A Data-Driven Approach for Predicting Elastic Properties of Inorganic Materials

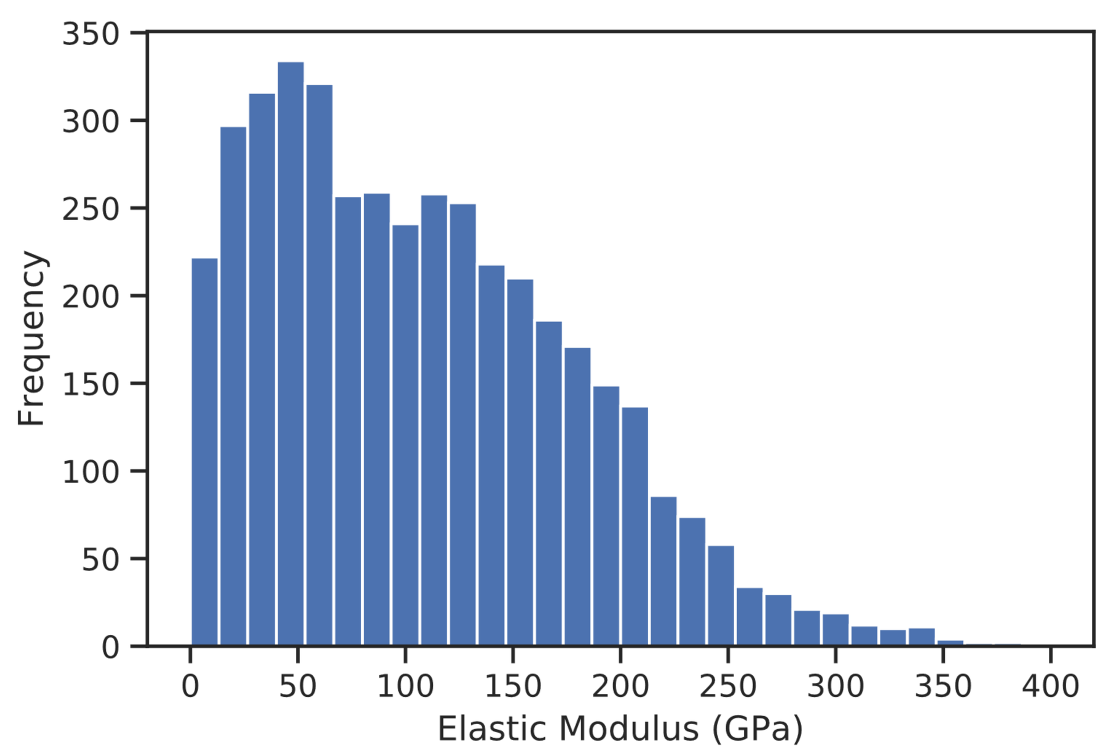
Camila Cendra (ccendra@stanford.edu)

Scope

- Materials discovery from first-principle computations is expensive[1,2].
- Identifying mechanical properties of new materials is crucial to determine their potential functionality.
- The elastic modulus measures a material's resistance to deformation.
- We use machine learning (ML) methods to predict the elastic modulus (y) from common chemical properties, bypassing the need to use more expensive computational methods.

Data and Features

- Dataset[3] consists of 4208 x 136.
- Y data for training is elastic modulus



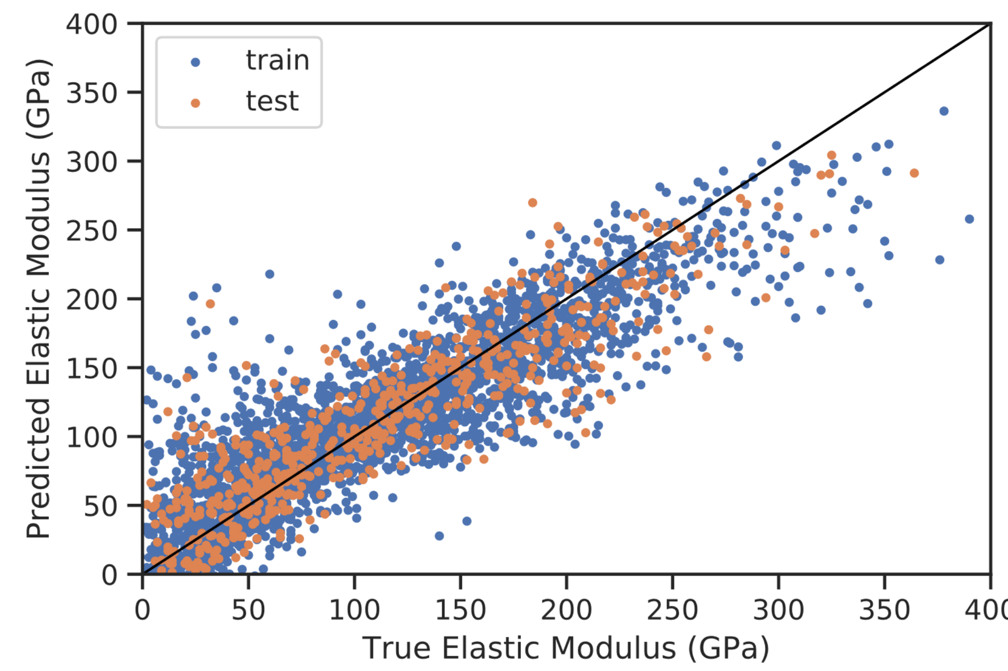
- 3039/537/632 train/dev/test split.
- Features (X) of the model are 135 descriptive attributes.
  - ➢ 118 encode chemical composition
  - ➢ 17 encode heuristic quantities[2]
    - ❖ i.e.: electronegativity, valence electrons, atomic mass and size.
- X was standardized to zero mean and unit variance using training data.

Supervised Learning Models

1) Linear Regression (LR) + Regularization

- Linear model that minimizes least squares loss while penalizing the size of coefficients (ω).

$$\min_{w} ||Xw - y||_2{}^2 + \alpha ||w||_2{}^2$$



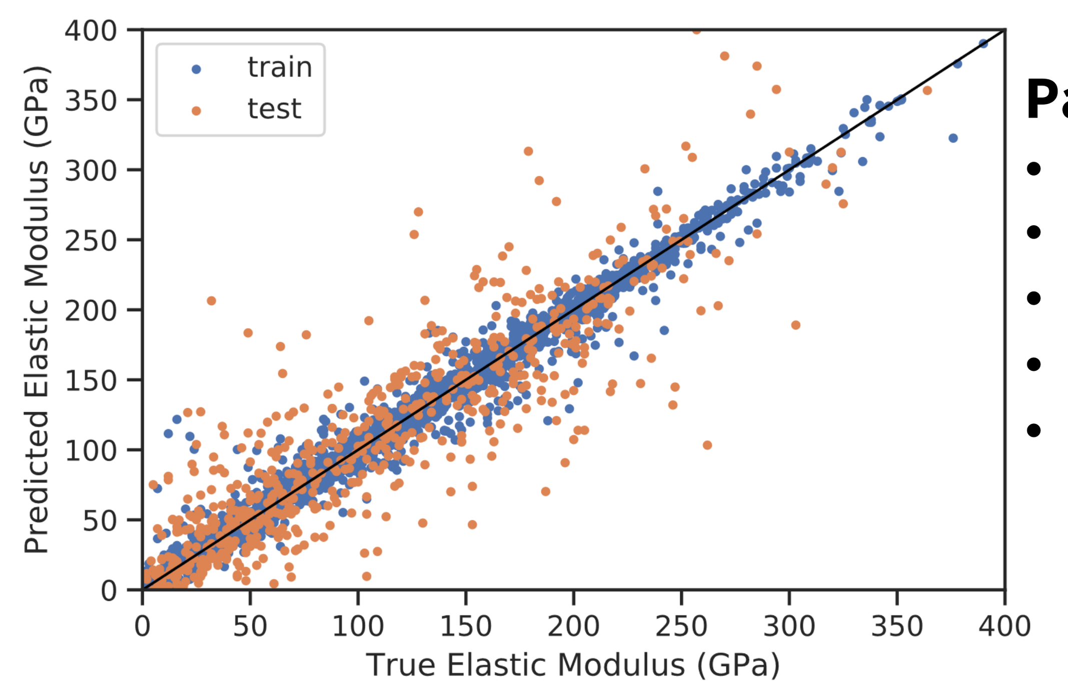
Parameters

- α: 0.5

2) Multi Layer Perceptron (MLP)

- Neural network with one hidden layer. Loss minimization with stochastic gradient descent.

$$Loss(\hat{y}, y, W) = \frac{1}{2}||\hat{y} - y||_2^2 + \frac{\alpha}{2}||W||_2^2$$

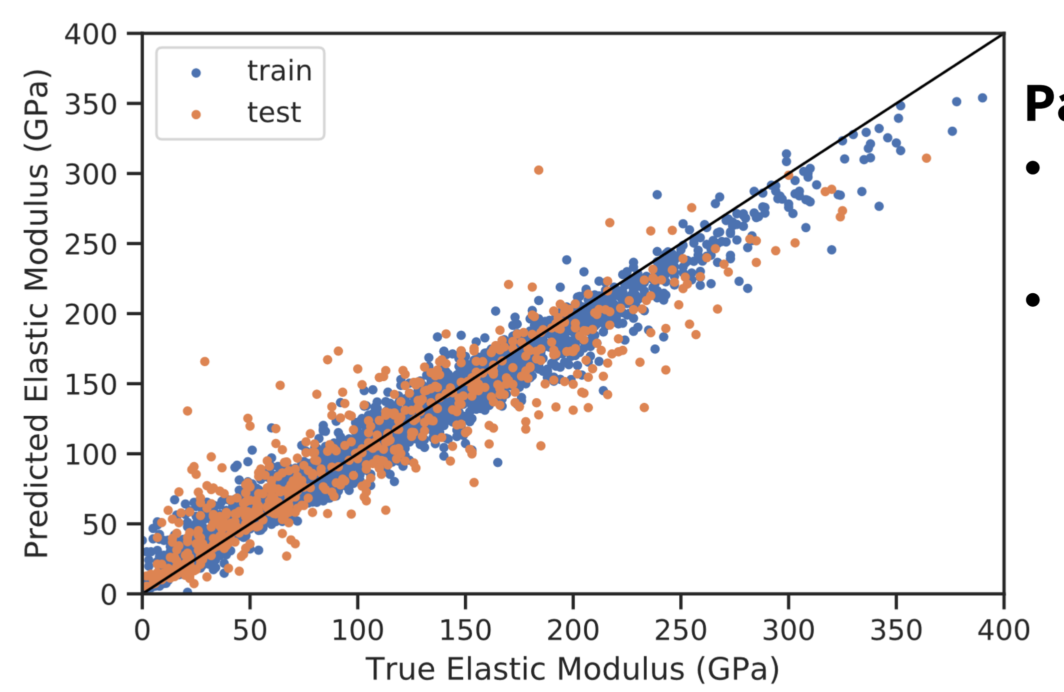


Parameters

- α: 0.0001
- l_rate: 0.001
- activation: relu
- output: identity
- 100 neurons in hidden layer

3) Random Forest Regressor (RFR)

- Bootstrapped meta estimator that fits classifying decision trees.



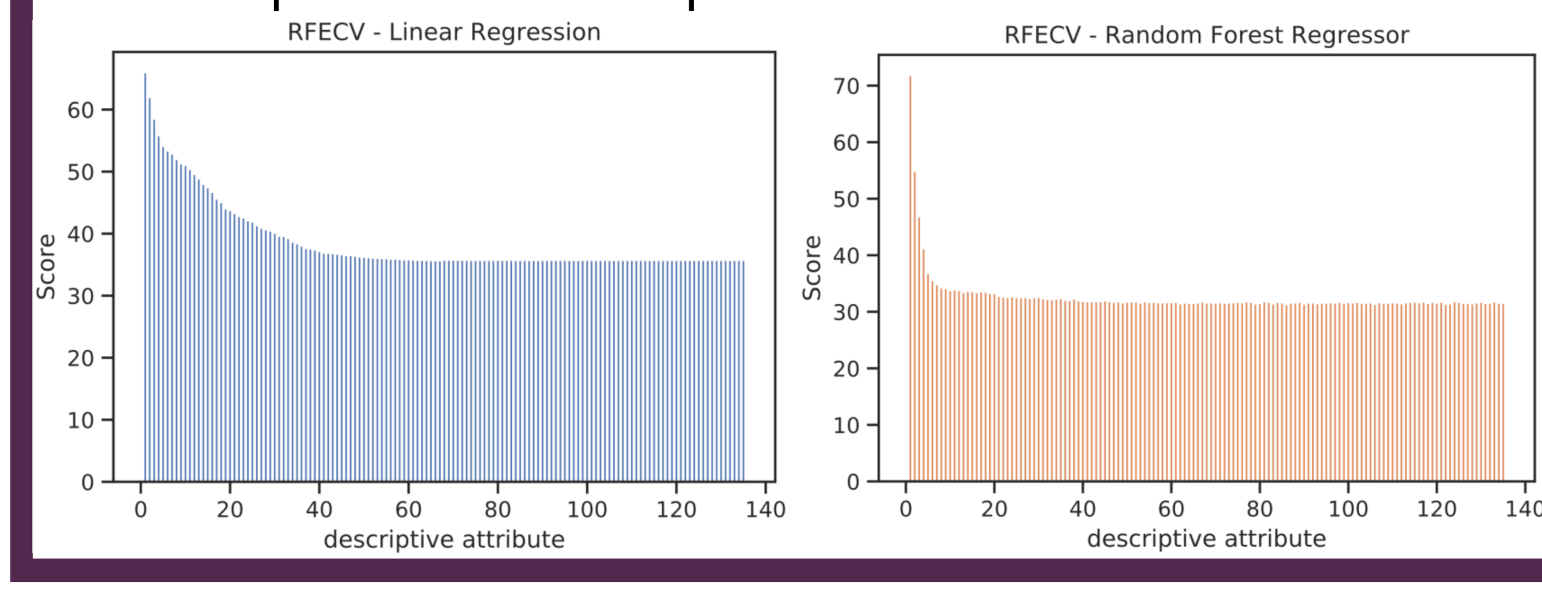
Parameters

- Number trees: 100
- Maximum depth: 15

Results

Model	RMSE train	RMSE test	RMSE 10-fold CV	Pearson, r [p-value] train	Pearson, r [p-value] test
LR	35	42	36 ±16	0.88 [< 0.01]	0.84 [< 0.01]
MLP	28	37	32 ± 16	0.92 [< 0.01]	0.88 [< 0.01]
RFR	27	38	32 ± 16	0.93 [< 0.01]	0.89 [< 0.01]

Feature Importance: as determined by 10-fold recursive feature elimination with cross-validation (RFECV), heuristic and compositional descriptive attributes are both useful.



Discussion

- Using descriptive attributes, which are readily obtained analytically for any given composition, we have predicted the elastic modulus of a diverse set of materials with high accuracy.
- 10-fold CV shows that the prediction performance of the ML models is consistent.
- Both heuristic and compositional features contribute to the models' high performance.

Future Work and References

- Predict other crucial mechanical properties, such as shear modulus and fracture toughness, using the developed methodology.
- Implement advanced ensembling algorithms to achieve higher predictive accuracy.

References:

[1] B. Meredig, A. Agrawal, S. Kirklin, J. E. Saal, J. W. Doak, A. Thompson, K. Zhang, A. Choudhary, C. Wolverton, *Phys. Rev. B - Condens. Matter Mater. Phys.* 2014, *89*, 1.

[2] M. de Jong, W. Chen, H. Geerlings, M. Asta, K. A. Persson, *Sci. Data* 2015, *2*, 150053.

[3] Dataset obtained from the Open Citrination Platform and is publicly available. https://citrine.io/research/open-citrination-platform/