Bayesian Optimization-Guided Experimental Design for Graphene-Reinforced Cementitious Composites

Ben Yuan[1] Abhimanyu Goel[2] Virgil Gavriliuc[1] Artem Maevskiy[1] Harn Wei KUA[2] Maxim Rybin[1] Konstantin S. Novoselov[1]

[1]*[Institute for Functional Intelligent Materials (IFIM), National University of Singapore, Singapore]*

[2]*[Department of the Built Environment, National University of Singapore, Singapore]*

Correspondence to: [Maxim Rybin] maxim-r@nus.edu.sg

1. Background and Motivation

Graphene has attracted extensive attention due to its outstanding physical and chemical properties[1], with strong potential to enhance the mechanical[2] and functional performance of cement-based composites[3]. However, the mixture design space is high-dimensional (e.g., water/sand ratio, plasticizer dosage, and graphene type/dosage), and the resulting strength can vary substantially even under nominally similar mixture parameters. This complexity makes trial-and-error optimization inefficient and limits reliable performance prediction, motivating the integration of data-driven and uncertainty-aware AI methods to accelerate experimental decision-making for graphene-reinforced cementitious materials.

2. Objective

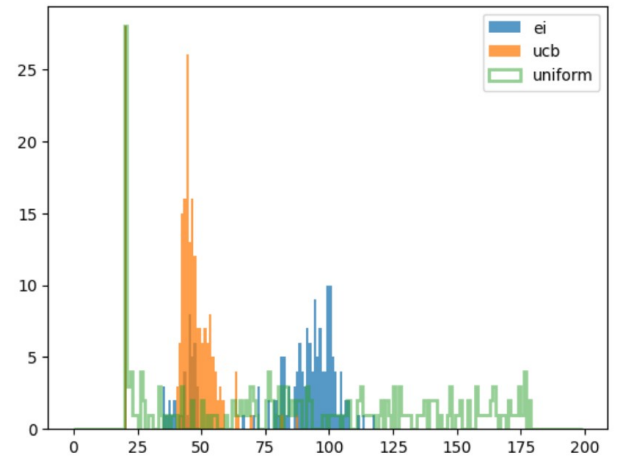
This study aims to experimentally investigate the influence of graphene incorporation on the mechanical performance and hydration behavior of cement-based materials. Building upon the experimental observations, an AI-assisted performance prediction framework is developed to support data-driven mixture design and enable more efficient exploration of optimal composition parameters.

3. Methods

Cement-based composites incorporating different types and dosages of graphene are experimentally prepared and characterized to obtain compressive strength and related mechanical properties. Isothermal calorimetry and nuclear magnetic resonance (NMR) analyses are employed to examine the influence of graphene on cement hydration and microstructural evolution. Based on the experimental dataset, machine learning models are developed to predict compressive strength from mixture parameters. In particular, Gaussian Process regression is adopted to provide uncertainty-aware predictions, enabling Bayesian optimization strategies[4] with acquisition functions such as Expected Improvement (EI) and Upper Confidence Bound (UCB) to guide the selection of subsequent mixture experiments.This data-driven approach aligns with recent advances in AI-assisted materials discovery and active learning frameworks [5].

4. Results

Experimental results show that appropriate graphene incorporation leads to an approximately 30% enhancement in compressive strength compared with the reference mixtures. In addition, hydration characterization using isothermal calorimetry and solid-state nuclear magnetic resonance (ss-NMR) reveals that mixtures containing graphene oxide (GO) exhibit significantly stronger early-stage hydration reactions compared with control samples without additives. Machine learning models effectively capture the relationship between mixture parameters and strength performance, achieving prediction errors significantly lower than random baselines. More importantly, Bayesian optimization simulations demonstrate that AI-guided strategies can identify high-performance mixtures with substantially fewer experimental trials, where the Upper Confidence Bound (UCB) acquisition function consistently outperforms alternative strategies.



- horizontal axis: number of measurements before reaching the largest compressive strength
- vertical axis: counts

Fig. 1: Comparison of Bayesian optimization strategies (EI and UCB) and random sampling for experimental mixture design efficiency

5. Significance

This work demonstrates that uncertainty-aware Bayesian optimization can effectively guide experimental mixture design and reduce the number of experiments required to identify high-performance graphene-reinforced cement composites. By integrating experimental validation with AI-driven decision-making, the proposed framework establishes a practical pathway toward data-driven and closed-loop materials optimization. The study highlights how artificial intelligence can move beyond prediction to actively support experimental planning and accelerate the development of advanced civil engineering materials.

Acknowledgments

The authors acknowledge the support from the Institute for Functional Intelligent Materials (IFIM), National University of Singapore, and collaborators who contributed to this work.

References

[1]K. S. Novoselov, V. I. Fal'ko, L. Colombo, P. R. Gellert, M. G. Schwab, and K. Kim, "A roadmap for graphene," Nature, vol. 490, no. 7419, pp. 192–200, 2012.
[2]Pan, Z., He, L., Qiu, L., et al., "Mechanical properties and microstructure of a graphene oxide–cement composite," Cement and Concrete Composites, 2015.
[3]Lv, S., Ma, Y., Qiu, C., et al., "Effect of graphene oxide nanosheets on microstructure and mechanical properties of cement composites," Construction and Building Materials, 2013.
[4] J. Snoek, H. Larochelle, and R. P. Adams, "Practical Bayesian Optimization of Machine Learning Algorithms," in Advances in Neural Information Processing Systems (NeurIPS), 2012
[5] T. Lookman, F. J. Alexander, and K. Rajan, "Information science for materials discovery and design," Nature Reviews Materials, vol. 1, no. 9, 2016

Appendix A. [how Gaussian Process model works]

Gaussian Process (GP) regression predicts the value at any new input by treating the observed data as points drawn from a smooth underlying curve; for each new point it computes a predicted mean (the best-guess value) and a predicted standard deviation (how uncertain that guess is). These two quantities are derived from a covariance (kernel) function that measures how similar any two inputs are — points that are close together in the input space influence each other more strongly, pulling the mean toward nearby observed values and reducing the uncertainty, while isolated points retain a higher standard deviation. The result is a smooth estimate of the target variable together with a confidence band that reflects how much data support each prediction.

To select the next best measurement, the two acquisition functions are defined as follows:

• Upper Confidence Bound (UCB):
UCB suggests the next sampling point by maximizing the value of (mean + k · standard deviation). In this work, k = 1 is used.

• Expected Improvement (EI):
For each input x, EI calculates the expectation
$E[\max(0, \hat{y}(x) - y_max)]$,
where y_max is the highest target value observed so far, and $\hat{y}(x)$ is the prediction at x modeled as a normal distribution:
$\hat{y}(x) \sim N(\hat{y}_mean(x), \sigma^2(x))$.