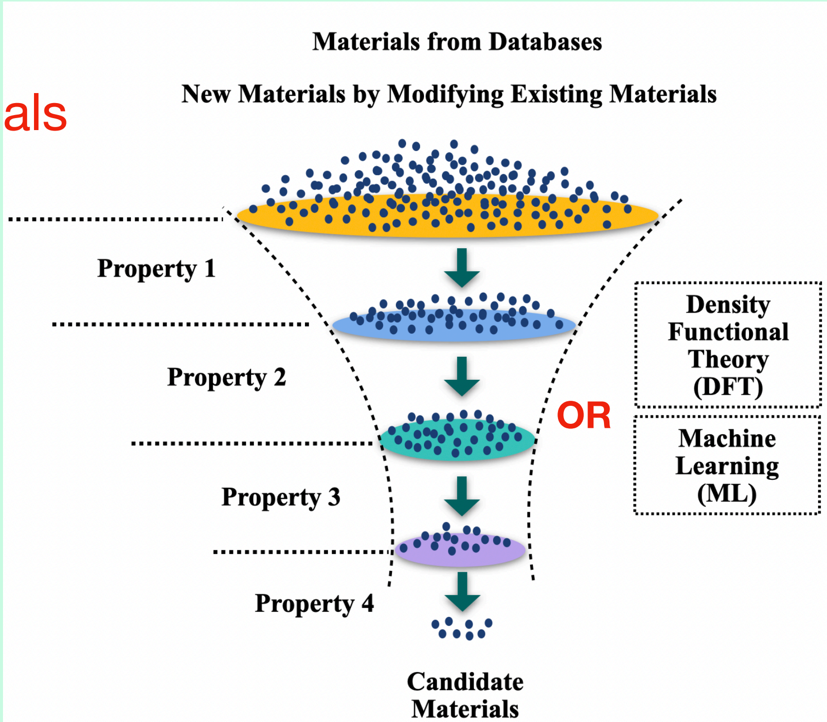
Materials screening

1. Experiments are time consuming, expensive and uncertain
    - Each sample may take ~2 days to ~2 months for synthesis
    - High-purity elements can be very expensive

2. High throughput experiments: still expensive
3. High throughput (DFT) computations
    - Expensive for hundreds of thousands of materials

4. Here comes machine learning
    - Hierarchical filtering/screening of materials



Applications of ML in materials

A few hand-picked examples

Journal of Magnetism and Magnetic Materials 589 (2024) 171590

Contents lists available at ScienceDirect

Journal of Magnetism and Magnetic Materials

journal homepage: www.elsevier.com/locate/jmmm

ELSEVIER

Research article

Leveraging available data for efficient exploration of materials space using Machine Learning: A case study for identifying rare earth-free permanent magnets

Sourav Mal, Prasenjit Sen *

Harish-Chandra Research Institute, a CI of HBNI, Prayagraj 211019, India

ARTICLE INFO

Keywords:
Machine learning
Permanent magnets
Magnetic anisotropy energy

ABSTRACT

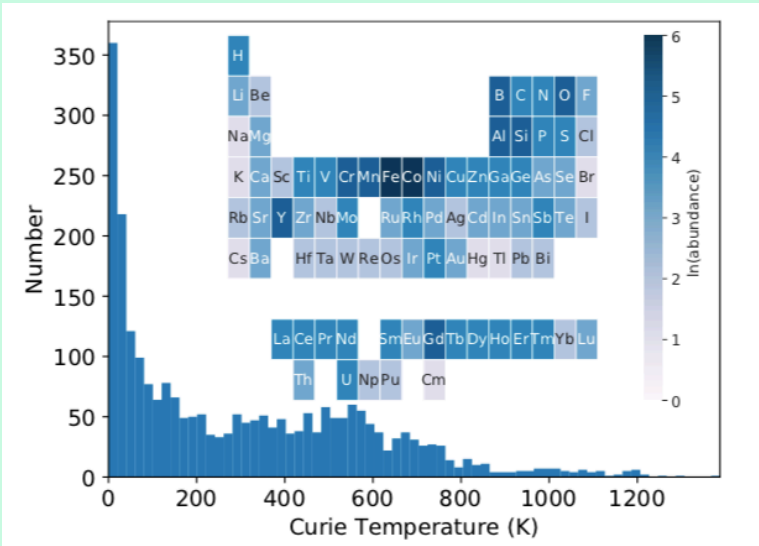
It is shown that even the relatively small amount of available materials data can be innovatively utilized to explore the materials space in order to identify materials with desired target properties. As an example of this, data from the novomag and Novamag databases are used to train random forest and neural network models which can predict thermodynamic stability, and magnetic properties of materials. Performance of these models are tested thoroughly and are found satisfactory. These models are subsequently used to interpolate within the above databases, and to extrapolate to parts of the materials composition and structure space not covered in these databases, to identify stable, magnetic materials that have large saturation magnetization and large easy-axis anisotropy. Screening 686 materials via the trained models, and subsequently performing first principles calculations, 21 new candidate materials for rare earth free permanent magnet are identified. Some of these materials have anisotropy constants as large as 5 and 6 MJ m^{-3}, larger than that of the most widely used permanent magnet $Nd_2Fe_{14}B$. This simple approach can be used to screen materials with other functionalities in future.

Example problem

- Computer and related devices
- Household appliances
- Industrial/large scale equipments
- Strategic applications

- Most use **rare earth** elements Nd, Sm
  - Supply issues, geopolitical sensitivity
  - Most widely used $Nd_2Fe_{14}B$

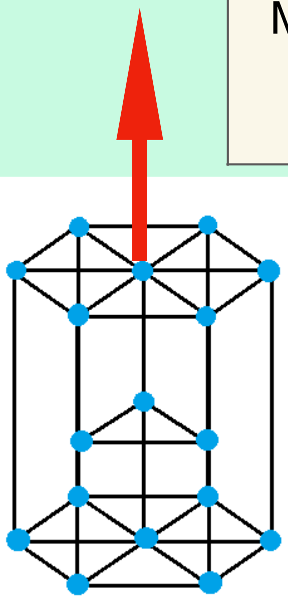
- Not many to use at room temperature
  - Need large magnetization, anisotropy
  - Very few FM magnets with $T_c > 600$ K



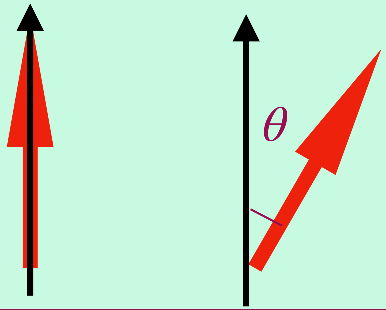
Example problem

- Discover/design magnetic materials with RE elements, satisfying

Define the problem	Discover/design **rare earth free strong permanent magnets**
Figures of merit	**Stable**, **M_s > 1 T, K_1 > 1 MJ/m^3**, T_C > 600 K, low cost, easy manufacturability
Materials to search through	Existing, Combinatorial, Heuristics —> screening Better ways? Machine learning



Hexagonal closed structure (HCP)



MAE

$$\Delta E = K_1 \sin^2 \theta$$

$K_1 > 0$: uniaxial anisotropy

$K_1 < 0$: planar anisotropy

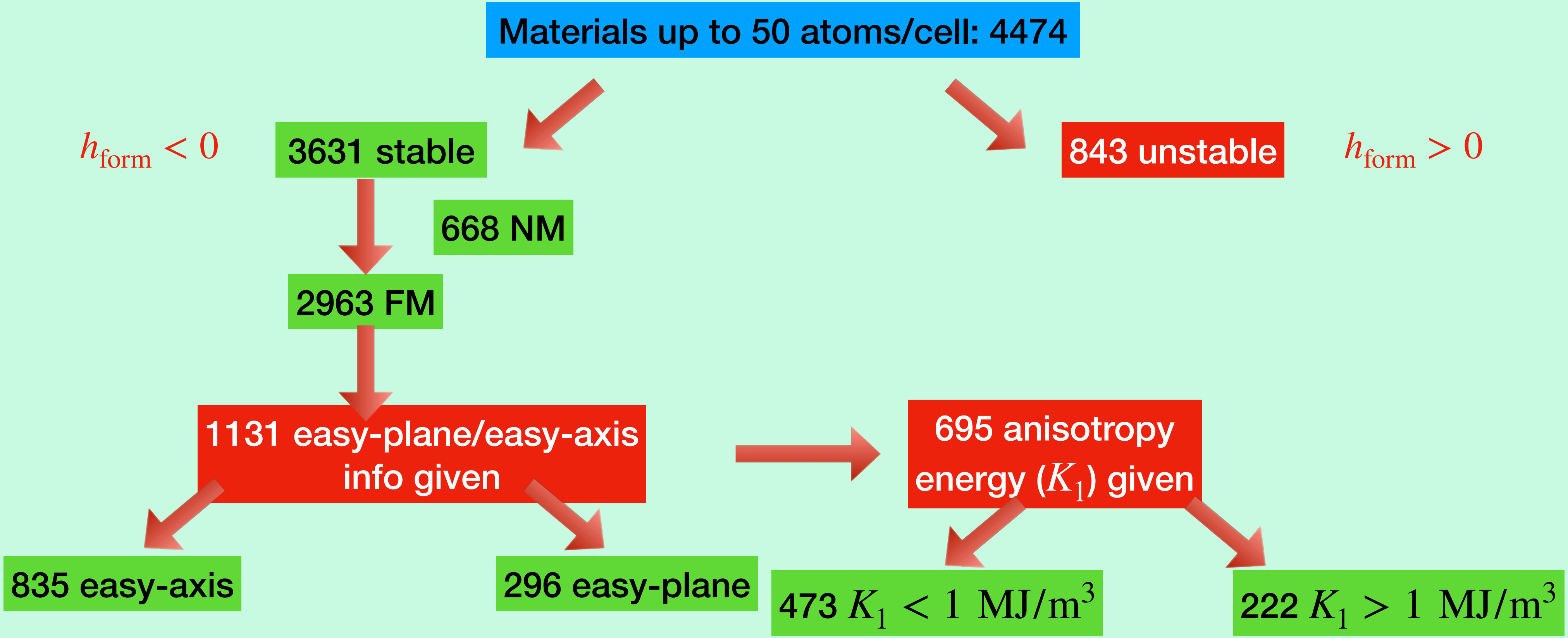
Calculating K_1 via DFT is very expensive and challenging

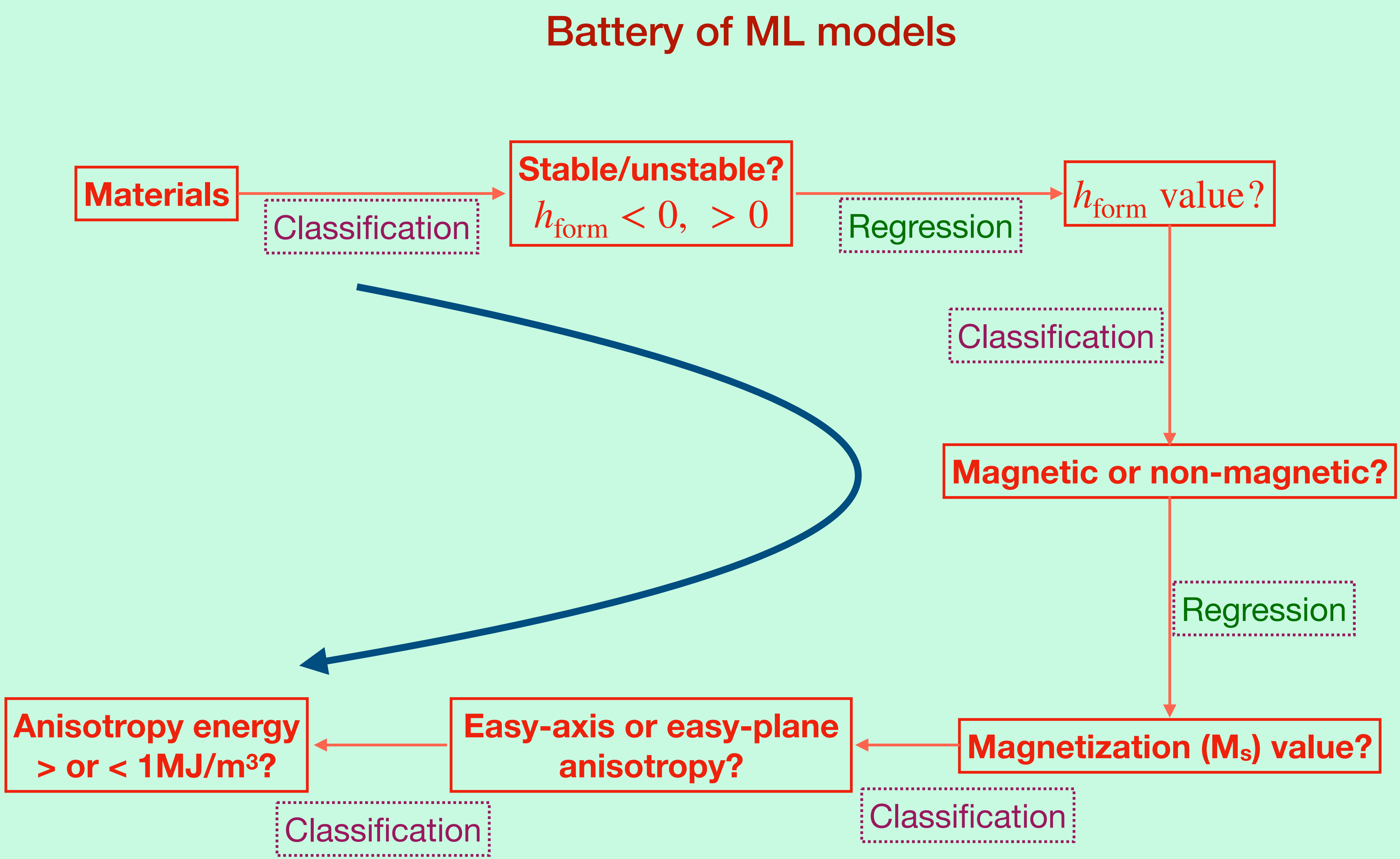
Paucity of data
Particularly data on magnetic properties

- Data from public databases

- ***NovoMag***
- ***Novamag***

DFT generated databases for magnetic materials




Battery of ML models
Materials
Classification
Stable/unstable?
$h_{\text{form}} < 0, \ > 0$
Regression
h_{form} value?
Classification
Magnetic or non-magnetic?
Regression
Magnetization (M_S) value?
Classification
Easy-axis or easy-plane anisotropy?
Classification
Anisotropy energy > or < 1MJ/m^3?

Features used

Table 2: Complete list of features for all the machine learning tasks

	Features	Symbol	Dimension
(I)	L^p stoichiometry norm(p=2,3)	L^2, L^3	2
	stoichiometry entropy	S_e	1
(II)	composition-weighted atomic number(Z)	$\bar{Z}$, $\lvert\delta Z\rvert$, $Mo(Z)$	3
	composition-weighted electronegativity(e)	$\bar{e}$, $\lvert\delta e\rvert$, $Mo(e)$	3
	composition-weighted period(P)	$\bar{P}$, $\lvert\delta P\rvert$, $Mo(P)$	3
	composition-weighted group(G)	$\bar{G}$, $\lvert\delta G\rvert$, $Mo(G)$	3
	composition-weighted valence electron number(v)	$\bar{v}$, $\lvert\delta v\rvert$, $Mo(v)$	3
(III)	Sine matrix eigenvalues	$E_i^{SM}(i = 1-50)$	50
	total		68

Features

$$L^p = \left(\sum_i x_i^p\right)^{1/p}$$

$$S_e = -\sum_i x_i \ln x_i$$

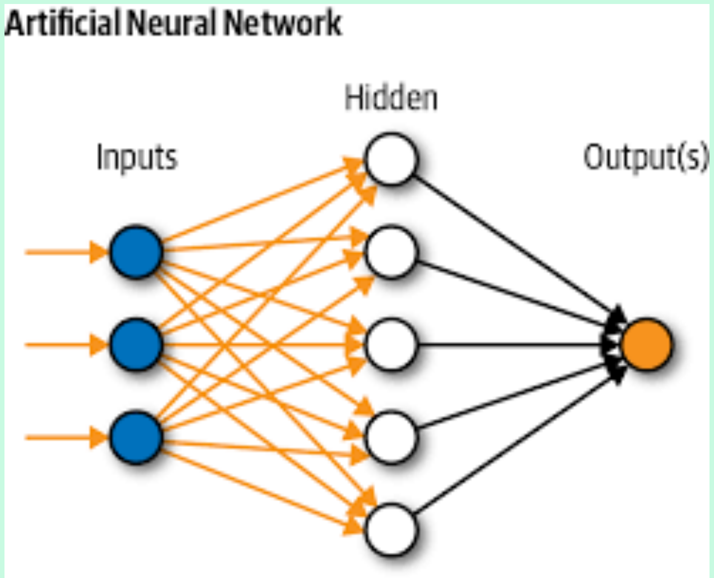
$$\bar{Z} = \frac{1}{n}\sum_i Z_i; \quad |\delta Z| = \sum_i x_i |Z_i - \bar{Z}|$$

$$M_{ij}^{\text{sine}} = 0.5 Z_i^{2.4} \quad i = j$$

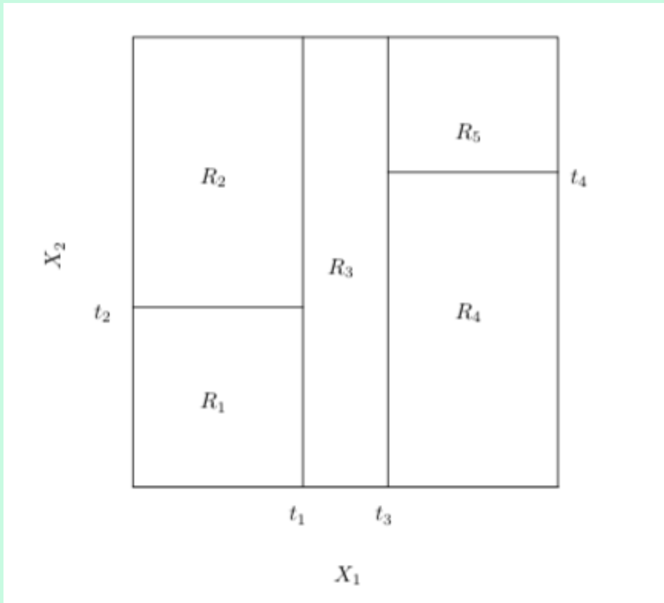
$$M_{ij}^{\text{sine}} = Z_i Z_j \left|\mathbf{B} \, . \sum_k \hat{e}_k \sin^2(\pi \hat{e}_k^T \, . \, \mathbf{B} \, . \, (\vec{R}_i - \vec{R}_j))\right|^{-1} \quad i \neq j$$

Models

Neural network



Random Forest



Model performance

Stability classifier

Model	Accuracy	Precision	Recall	f1
NN	0.88	0.91	0.95	0.93
RF	0.910	0.93	0.96	0.95

Magnetic classifier

Model	Accuracy	Precision	Recall	f1
NN	0.86	0.90	0.94	0.92
RF	0.89	0.93	0.93	0.93

Axis classifier

Model	Accuracy	Precision	Recall	f1
NN	0.74	0.78	0.89	0.83
RF	0.76	0.82	0.86	0.84

h_{form} predictor

Model	**MAE (eV/atom)**	R^2	**r**
NN	0.064	0.82	0.91
RF	0.057	0.86	0.93

M_s predictor

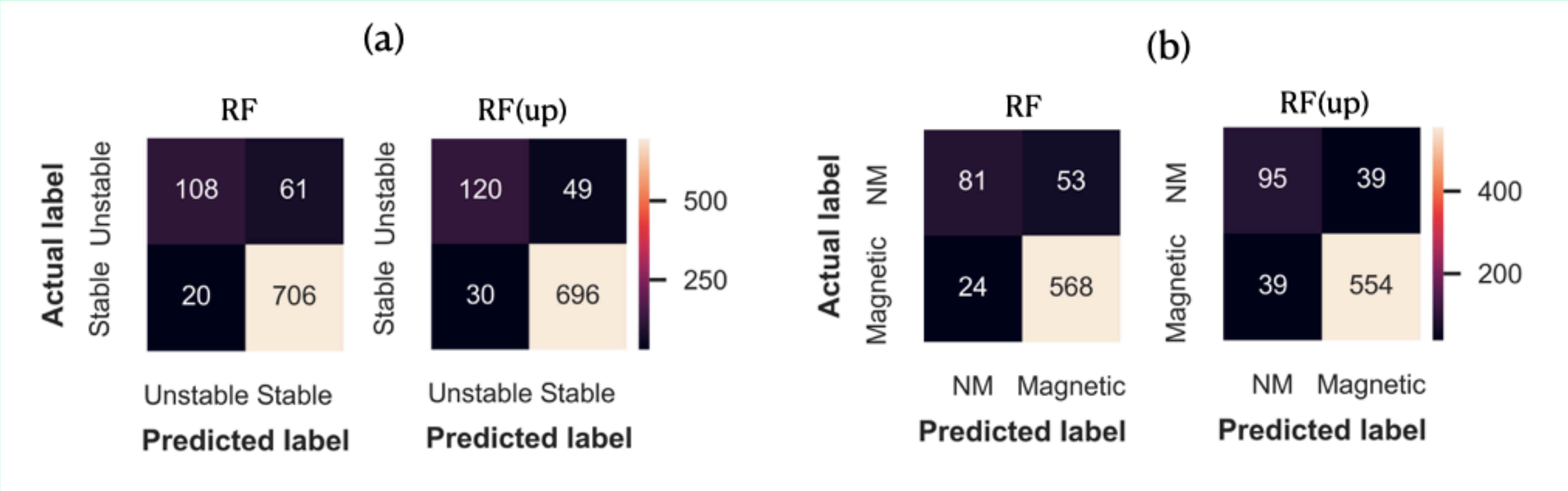
Model	MAE $(\mu_B/\mathrm{\AA}^3)$	R^2	r
NN	0.014	0.84	0.92
RF	0.013	0.85	0.94

Anisotropy classifier

Model	Accuracy	Precision	Recall	f1
NN	0.68	0.68	0.68	0.68
RF	0.75	0.61	0.73	0.66

Understanding the models

(a) stable-unstable classifier (b) magnetic-nonmagnetic classifier



In RF(up), the minority class is up-sampled

(a) Unstable materials: misclassified by RF 36%; misclassified by RF(up) 29%.

(b) Non-magnetic materials: misclassified by RF 39%; by RF(up) 29%

Significantly reduces the misclassification rates.

Understanding the models

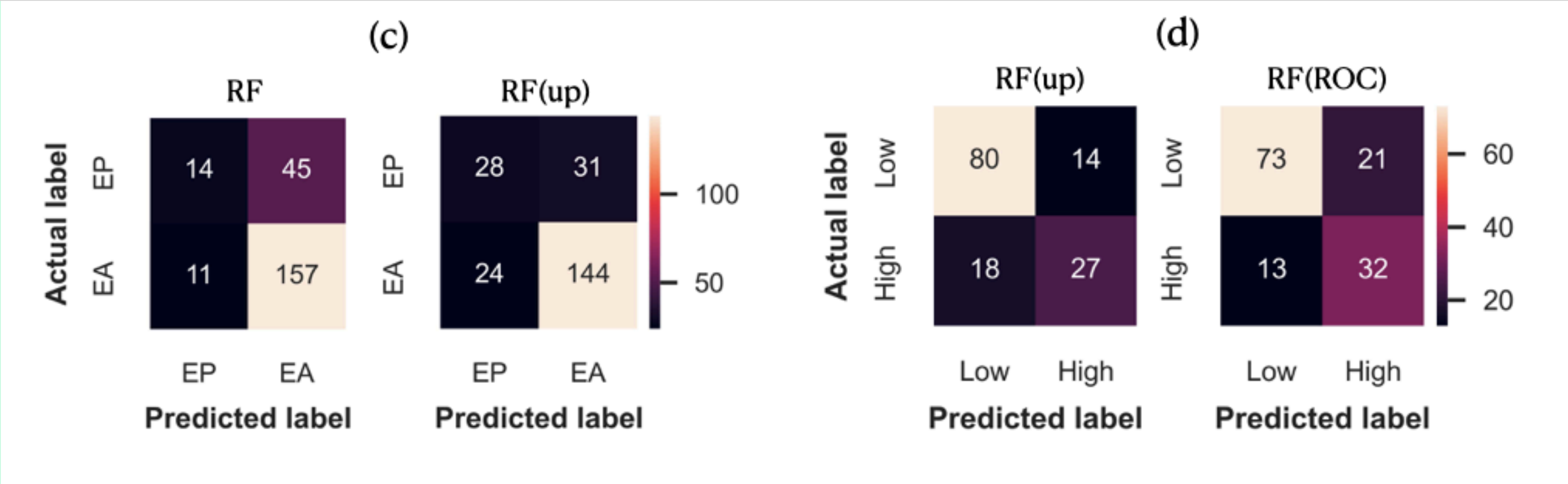
(c) easy-axis vs easy-plane (b) $K_1 < 1$ vs $K_1 > 1$

In RF(up), the minority class is up-sampled

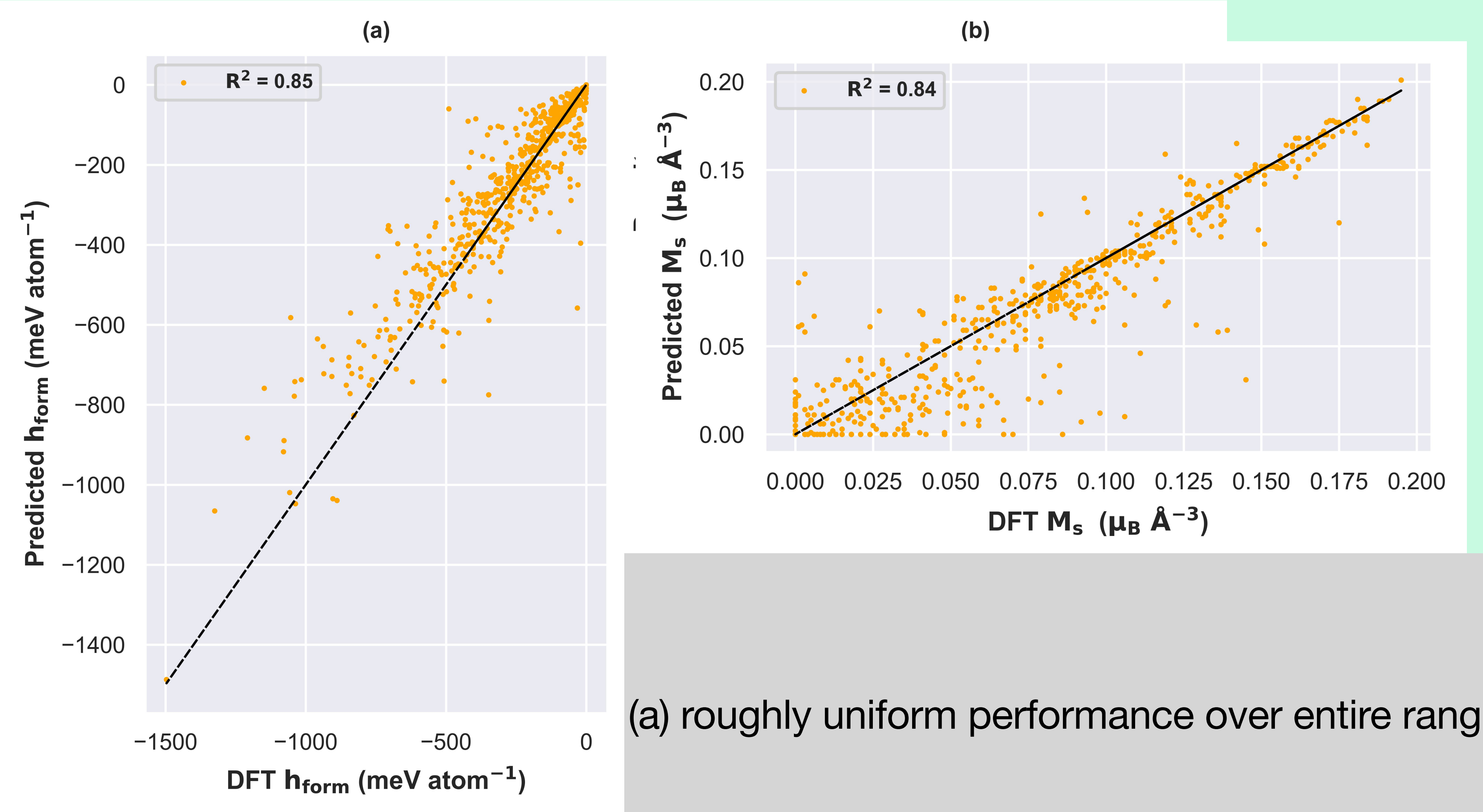
(c) easy-plane: misclassified by RF 76%; misclassified by RF(up) 52%.

(c) high K_1 materials: misclassified by RF 46%; by RF(up) 40%; by RF(ROC) 28%.

Significantly reduces the misclassification rates.



(a) h_{form} regression (b) M_s regression



(a) roughly uniform performance over entire range

(b) better performance for larger M_s

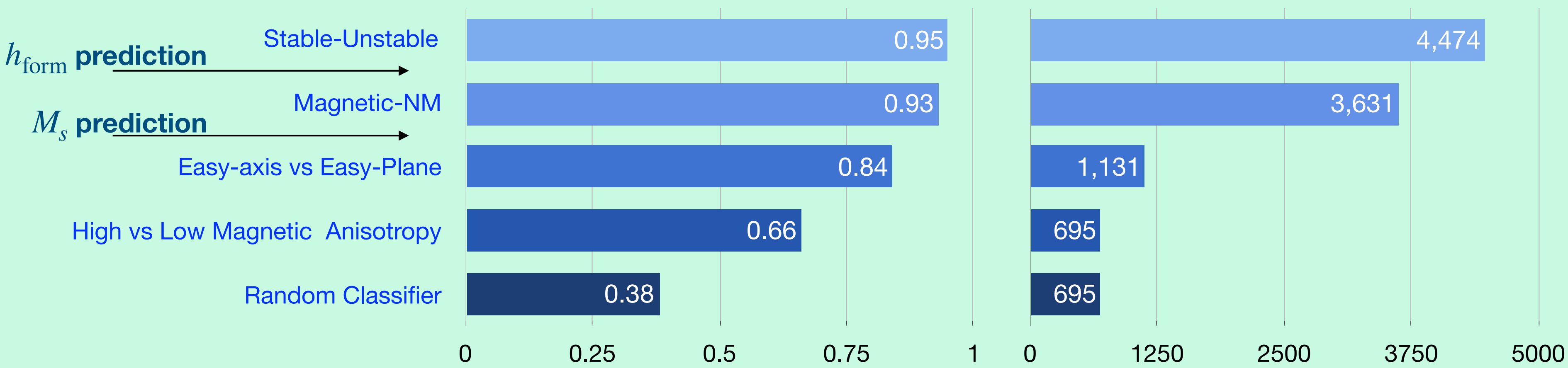
Machine Learning Model Performances

Performance of Regression Models

Quantity	R2 score	MAE	Unit	Data Set Size
h_{form}	0.86	0.057	eV/atom	3631
M_s	0.85	0.15	Tesla	2963

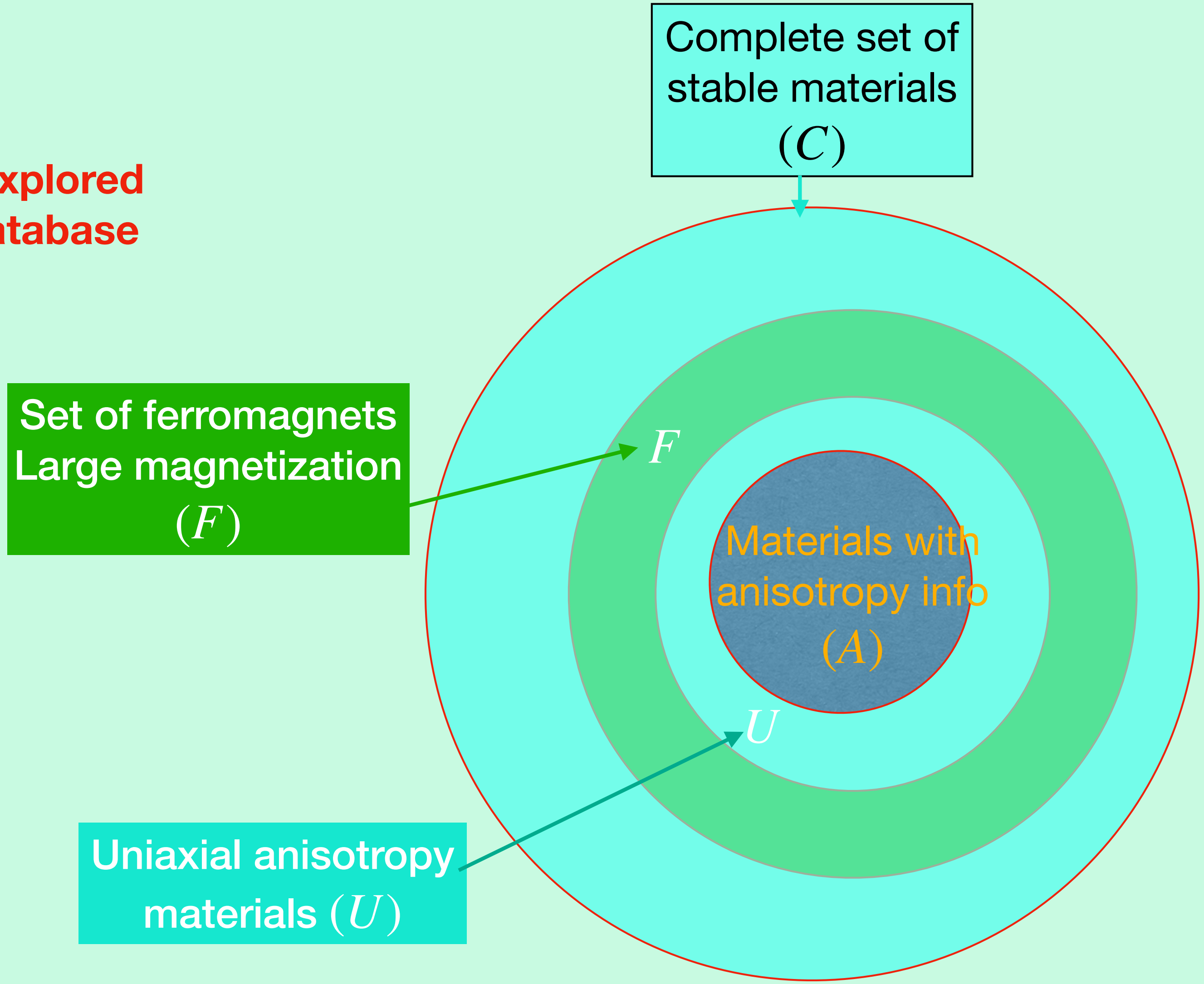
Performance of Classifier Models *
(Based on F1 score)

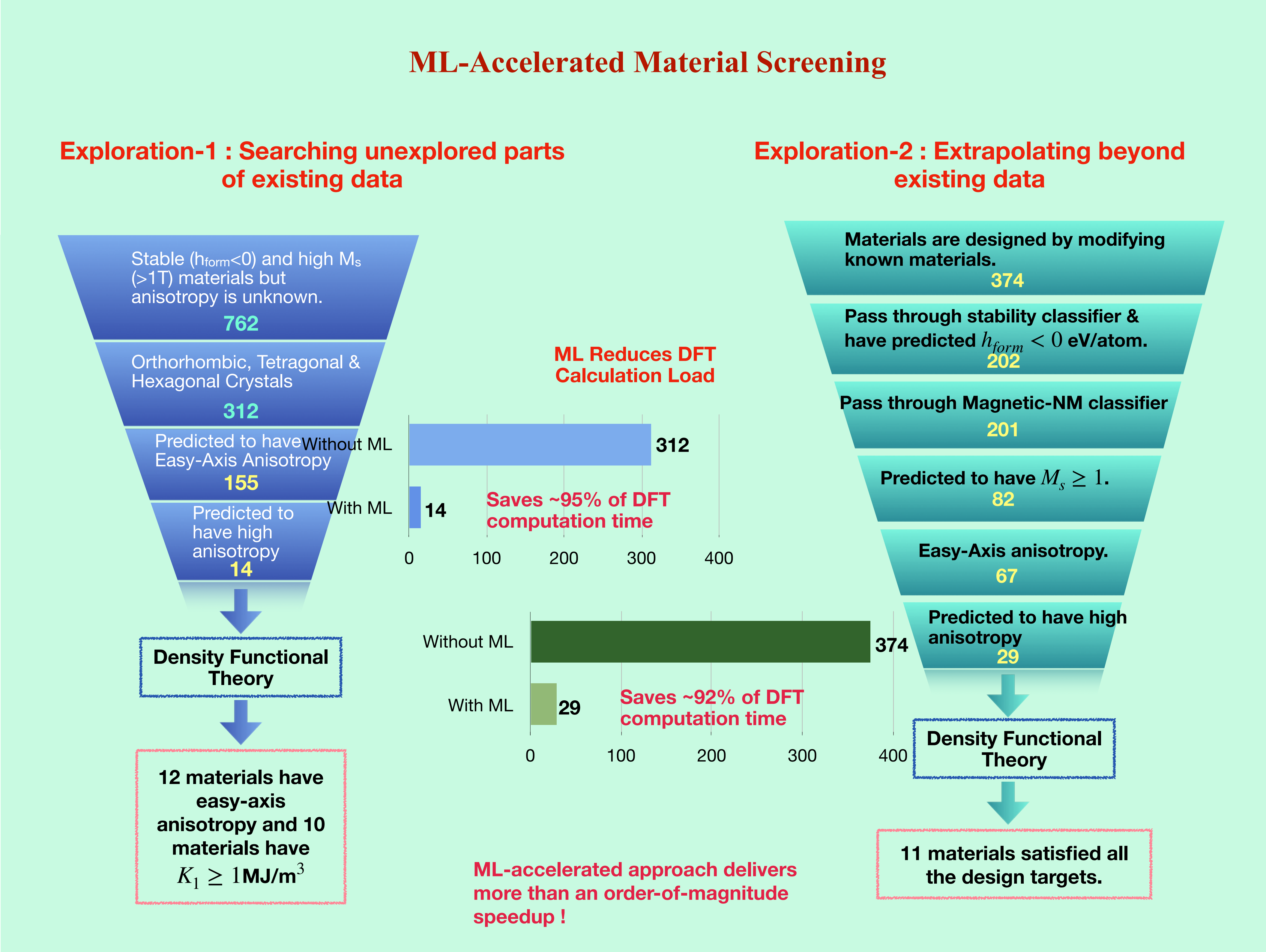
Data Set Sizes



Putting the models to work

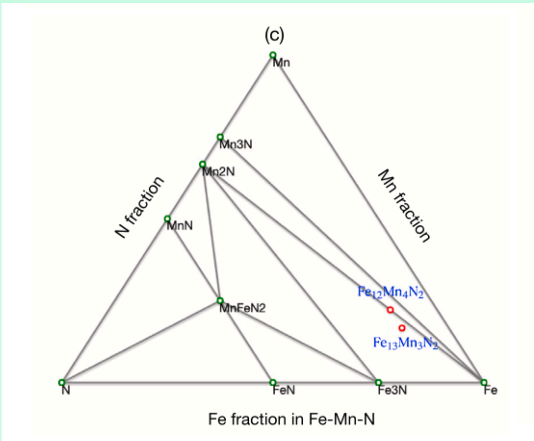
Screening unexplored parts of the database

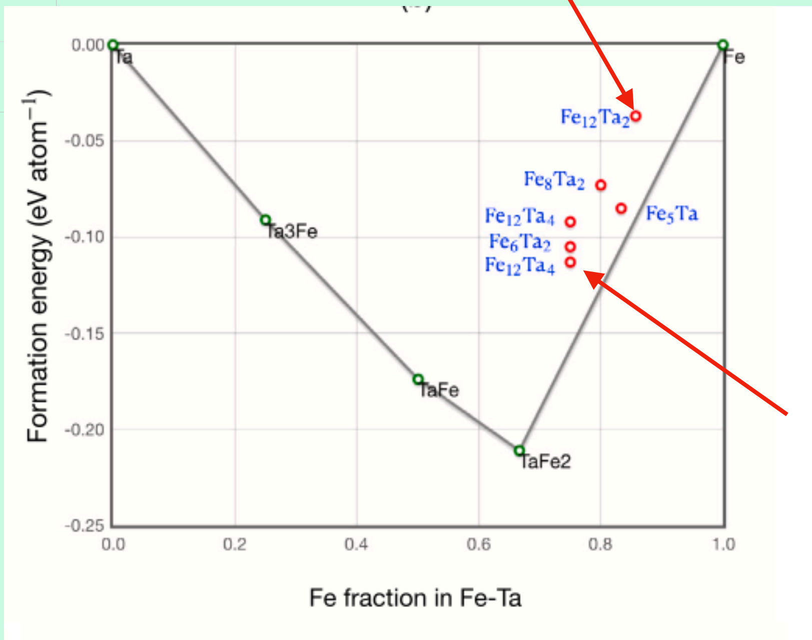



ML-Accelerated Material Screening
Exploration-1 : Searching unexplored parts of existing data
Stable (hform<0) and high Ms (>1T) materials but anisotropy is unknown.
762
Orthorhombic, Tetragonal & Hexagonal Crystals
312
Predicted to have Easy-Axis Anisotropy
155
Predicted to have high anisotropy
14
Density Functional Theory
12 materials have easy-axis anisotropy and 10 materials have $K_1 \geq 1$MJ/m^3
ML Reduces DFT Calculation Load
Without ML
312
With ML
14
Saves ~95% of DFT computation time
0
100
200
300
400
Without ML
374
With ML
29
Saves ~92% of DFT computation time
0
100
200
300
400
ML-accelerated approach delivers more than an order-of-magnitude speedup !
Exploration-2 : Extrapolating beyond existing data
Materials are designed by modifying known materials.
374
Pass through stability classifier & have predicted $h_{form} < 0$ eV/atom.
202
Pass through Magnetic-NM classifier
201
Predicted to have $M_s \geq 1$.
82
Easy-Axis anisotropy.
67
Predicted to have high anisotropy
29
Density Functional Theory
11 materials satisfied all the design targets.

Interesting materials
Obtained from DFT

	h_{form} (meV/atom)	M_s (T)	Anisotropy (MJ/m³)	E_{hull} (meV/atom)
	From the data set			
$Fe_{12}Ta_2$	-37	1.63	6.14	53
$Fe_{12}Ta_4$	-113	1.08	5.38	45
	New materials			
$Fe_{16}N_2$	-30.2	2.20	0.72	
$Fe_{13}Mn_3N_2$	-356	2.09	1.47	108
$Fe_{12}Mn_4N_2$	-468	1.99	2.19	139





How about T_C?

Data ?

- No theory-generated database for T_c
- Experimental data
  - Multiple values for the same material
  - Structure not always reported
- Data source
  - AtomWork, Springer Materials, Handbook of Magnetic Materials (Nelson & Sanvito, PRMat3-104405, Halder et al. Phys. Rev. Appl. 14, 034024)
  - Database created by parsing published literature using NLP models (Court & Cole, Sc. Data DOI: 10.1038/sdata.2018.111)
- How much can be done with out structure information?
  - Nelson-Sanvito, composition weighted chemical features
  - RF model: $R^2 = 0.87$; MAE = 57 K.