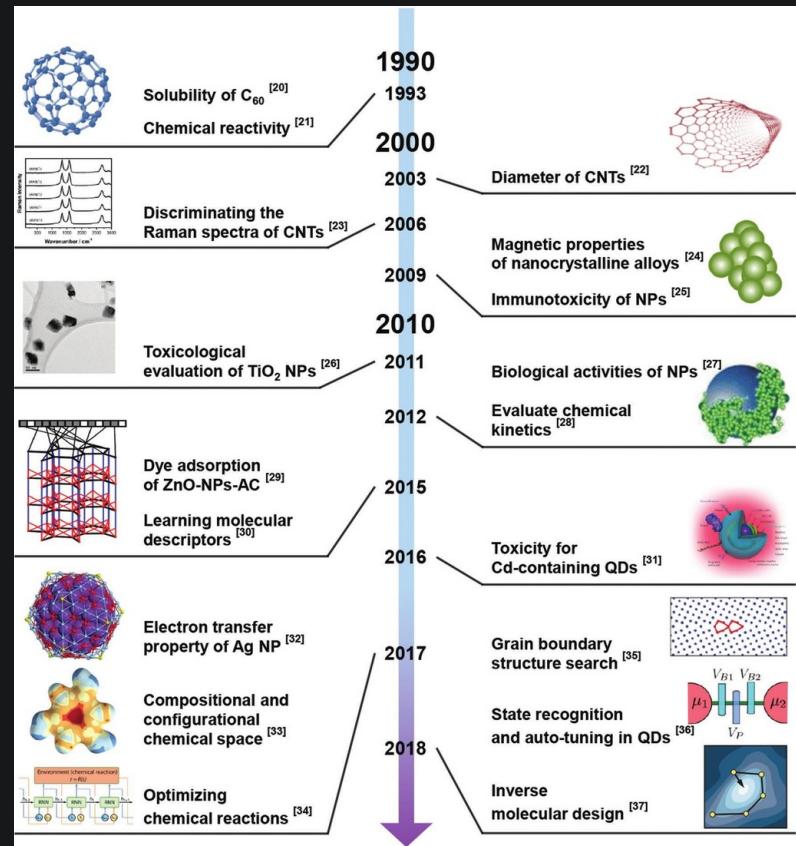


Paper presentation on "Nanomaterials Discovery and Design through Machine Learning" by Wang et al

Parth Shah
180122030

Introduction

- Machine Learning
- NN to predict which bonds will break preferentially (1993)
- Morphology estimation and Quality assessment of CNTs using artificial neural networks (2003)
- Raman spectroscopy and SVMs for toxicological evaluation of nanoparticles (2006)
- Modelling biological activities of nanoparticles (2012) (using FCNN)
- Optimizing Chemical Reactions with Deep Reinforcement Learning (2017)
- Inverse molecular design using machine learning (2018)



The general process

- feature engineering from the database to features
- model building from features to models

Feature engineering

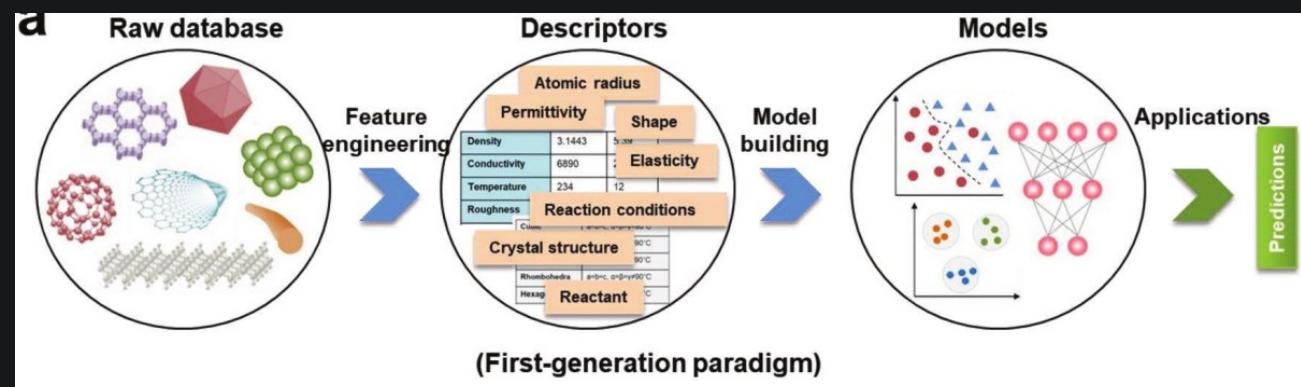
- descriptors to represent the raw nanomaterials.
- structure properties,
- mechanical properties,
- electrical properties,
- magnetic properties

Applications

- QSPR,
- QSAR,
- composition
- and configuration of chemical space

Difficulties

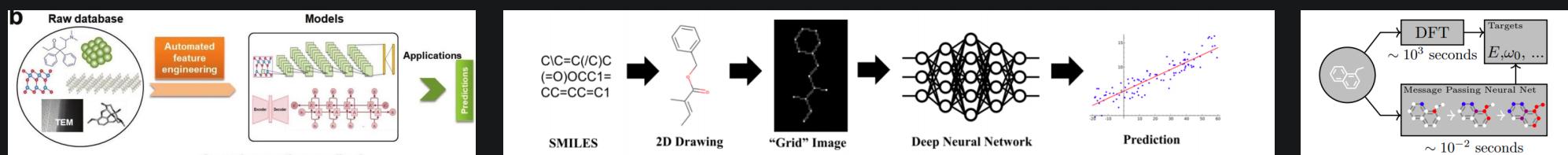
- Feature engineering
- Model building



Tasks	Specific objectives	Materials	Machine learning algorithms	References
Macroscopic performance prediction	Solubility	C ₆₀ , SF ₆	Regression	[20,39]
	Diameter, morphology	CNT, fibers	Artificial neural network (ANN)	[22,40,41]
	Magnetic properties	Nanocrystalline alloys	ANN	[24]
	Toxicity	TiO ₂ NPs, Cd-containing QDs	Decision tree, support vector machine (SVM), random forest	[25,26,31]
	Biological activities	NPs	Linear regression, ANN	[27]
	Catalytic activity, active sites	NPs	ANN, Bayesian linear regression	[49–52]
	Dye adsorption	ZnO-NPs-AC	ANN	[29]
	Solubility, drug efficacy, photovoltaic efficiency	Moleculars	Convolutional neural network (CNN)	[30,43]
Microscopic performance prediction	Synthesis recipes	TiO ₂ nanotubes	Decision tree, SVM	[45]
	Electron transfer property	Ag NPs	K-means, regression, random forest, ANN	[32]
	Chemical kinetic parameters	Chemical kinetics	ANN, RNN	[21,28,34]
	3D geometry	Pt NPs	ANN	[42]
	HOMO, LUMO, electron energy gap	Moleculars	CNN	[44]
	Grain boundary structure, defect and phase evolution	Cu, doped WS ₂	Clustering, CNN	[35,53]
	Raman spectra	Functionalized MWNTs, NPs	Principal component analysis (PCA), least squares, non negative matrix factorization (NMF)	[23,46,47]
	State recognition and auto-tuning	QDs	Deep neural network (DNN)	[36]
New nanomaterials discovery	Compositional and configuration chemical space	Molecules	DNN	[33]
	Inverse molecular design	Molecules	Generative machine learning models	[37]
	Composition and atomic ordering	Alloy nanoparticles	ANN	[53]

Automated Feature engineering

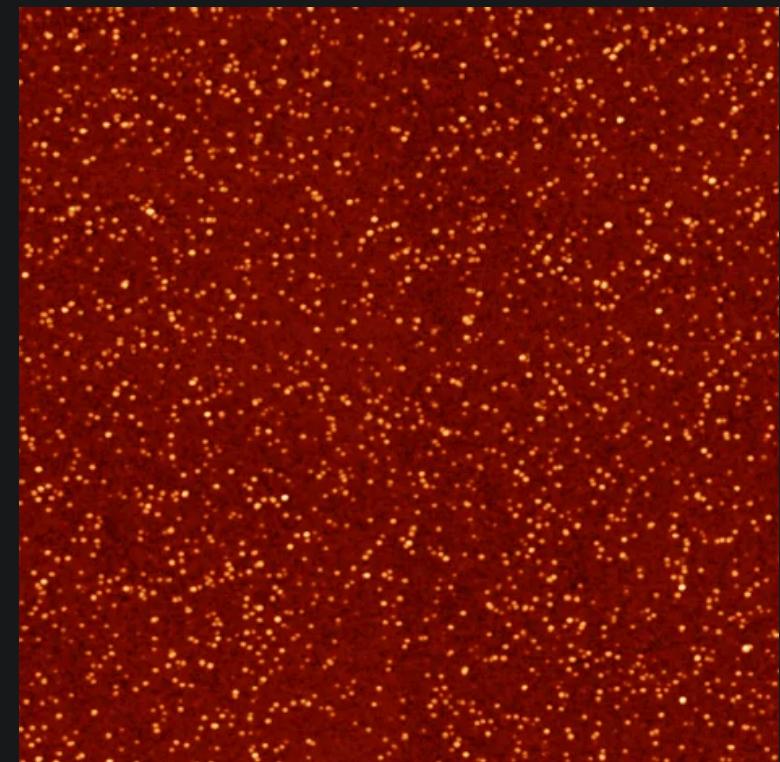
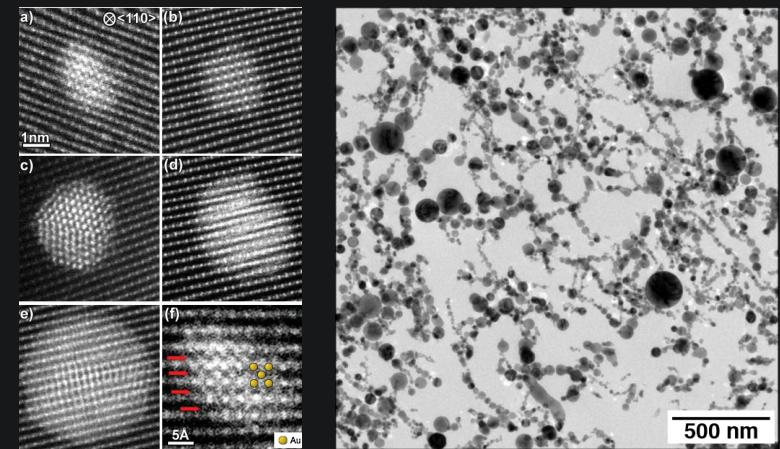
- The emergence of Deep Learning
- Capable of developing their own set of internal features
- Chemception
- Message Passing Neural Network



Learning from Multidimensional Characterization

- Earlier input were physical and chemical properties.
- Microscopic images like STEM, Cryo-EM, and AFM provide more effective inputs for nanomaterials design - spatially distributed information - morphology and structure.
- Raw molecular graphs can provide insight into the physical and chemical behaviour of materials.

ML system in nanomaterials could learn from much richer data modalities to enhance the prediction performance of models.



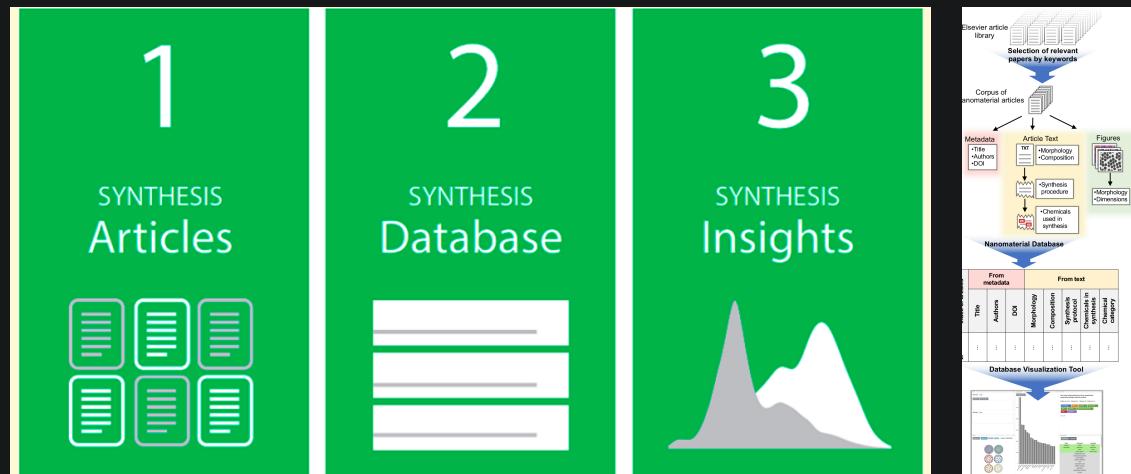
End-to-End research methodology

- Scientists extract information through going through scientific literature.
- The recent proliferation of journals and articles.

Recent advances:

- Text Mining
- Information retrieval

Kim et al: Synthesis conditions => special database => ML for predicting Nanomaterial properties



Technical Challenges

- Training-application gap - Specific Database needed
- Model interpretability - knowledge learned by these models is difficult to understand
- Difficult to transfer learned knowledge from a specific task to other relevant tasks
- To better utilize nanomaterials, developing the flexible and transferrable descriptors that can represent nanomaterials well is also urgent and challenging.

Perspectives

- Production of high-quality data for machine learning through nanocombinatorics
- Develop machine learning algorithms that can train the data on a small database, especially when the amount of data is intrinsically small in nanomaterials domain
- Machine learning can enhance existing theoretical computational approaches. The traditional quantum/molecular mechanics, such as density functional theory, molecular dynamics, and Monte Carlo techniques, have been combined with machine learning for materials research.

