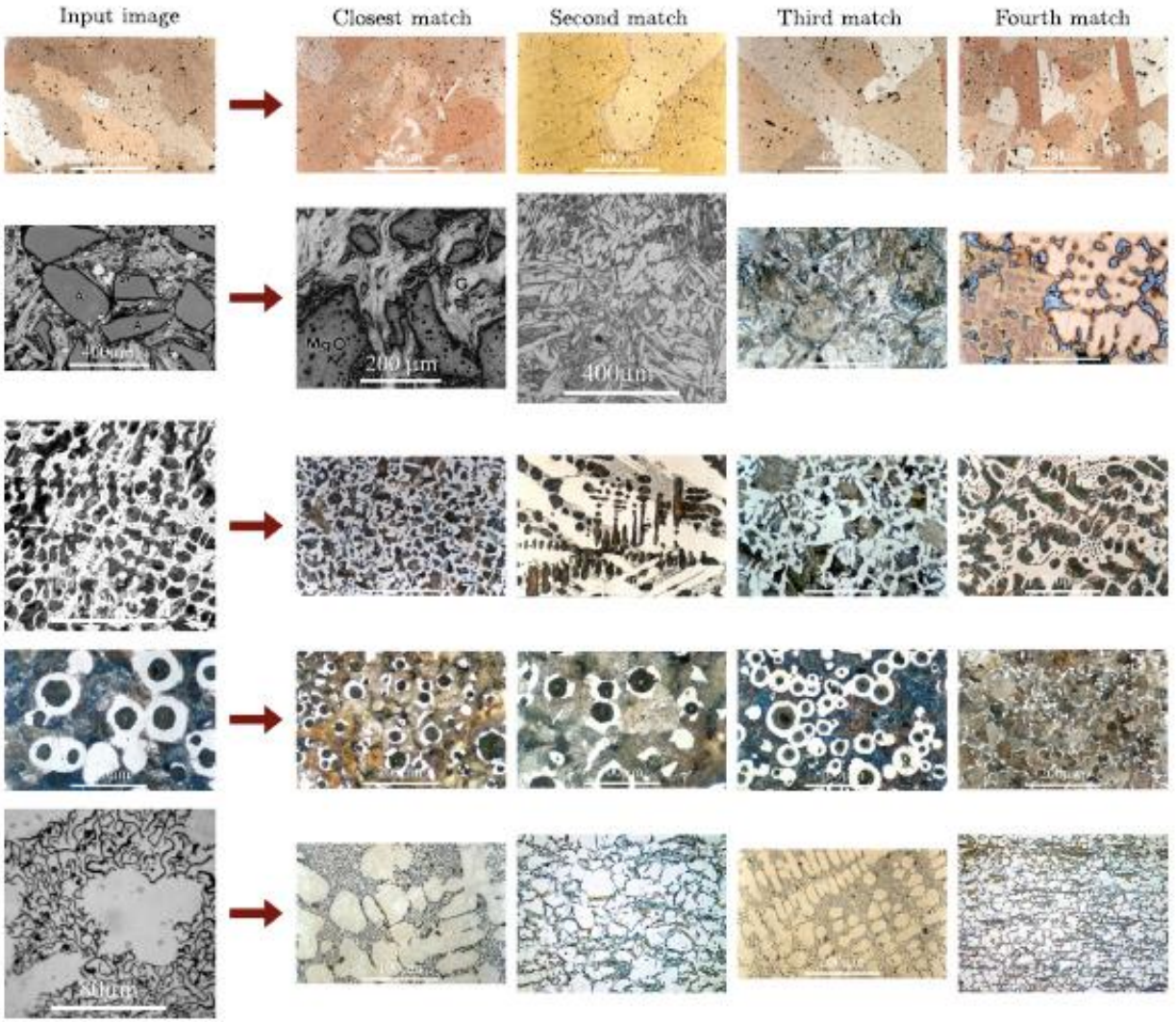
w/o Active Learning



Active Learning

Van Valen, DA, et al. *PLoS Comput Biol* 12.11 (2016): e1005177.

Computer Vision for metallurgical microstructure

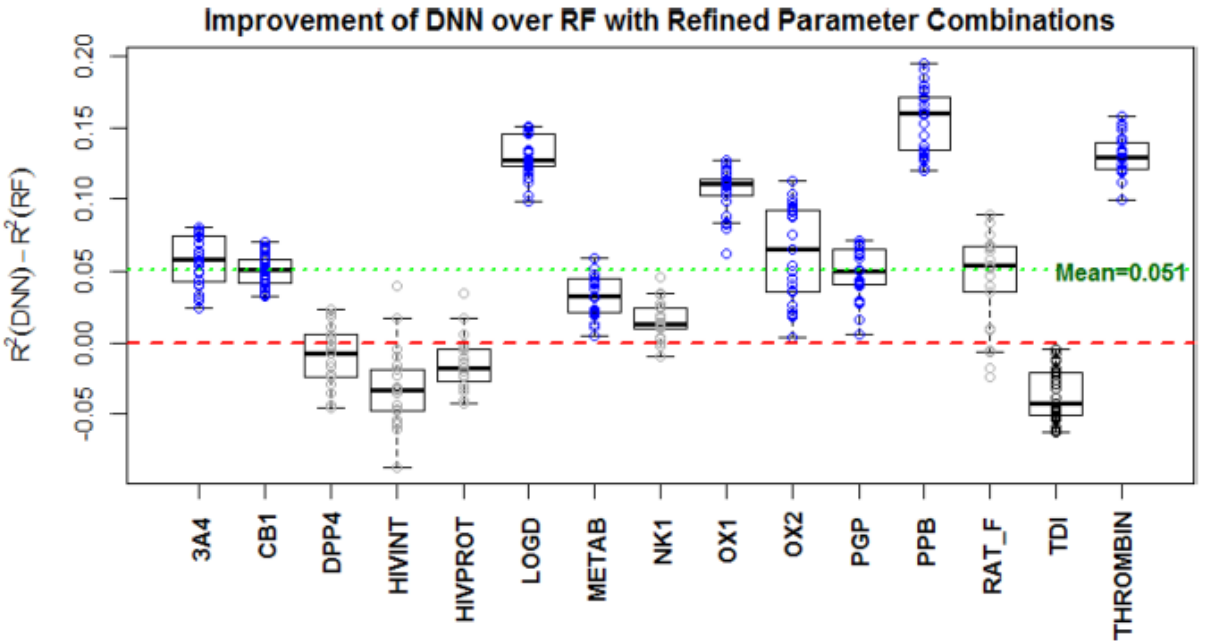


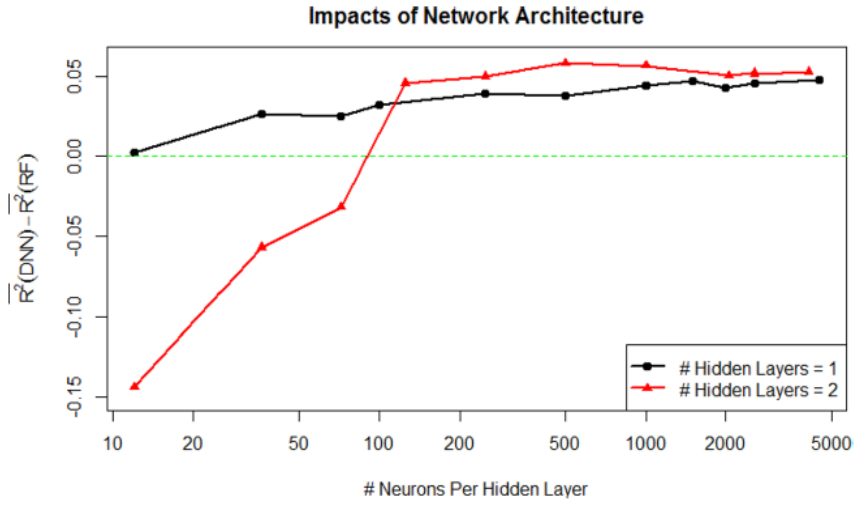
DeCost, Holm, Computational Materials Science 110 (2015) 126-133;
dx.doi.org/10.1016/j.commatsci.2015.08.011

DNN for quantitative structure-activity relationships (QSAR)

Ma, Sheridan, et al J. Chem. Inf. Model. 2015 55, 2, 263-274

doi 10.1021/ci500747n





Learning models of materials properties

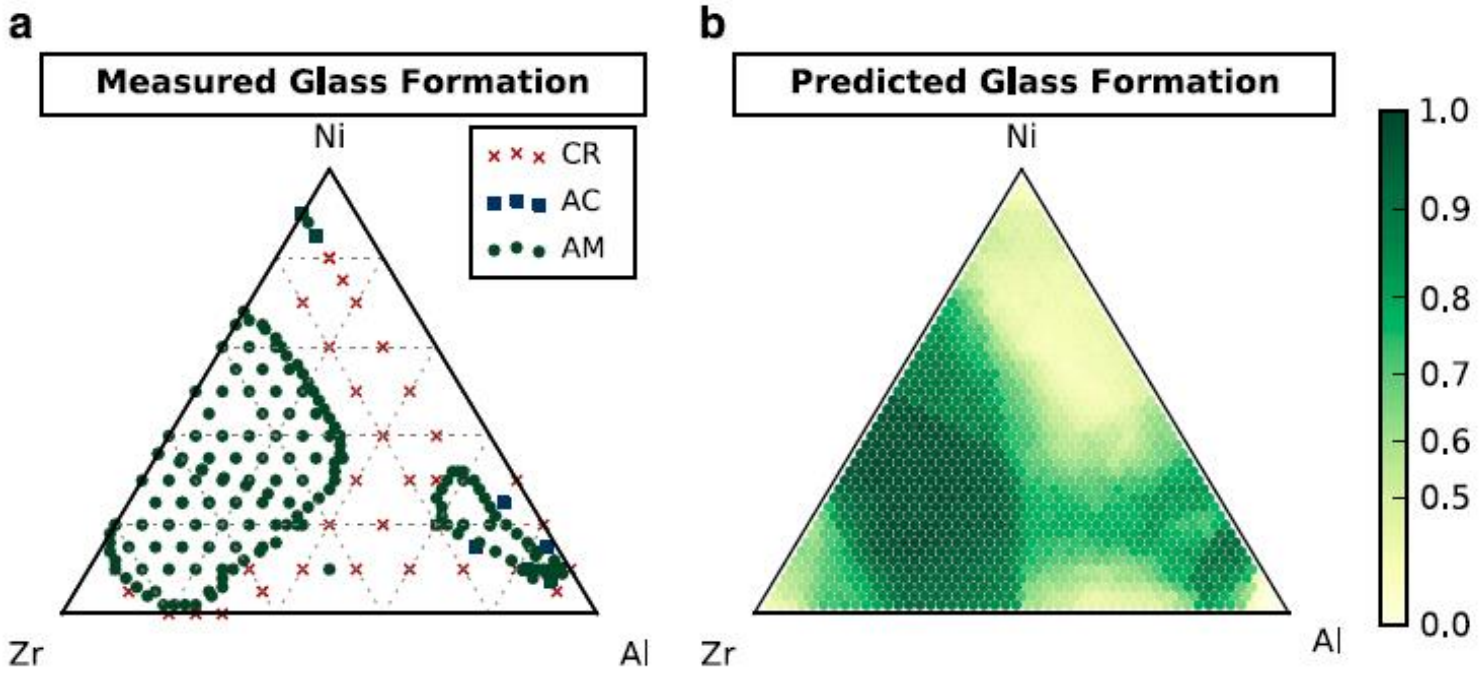
Ward, et al, npj Computational Materials 2016 2, 16028;

doi:10,1038/npjcompumats.2016.28

Table 1. Comparison of the ability of several machine learning algorithms to predict properties of materials from the OQMD

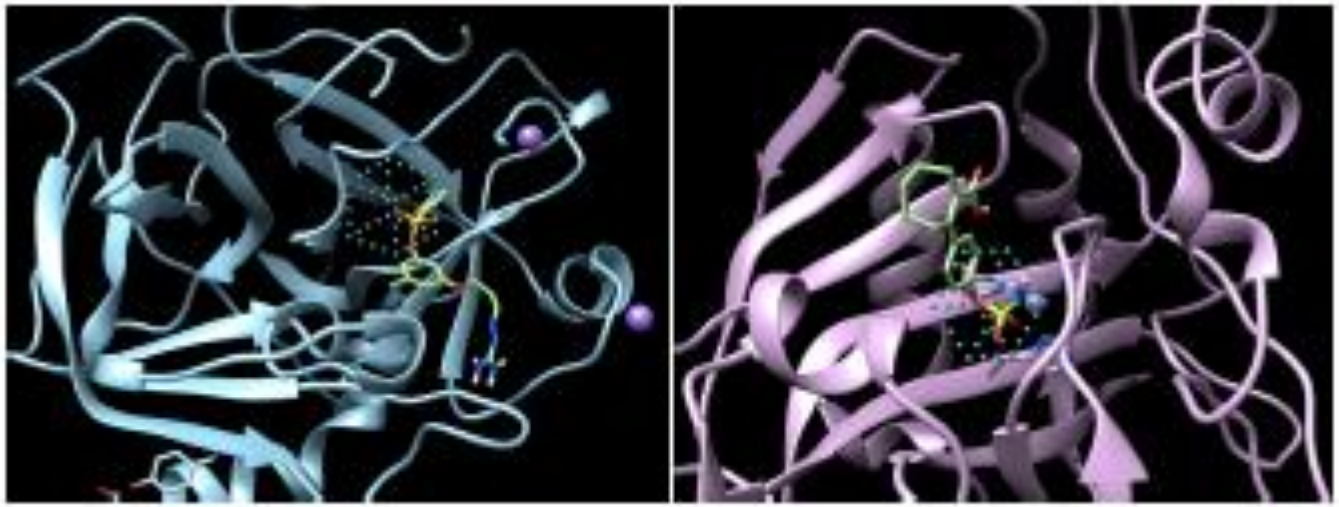
	Machine learning algorithm			
	Linear regression	*Reduced-error pruning tree (REPTree)*	*Rotation forest*[46]*+REPTree*	*Random subspace*[55]*+REPTree*
Property				
Volume (Å^3 per atom)	1.22	0.816	0.593	0.563
Formation energy (eV per atom)	0.259	0.126	0.0973	0.0882
Band gap energy (eV)	0.202	0.0701	0.0643	0.0645

Abbreviations: DFT, density functional theory; OQMD, Open Quantum Materials Database.
Data represents the mean absolute error in a 10-fold cross-validation test of a single model trained on the properties predicted using DFT of 228,676 crystalline compounds.



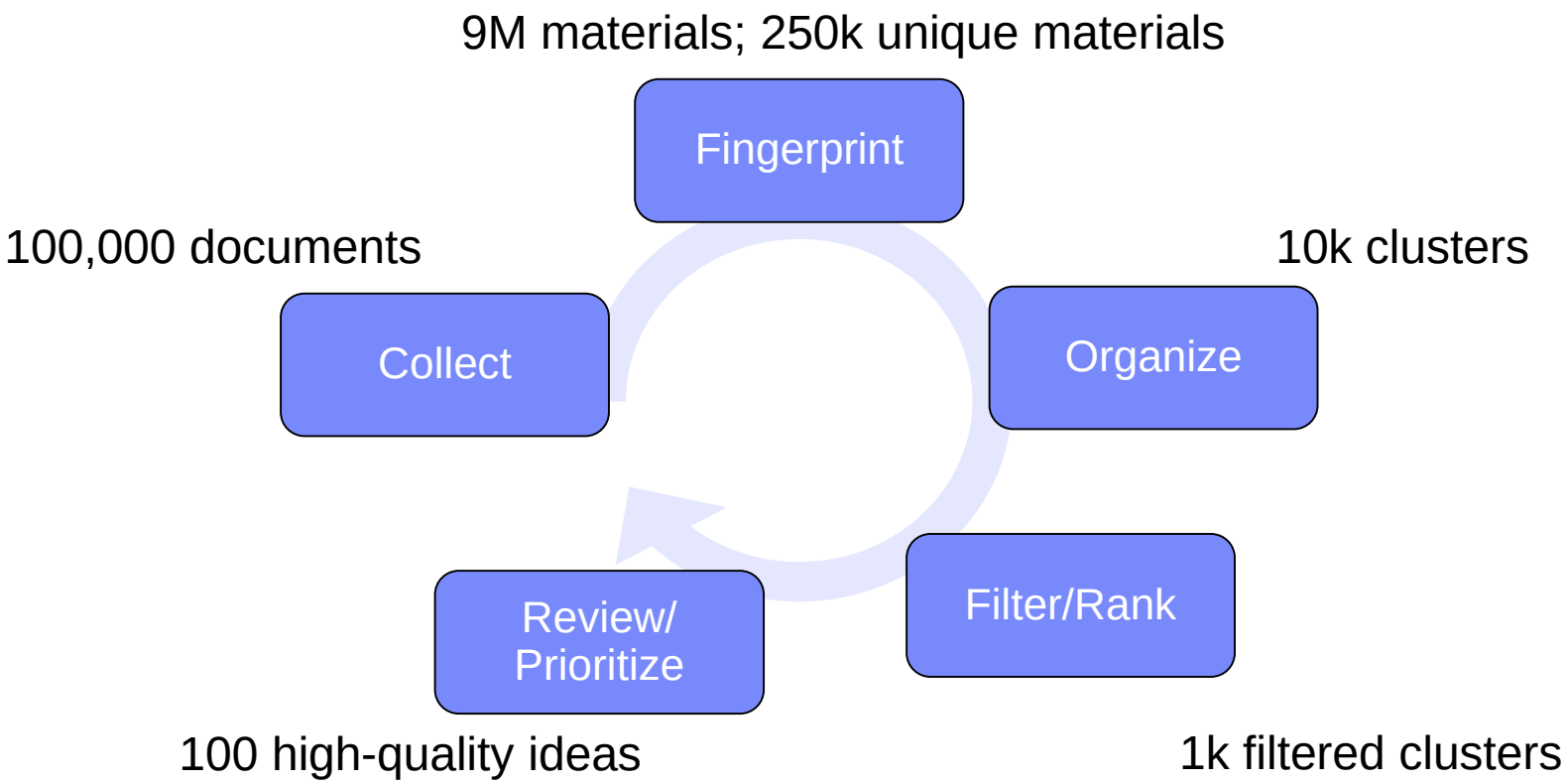
Convolutional DNNs for drug-target binding prediction

Wallach, Dzamba, Heifets, arXiv preprint arXiv:1510.02855, 2015



Comparator	# of DUDE targets tested	Comparator mean AUC	AtomNet mean AUC
Surflex-Dock	10	0.76	0.93
DOCK3.7	full set	0.70	0.895
DOCK6.7	5	0.72	0.85

Sample Informatics Cycle



AI/Machine learning in Discovery applications

Improve/speed experiment

Manage complex data

Learn predictive models

Encode human expertise