

# Hybrid GA–PSO Optimization of Ambient-Cured Low-Clinker Alkali-Activated Mortars: Model Development and Experimental Verification

S Cho<sup>1</sup>, Y Chae<sup>1</sup> and M O Kim<sup>1,\*</sup>

<sup>1</sup> Department of Civil Engineering, Seoul National University of Science and Technology, Seoul, Republic of Korea

\* Contact author: minookkim@seoultech.ac.kr

**Abstract.** Literature-trained surrogates increasingly guide low-clinker binder design for ocean structures, but their recommendations are seldom stress-tested. This study tests an XGBoost–hybrid GA–PSO framework for ambient-cured fly ash mortars with OPC and silica fume (SF), built on 331 records from 15 studies (random-split  $R^2 = 0.951$ ) and verified with six mortars at OPC + SF = 15 wt%. Source-held-out validation of a re-trained surrogate raised the RMSE from 3.65 to 11.6 MPa. At equal budget, GA and GA–PSO reached the reference optimum in 93 % and 90 % of runs and PSO in 30 %; the hybrid's PSO stage added no measurable benefit. Resampled surrogates favored an OPC-rich near-optimal region rather than the returned OPC12.5–SF2.5 point, and the tests (maximum 49.8 MPa at OPC12–SF3) did not separate OPC12–SF3 from OPC9–SF6. Over-prediction reached +51 % where SF-bearing training records were sparse. Source-aware validation, region-level reporting and constituent-level coverage checks are recommended.

## 1. Introduction

Harbor works, breakwaters and offshore foundations consume large volumes of concrete whose clinker carries most of its embodied CO<sub>2</sub> [1]. Fly ash alkali-activated binders replace most of that clinker [2,3], but low-calcium ash reacts slowly at room temperature, and heat curing is rarely practicable for cast-in-place marine work. Small additions of OPC supply calcium that accelerates early reaction [4,5], and silica fume (SF) supplies reactive silica [6]; dividing a limited replacement budget between them is a coupled design problem.

Surrogates trained on literature data are now routinely coupled with metaheuristic search to propose such mixtures [7–12], usually supported by a random-split test score and, less often, one fabricated optimum. This leaves the robustness of the recommendation untested. Literature databases are clustered by source, so a random split scores interpolation within known laboratories rather than prediction for a new one [13,14]; optimizers are compared, if at all, on test functions rather than on the surrogate actually used; equally accurate tree ensembles can place their maxima differently [15]; and an optimized mixture may lie where the training data contain few records with the constituents being optimized.

This paper therefore asks how robust a literature-trained optimization is to source dependence, optimizer choice, surrogate uncertainty and limited training coverage. An existing XGBoost–GA–PSO design of ambient-cured OPC–SF fly ash mortars, with its six-mixture experimental verification, is used as the test case and is subjected to four diagnostics: source-held-out validation (D1), an equal-budget

optimizer benchmark (D2), optimum identifiability (D3) and coverage of the verification mixtures by the training data (D4).

## 2. Methods

### 2.1. *Database and the two surrogates*

The database holds 331 strength records from 15 studies on ambient-cured fly ash mortars with sodium-based activators. Eleven inputs describe each record: ash CaO, SiO<sub>2</sub> and Al<sub>2</sub>O<sub>3</sub>; OPC and SF replacement; the silicate-to-hydroxide ratio (SS/SH); hydroxide molarity (SH); sand-to-binder (S/B) and activator-to-binder (A/B) ratios; superplasticizer dosage; and age. Because the database carries no study identifier, records were grouped by ash oxide composition into 15 groups that serve as source labels; this grouping is a proxy and need not map one-to-one onto the 15 studies. SH is unreported for 64 records (19.3 %), which the tree learners route natively.

Two surrogates are distinguished throughout. The original surrogate is the XGBoost model selected from five learners (XGBoost, LightGBM, CatBoost, random forest, support vector regression) on an 8:2 random split with Bayesian hyperparameter search under five-fold cross-validation [16,17] and interpreted with SHAP [18]; it drove the design and the experiments and supplies the predictions in section 3.4. The re-trained surrogate is a second XGBoost model built independently on the same records and inputs, tuned by 120 Parzen-estimator trials under five-fold cross-validation of an 8:2 training split (test  $R^2 = 0.966$ , RMSE = 2.88 MPa) and refitted to all 331 records for D2 and D3. It does not replace the original model; it tests whether comparable random-split accuracy yields the same engineering recommendation. **Model selection was based solely on random-split performance; LightGBM, CatBoost, random forest and support vector regression were not evaluated under source-held-out validation.**

### 2.2. *Constrained design problem*

The objective is the predicted 28-day strength. OPC + SF is fixed at 15 wt% of the binder, since Vicat tests gave an initial set of 60 min at 15 % OPC and 35 min at 20 %. The decision variables are OPC (0–15 wt%), SH (6–14 mol/L), SS/SH (1.5–3.5), A/B (0.45–0.60, the lower bound from mixing trials) and S/B (1.0–2.0) (SH and SS/SH after [19], S/B after [20]); the ash oxides are fixed at the XRF values of the experimental ash (CaO 15.94, SiO<sub>2</sub> 43.80, Al<sub>2</sub>O<sub>3</sub> 25.74 wt%), superplasticizer at 0.02 wt% and age at 28 d. In the diagnostics the equality is enforced by writing  $SF = 15 - OPC$ .

### 2.3. *Robustness diagnostics*

Validation (D1). **For the re-trained XGBoost surrogate, repeated random five-fold cross-validation (five repeats) was compared** with leave-one-source-out validation, in which each source is predicted by a surrogate trained on the other 14 [14]. Both used the re-trained hyperparameters without nested re-tuning, so the source-held-out estimate may still be optimistic. Because two pairs of groups share identical mixture designs, and three groups have nearly identical ash (CaO  $\approx$  25.8 wt%), the analysis was repeated with those groups merged (13 and 11 groups).

Optimizer benchmark (D2). A real-coded GA (population 40, tournament selection, blend crossover  $\alpha = 0.5$ , Gaussian mutation  $p = 0.2$ , two elites), a PSO (40 particles, inertia 0.9→0.4,  $c_1 = c_2 = 1.5$ ) and the hybrid (25 GA generations seeding 25 PSO iterations) each received 2000 evaluations of the re-trained surrogate [21] over 30 seeds with common random numbers, so the GA and the GA stage of the hybrid coincide up to the switch. A run succeeds if it ends within 0.1 MPa of the best value found in all 90 runs.

Identifiability (D3). Each surrogate was profiled along the partition (OPC 0–15 wt%, step 0.25) with the other inputs at the original optimum (SH = 12 mol/L, SS/SH = 2.23, A/B = 0.48, S/B = 1.2), and the location of the maximum was recorded for 200 record-level bootstrap refits [22] and for the 15 source-held-out refits. Record-level resampling ignores within-source clustering and likely understates between-source variability, hence the source-held-out complement.

Coverage (D4). Inputs were z-standardized with the database mean and standard deviation, and each verification mixture was assigned the mean distance to its five nearest database records [23],

$$d(x) = (1/k) \sum_{j=1}^k \|z(x) - z(x_j)\|, \quad k = 5 \quad (1)$$

using a NaN-aware Euclidean distance in which missing SH values are omitted and the distance is rescaled by the square root of the ratio of total to observed coordinates [24]. Reference distances were computed for each database record excluding itself, and excluding its whole source.

#### 2.4. *Experimental verification*

Six mortars traced the partition from OPC15–SF0 to OPC0–SF15 in 3 wt% steps (fly ash 85 wt% of binder; activator and aggregate at the optimum values), sealed-cured at  $25 \pm 2$  °C and tested at 28 days on prism halves. Companion pastes of identical binder were examined at 28 days by thermogravimetry (to 800 °C at 10 °C/min under N<sub>2</sub>) and mercury intrusion porosimetry.

### 3. Results and discussion

#### 3.1. *Model development and source-held-out validation*

In the original cross-validation, XGBoost gave the lowest RMSE in every fold (figure 1(a);  $4.03 \pm 0.97$  MPa, against  $4.70 \pm 0.83$  MPa for LightGBM and  $7.85 \pm 1.34$  MPa for SVR) and the **best random-split test score** ( $R^2 = 0.951$ , RMSE = 3.07 MPa). **These results support model selection under the random-split protocol only; the relative performance of the five learners on an unseen source remains untested.**

In 65 refits of the re-trained surrogate (50 record-level bootstrap, 15 source-held-out), SHAP placed OPC first among compositional inputs in median rank (overall rank 2–6), whereas SF ranked 5th–10th, so the order below OPC is unstable. **OPC's higher importance is consistent with calcium-assisted reaction and the formation of calcium-bearing binding gels under ambient curing [4,5]. SF supplies reactive silica and can contribute to particle packing, but its effect depends on calcium availability and activator conditions [6]. Uneven coverage also matters: SF occurs in 96 records from three sources, and combined OPC–SF mixtures in only 27 records from one source (section 3.4). Resampling or omitting a source therefore changes the evidence available for SF. Correlated ash oxides and co-varying activator inputs can also carry overlapping predictive information; refits may allocate their contributions differently, shifting the order below OPC. These are plausible explanations for the rank variability, not mechanisms established by this analysis. SHAP explains fitted predictions [18]; a lower or unstable SF rank does not establish a negligible physical role, and the ranks should not be interpreted as causal effects or a universal ordering of material importance.**

Holding out whole sources changes the picture (table 1, figure 1(b)). For the re-trained surrogate, the pooled RMSE rises from 3.65 MPa under random cross-validation to 11.6 MPa (+218 %), and  $R^2$  falls from 0.94 to 0.42; with the merged groupings it is 11.5 and 11.1 MPa, so the result does not depend on the exact source definition. The per-source RMSE has a median of 10.2 MPa (IQR 5.6–15.8 MPa), and the mean per-source bias spans –26.9 to +16.3 MPa, indicating systematic offsets rather than scatter. CaO content alone does not explain them: three sources with nearly identical ash ( $\text{CaO} \approx 25.8$  wt%) are each predicted with 14.6–18.5 MPa error when held out. Unencoded attributes such as ash fineness, glass content or mixing and curing practice may contribute [25] but were not tested.

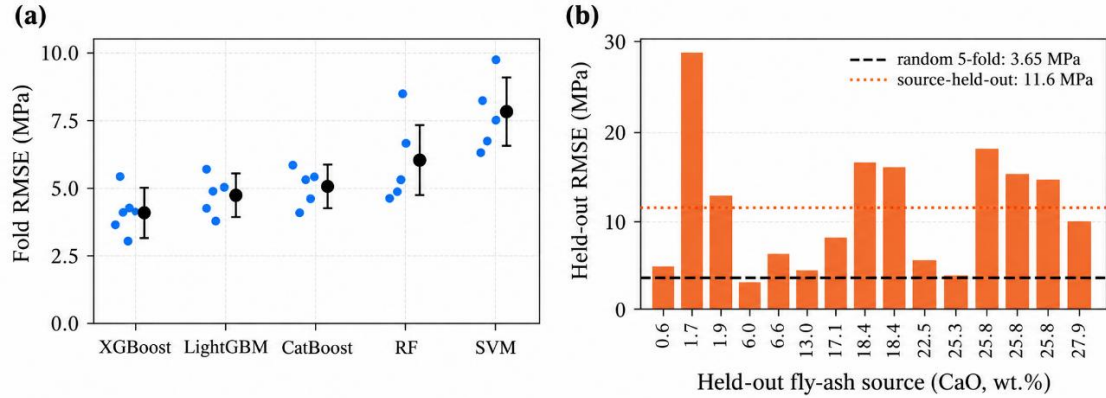


**Figure 1.** (a) Fold-level RMSE of the five learners in the original five-fold cross-validation; dots are folds and black markers the mean  $\pm$  SD. (b) RMSE of the re-trained surrogate for each fly-ash source when that source is held out, ordered by the CaO content of the ash; the horizontal lines mark the random five-fold and the pooled source-held-out RMSE.

**Table 1.** Surrogate accuracy under different validation protocols. Values after  $\pm$  are standard deviations across folds or repeats;  $\Delta$ RMSE is the change relative to the random five-fold RMSE of the same surrogate.

| Protocol                    | Surrogate  | R <sup>2</sup>    | RMSE (MPa)      | MAE (MPa)       | $\Delta$ RMSE (%) |
|-----------------------------|------------|-------------------|-----------------|-----------------|-------------------|
| Random 8:2, test set        | Original   | 0.951             | 3.07            | 2.25            | –                 |
| Random five-fold (training) | Original   | 0.928 $\pm$ 0.030 | 4.03 $\pm$ 0.97 | 2.79 $\pm$ 0.57 | reference         |
| Random 8:2, test set        | Re-trained | 0.966             | 2.88            | 2.18            | –                 |
| Random five-fold, 5 repeats | Re-trained | 0.943 $\pm$ 0.006 | 3.65 $\pm$ 0.18 | 2.42 $\pm$ 0.13 | reference         |
| Leave-one-source-out (15)   | Re-trained | 0.42              | 11.62           | 8.88            | +218              |
| Six verification mixtures   | Original   | $\rho = 0.94^a$   | 7.81            | 6.61            | +94               |

<sup>a</sup> Spearman rank correlation between predicted and measured means of six mixtures; R<sup>2</sup> is not reported because six mixture means do not constitute a conventional test set.

### 3.2. Optimizer benchmark and surrogate uncertainty

Under the tested budget, GA and the hybrid were both reliable, succeeding in 93 % and 90 % of runs (figure 2, table 2). PSO approached the high-fitness region fastest (99.5 % of the reference after a median of 198 evaluations, against 305 for GA) but mostly stalled 0.3 % below it, and only 30 % of its runs succeeded. Because the hybrid shares its first 1000 evaluations with the GA, the comparison isolates its PSO stage, which produced no measurable improvement. These results are specific to this configuration and budget.

Optimizer-induced variation was small: mean final values of the three algorithms differed by 0.4 MPa, and the run-to-run standard deviation was 0.05–1.03 MPa. Surrogate-induced variation was larger. All successful runs placed the binder in the OPC-rich part of the partition (OPC 12.0–15.0 wt%; GA 13.9  $\pm$  0.8 wt%), in line with the original optimum, and at the original design point the two surrogates predict 50.5 and 52.4 MPa. Their optima elsewhere differ, however: the re-trained surrogate moves A/B near its lower bound (0.46), SH to about 13 mol/L and S/B to 1.64 and predicts 63.9 MPa, against A/B = 0.48, SH = 12 mol/L, S/B = 1.2 and 50.5 MPa for the original surrogate. At a fixed composition, the 5–95 % band of the bootstrap refits spans about  $\pm 6$  MPa. Precise convergence of the optimizer should therefore not be confused with certainty of the surrogate.

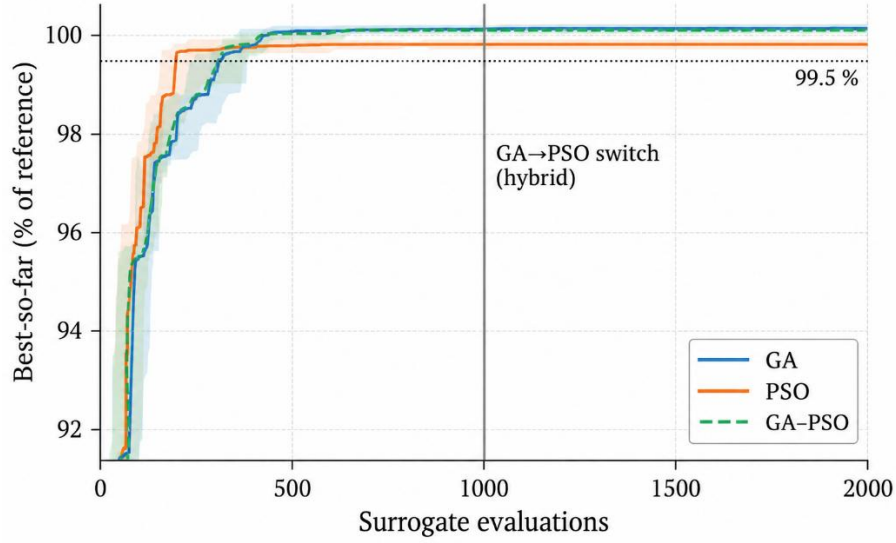


**Figure 2.** Best-so-far fitness against surrogate evaluations for GA, PSO and the GA–PSO hybrid, as a percentage of the reference optimum (median and interquartile band of 30 seeded runs). The hybrid shares its first 1000 evaluations with the GA through common random numbers; the dotted line marks 99.5 %.

**Table 2.** Equal-budget optimizer benchmark (2000 evaluations, 30 seeds). Fitness is a percentage of the reference optimum; evaluation counts are medians [interquartile range] over the runs that reached each threshold.

| Algorithm | Final fitness (%) | Success (%) | Evals to success | Evals to 99.5 %            | OPC returned (wt%) |
|-----------|-------------------|-------------|------------------|----------------------------|--------------------|
| GA        | 99.98 ± 0.08      | 93.3        | 420 [356–531]    | 305 [238–388]              | 13.9 ± 0.8         |
| PSO       | 99.30 ± 1.62      | 30.0        | 500 [394–605]    | 198 [146–272] <sup>b</sup> | 13.3 ± 2.3         |
| GA–PSO    | 99.97 ± 0.09      | 90.0        | 420 [354–514]    | 305 [238–388]              | 13.8 ± 0.8         |

<sup>b</sup> 27 of 30 runs reached 99.5 %. Worst final fitness: GA 99.70 %, PSO 93.63 %, GA–PSO 99.70 %. Success denotes ending within 0.1 MPa of the reference.

### 3.3. *Identifiability of the optimum along the OPC/SF partition*

The surrogates are nearly flat over the OPC-rich half of the partition (figure 3(a)). The original surrogate predicts 50.3, 50.3 and 50.2 MPa for OPC15–SF0, OPC12–SF3 and OPC9–SF6, a spread of 0.12 MPa, or 4 % of its test RMSE; it cannot rank these three compositions.

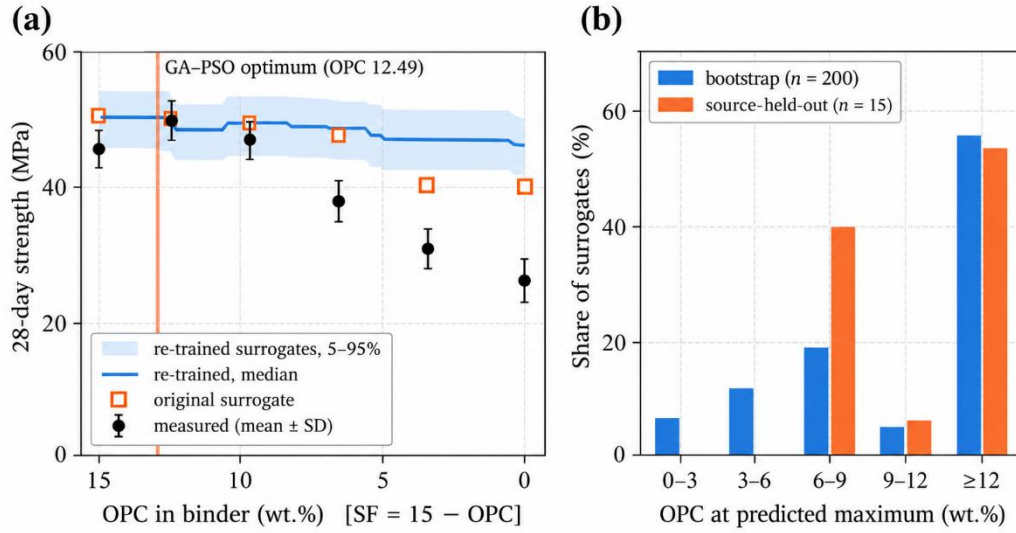


**Figure 3.** (a) Predicted 28-day strength along the OPC/SF partition at the optimized activator and aggregate settings: median and 5–95 % band of 200 record-level bootstrap refits of the re-trained surrogate, original-surrogate predictions at the six tested compositions, and measured means  $\pm$  SD. (b) Location of the predicted maximum for the bootstrap and source-held-out refits.

The maximum is consistently located in the OPC-rich portion but not at a single point (figure 3(b)): 55 % of bootstrap refits place it at OPC 12–15 wt% and 80 % at OPC  $\geq$  6 wt%, while 53 % of source-held-out refits place it at OPC 12–15 wt% and none below 6 wt%. The computational optimum is therefore identifiable primarily as a region. The experiments narrowed that region without reducing it to a point: OPC12–SF3 was significantly stronger than OPC15–SF0 and than the mixtures with OPC  $\leq$  6 wt%, but not distinguishable from OPC9–SF6, which in turn did not differ from OPC15–SF0 (table 3). The proximity of OPC12.49–SF2.51 to the strongest tested mixture is thus region-level agreement.

### 3.4. Experimental verification and coverage

Against the measurements (table 3), the original surrogate preserved the rank order largely (Spearman  $\rho = 0.94$ ,  $p = 0.005$ ) and was within 1.0 % at OPC12–SF3 ( $49.8 \pm 2.15$  MPa), but its over-prediction grew from +1.0 % to +50.5 % as SF replaced OPC. The six-mixture RMSE, 7.81 MPa, is 94 % above the original surrogate's random five-fold RMSE (4.03 MPa); being based on six mixture means, it indicates the size of the transfer error rather than a conventional test-set score.

**Table 3.** Measured and predicted 28-day compressive strength of the verification mixtures and the coverage distance  $d$  of each (equation (1)). Mixtures not sharing a Tukey letter differ at the 5 % level.

| Mixture   | Measured (MPa)  | Tukey | Original prediction (MPa) | Bias (%) | $d$ (–) |
|-----------|-----------------|-------|---------------------------|----------|---------|
| OPC15–SF0 | $45.5 \pm 1.71$ | b     | 50.3                      | +10.5    | 2.21    |
| OPC12–SF3 | $49.8 \pm 2.15$ | a     | 50.3                      | +1.0     | 2.11    |
| OPC9–SF6  | $46.6 \pm 1.46$ | ab    | 50.2                      | +7.7     | 1.94    |
| OPC6–SF9  | $38.3 \pm 2.55$ | c     | 47.2                      | +23.4    | 1.78    |
| OPC3–SF12 | $31.3 \pm 2.52$ | d     | 39.9                      | +27.5    | 1.69    |
| OPC0–SF15 | $26.3 \pm 1.38$ | e     | 39.6                      | +50.5    | 1.71    |

Global distance alone would not have flagged this bias. All six mixtures lie at  $d = 1.69$ – $2.21$ , beyond the 95th percentile of the within-database reference (1.32) and below the median of the out-of-source reference (2.37), yet  $d$  decreases toward the SF-rich end, where the bias is largest ( $r = -0.78$ ,  $n = 6$ ). The nearest neighbors of the three SF-richest mixtures contain neither SF nor OPC: SF occurs in 96 records

from 3 sources, and OPC together with SF in 27 records (8.2 %) from one source with a markedly different activator (SS/SH = 6.2, A/B = 0.74). The SF-rich half of the partition is thus predicted largely from SF-free neighbors, and coverage should be assessed for the variables being optimized as well as for the full input vector.

Paste descriptors that are not surrogate inputs also varied with composition: gel-related bound water (105–400 °C) was 3.40, 4.01 and 3.17 % and the fine-pore fraction (< 0.05 µm) 14.8, 48.3 and 24.3 % at OPC15–SF0, OPC9–SF6 and OPC0–SF15, both peaking one step on the SF-rich side of the strength maximum. Such changes in reaction products [26] are encoded only indirectly by mixture proportions.

#### 4. Implications for low-clinker binders in ocean structures

Literature-trained optimization should therefore report, besides a random-split score, a source-held-out error for transfer to a new site, supplier or laboratory; the optimum as a near-optimal region with its resampling distribution, spanned by the verification series; an equal-budget optimizer comparison on the surrogate in use, with the surrogate-induced spread; and coverage for the constituents being optimized.

The strongest tested mixture (12 wt% OPC in the binder) reached about 50 MPa at 28 days without heat curing. Compressive strength alone is insufficient to establish suitability for marine exposure; chloride ingress, sulfate and magnesium attack, wetting–drying, abrasion and other durability properties were not investigated. Limitations also include the use of a re-trained surrogate for the diagnostics, **the absence of source-held-out comparisons with the other four learners, source identification by ash composition**, a single ash, activator and curing regime, and one configuration per optimizer.

#### 5. Conclusions

- **For the re-trained XGBoost surrogate, random-split validation substantially underestimated source-transfer error:** holding out whole sources raised the RMSE from 3.65 to 11.6 MPa (+218 %).
- Under an equal budget, GA and the GA–PSO hybrid were both reliable (93 % and 90 % success) and PSO less so (30 %); the PSO stage of the hybrid added no measurable benefit, and optimizer-induced variation was small relative to surrogate-induced variation.
- The optimum is identifiable as an OPC-rich region rather than a precise point; the experiments narrowed the region, but OPC12–SF3 and OPC9–SF6 remained statistically indistinguishable.
- Prediction bias increased to +51 % in the SF-rich region, where constituent-specific training coverage was sparse, although global input-space distance did not indicate it.
- Robust ML-based mixture optimization therefore requires source-aware validation, optimizer benchmarking, region-level reporting of optima and constituent-specific coverage assessment.

#### Acknowledgements

This research was supported by the AI Seoul Tech Research Support Program of the Seoul Future Foundation.

#### References

- [1] Benhelal E, Zahedi G, Shamsaei E and Bahadori A 2013 Global strategies and potentials to curb CO<sub>2</sub> emissions in cement industry *J. Clean. Prod.* **51** 142–61
- [2] Duxson P, Fernández-Jiménez A, Provis J L, Lukey G C, Palomo A and van Deventer J S J 2007 Geopolymer technology: the current state of the art *J. Mater. Sci.* **42** 2917–33
- [3] Provis J L 2018 Alkali-activated materials *Cem. Concr. Res.* **114** 40–8
- [4] Nath P and Sarker P K 2015 Use of OPC to improve setting and early strength properties of low calcium fly ash geopolymer concrete cured at room temperature *Cem. Concr. Compos.* **55** 205–14
- [5] Mehta A and Siddique R 2017 Properties of low-calcium fly ash based geopolymer concrete incorporating OPC as partial replacement of fly ash *Constr. Build. Mater.* **150** 792–807

- [6] Wu H, He M, Cheng J, Wang T, Che Y and Du Y 2024 Study on the synergistic effects of OPC and silica fume on the mechanical and microstructural properties of geopolymer mortar *Constr. Build. Mater.* **434** 136740
- [7] Nguyen K T, Nguyen Q D, Le T A, Shin J and Lee K 2020 Analyzing the compressive strength of green fly ash based geopolymer concrete using experiment and machine learning approaches *Constr. Build. Mater.* **247** 118581
- [8] Peng Y and Unluer C 2022 Analyzing the mechanical performance of fly ash-based geopolymer concrete with different machine learning techniques *Constr. Build. Mater.* **316** 125785
- [9] Rathnayaka M, Karunasinghe D, Gunasekara C, Wijesundara K, Lokuge W and Law D W 2024 Machine learning approaches to predict compressive strength of fly ash-based geopolymer concrete: a comprehensive review *Constr. Build. Mater.* **419** 135519
- [10] Dong W, Huang Y, Cui A and Ma G 2023 Mix design optimization for fly ash-based geopolymer with mechanical, environmental, and economic objectives using soft computing technology *J. Build. Eng.* **72** 106577
- [11] Ji H, Lyu Y, Ying W, Liu J C and Ye H 2024 Machine learning guided iterative mix design of geopolymer concrete *J. Build. Eng.* **91** 109710
- [12] Sharma N, Seema, Paruthi S and Tipu R K 2025 Interpretable GA–PSO–optimised deep learning for multi-objective geopolymer concrete strength prediction *Asian J. Civ. Eng.* **26** 4679–706
- [13] Kaufman S, Rosset S, Perlich C and Stitelman O 2012 Leakage in data mining: formulation, detection, and avoidance *ACM Trans. Knowl. Discov. Data* **6** 15
- [14] Roberts D R et al 2017 Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure *Ecography* **40** 913–29
- [15] Breiman L 2001 Statistical modeling: the two cultures (with comments and a rejoinder by the author) *Stat. Sci.* **16** 199–215
- [16] Chen T and Guestrin C 2016 XGBoost: a scalable tree boosting system Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining pp 785–94
- [17] Akiba T, Sano S, Yanase T, Ohta T and Koyama M 2019 Optuna: a next-generation hyperparameter optimization framework Proc. 25th ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining pp 2623–31
- [18] Lundberg S M and Lee S-I 2017 A unified approach to interpreting model predictions *Adv. Neural Inf. Process. Syst.* **30** 4765–74
- [19] Ahmed H U, Mohammed A A, Rafiq S, Mohammed A S, Mosavi A, Sor N H and Qaidi S M A 2021 Compressive strength of sustainable geopolymer concrete composites: a state-of-the-art review *Sustainability* **13** 13502
- [20] Zailani W W A, Apandi N M, Adesina A, Alengaram U J, Faris M A and Tahir M F M 2024 Physico-mechanical properties of geopolymer mortars for repair applications: impact of binder to sand ratio *Constr. Build. Mater.* **412** 134721
- [21] Kennedy J and Eberhart R 1995 Particle swarm optimization Proc. ICNN'95 Int. Conf. Neural Networks vol **4** pp 1942–8
- [22] Efron B and Tibshirani R J 1993 An Introduction to the Bootstrap (New York: Chapman & Hall)
- [23] Meyer H and Pebesma E 2021 Predicting into unknown space? Estimating the area of applicability of spatial prediction models *Methods Ecol. Evol.* **12** 1620–33
- [24] Pedregosa F et al 2011 Scikit-learn: machine learning in Python *J. Mach. Learn. Res.* **12** 2825–30
- [25] Zhang Z, Provis J L, Zou J, Reid A and Wang H 2016 Toward an indexing approach to evaluate fly ashes for geopolymer manufacture *Cem. Concr. Res.* **85** 163–73
- [26] Garcia-Lodeiro I, Palomo A, Fernández-Jiménez A and Macphée D E 2011 Compatibility studies between N-A-S-H and C-A-S-H gels. Study in the ternary diagram Na<sub>2</sub>O–CaO–Al<sub>2</sub>O<sub>3</sub>–SiO<sub>2</sub>–H<sub>2</sub>O *Cem. Concr. Res.* **41** 923–31