Decoding the genome of cement by Gaussian Process Regression

Yu Song, Yongzhe Wang, Kaixin Wang,
Prof. Gaurav Sant, and Prof. Mathieu Bauchy*

Background and motivations

- Concrete (and cement) represents a typical engineering material that involves a heavy carbon footprint
- Top-one produced engineering material in human history
- Concrete is extensively used in modern construction

Concrete constructions and the environmental issues behind

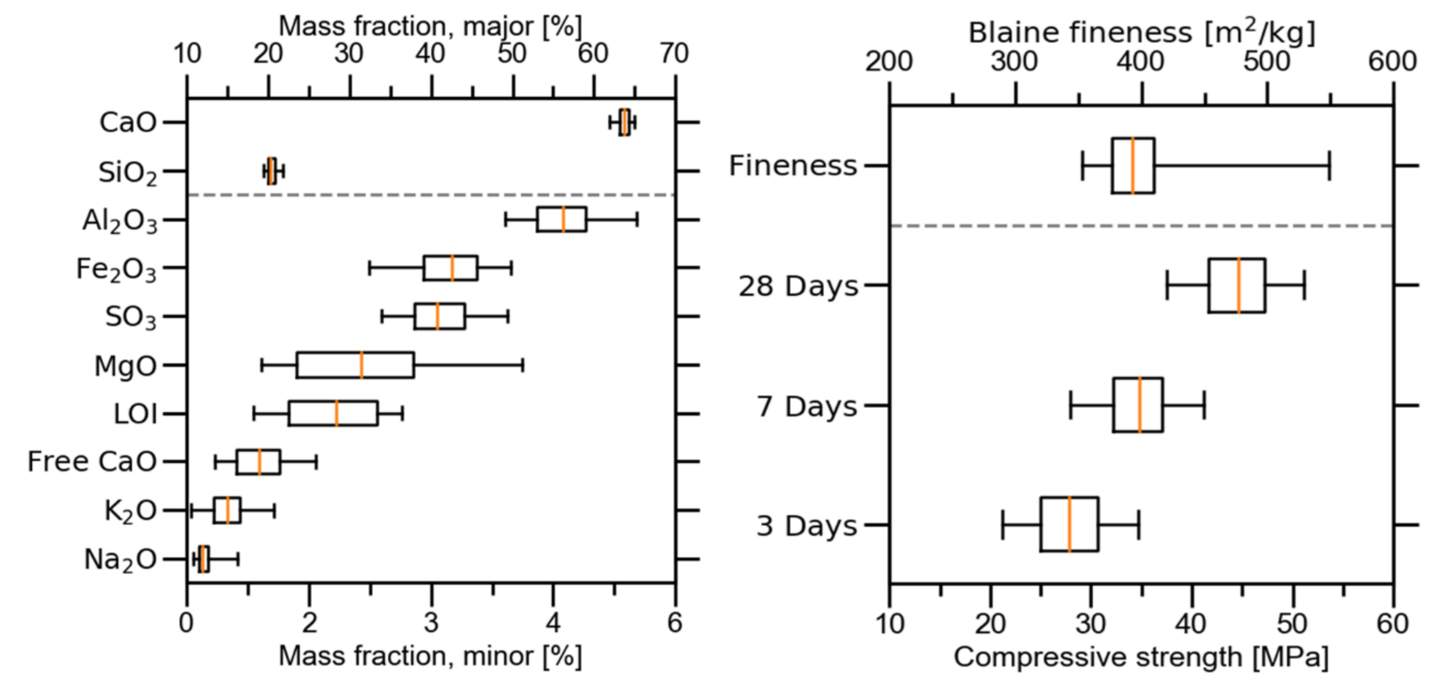
- Cement is a complex composite of physical and chemical phases formed by various oxides (i.e., the cement "genome")
- 1-ton cement ≈ 1-ton CO_2 and 4×10^9 tons produced annually
- Need to improve strength performance and reduce the carbon embodiment of cement
- Challenges: identify the effect of each oxide compounds
- A good manifestation of the common difficulties faced by many studies in engineering materials

Methods

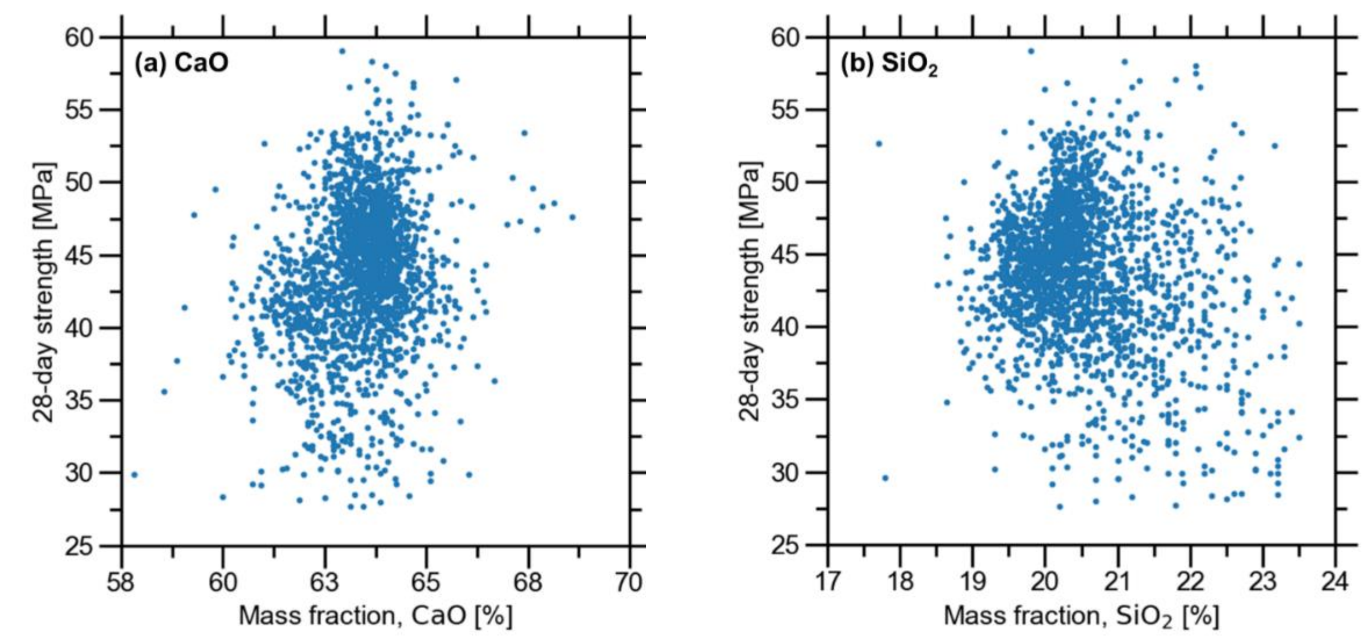
- Machine learning is more used to predict material properties
- This study aims to apply Gaussian process regression (GPR) to explain the composition-strength relationship of cement
- GPR is adopted because of several advantages:
  1. Non-parametric: good model flexibility
  2. Leveraging the prior knowledge via kernel selection
  3. Uncertainty estimation of the prediction: practical for the rational design of engineering materials

The cement dataset

- This dataset is collected over many U.S. cement testing institutes, consisting of 2060 cement samples
- Input features: oxide contents (CaO, SiO_2, etc.) and fineness
- Output target: 28-day strength (i.e., the standard reference)



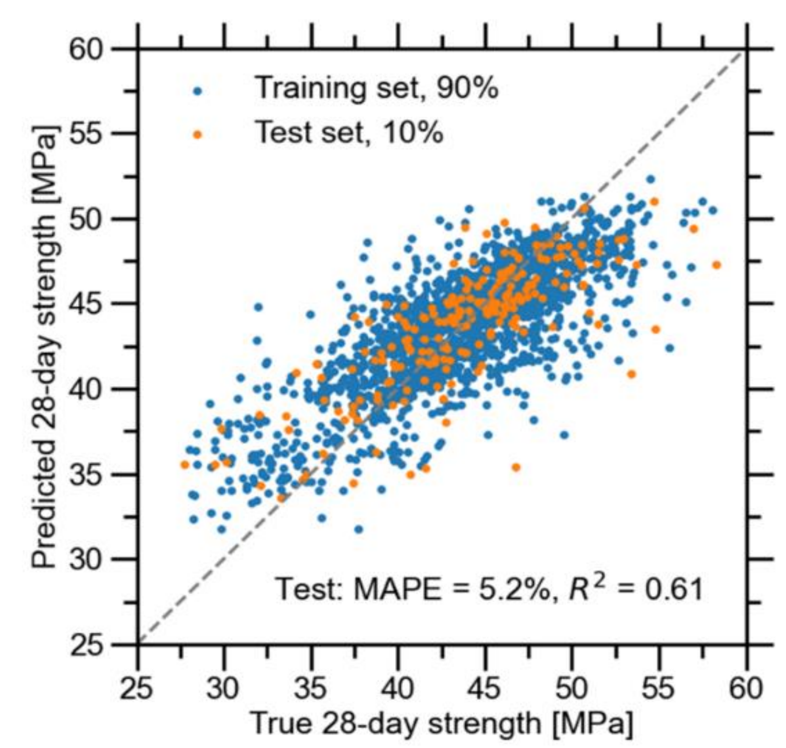
Variation of each item in the dataset.



No correlation between the strength and input features

GPR optimization and accuracy

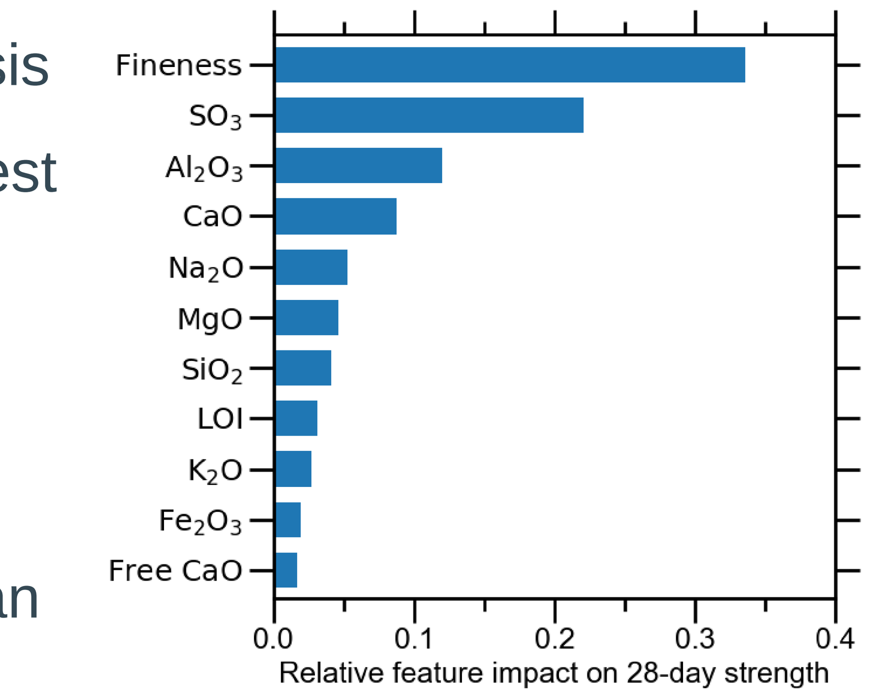
- [Linear + RBF] as the best kernel
- Noise is tuned for overfitting
- Test: 10% stratified samples
- The highest accuracy ever reported on this dataset
- An illustration of the common challenges for studying



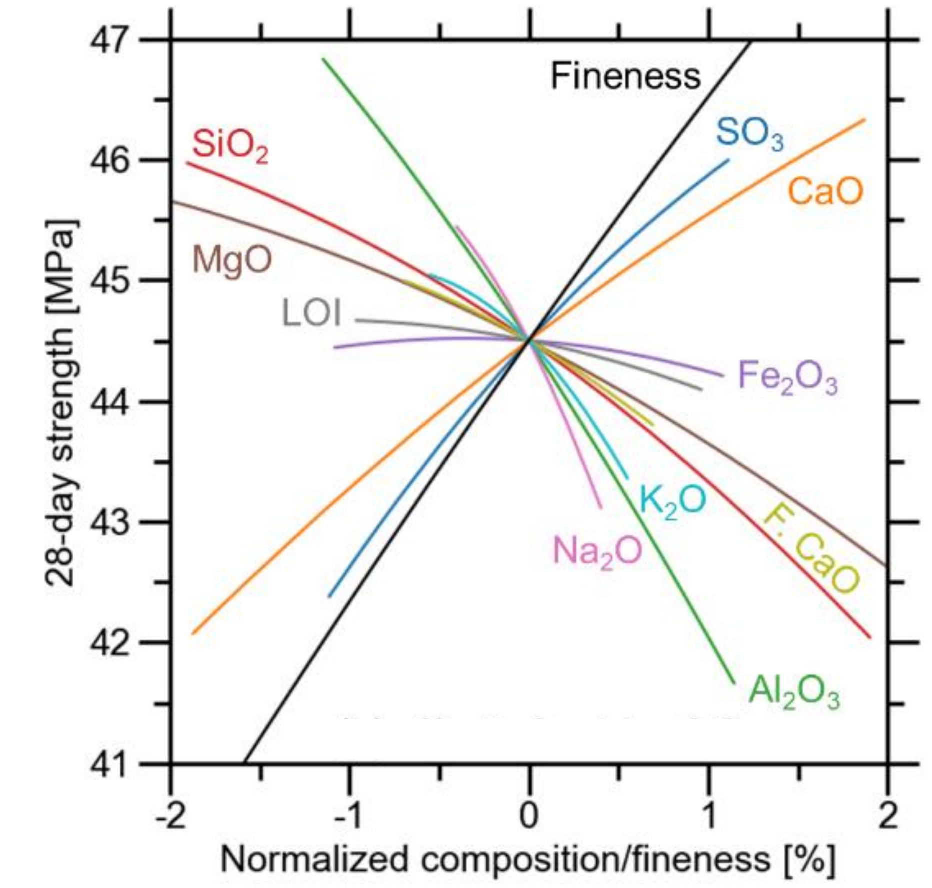
Predicted vs. true strength

Interrogation on the trained model

- The first feature impact analysis is based on the permutation test
- This analysis focuses on the magnitude of the influence
- The rank highlights that the fineness is more important than any oxide

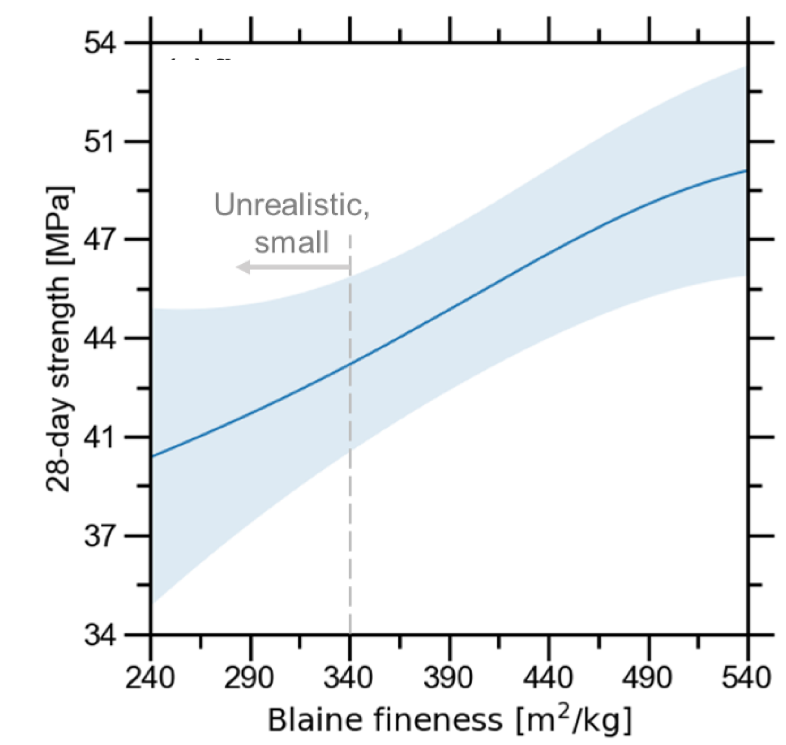


Feature impact analysis



Feature effect analysis

- Feature effect analysis
- Tune only one feature from its min to max
- The trends of the major oxides are broadly in line with cement chemistry
- Important insight for high-performance cement



Uncertainty for Fineness

- Uncertainty of GPR prediction
- The shadowed region represents the standard deviation
- The high uncertainty for small fineness is due to limited sample
- The awareness of uncertainty is very practical for the design of engineering materials

Remarks: The presented archetypal pipeline successfully maps the material property, and it can be applied to advance the study, design, and optimization of many other materials

Presented in the Workshop on machine learning for engineering modeling, simulation and design @ NeurIPS 2020, Dec 11th, 2020