Explainable Concrete Strength Prediction with Amortized Gaussian Process

Yifei Duan [1,2], Kai Gong [3], Kristen A. Severson [4], Jie Chen [5], Jeremy R. Gregory [6], Elsa A. Olivetti [2]

[1]Institute for Data, Systems, and Society, Massachusetts Institute of Technology, Cambridge, MA, USA
[2]Department of Materials Science and Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA
[3]Department of Civil and Environmental Engineering, Rice University, Houston, TX, USA
[4]Microsoft Research, Cambridge, MA, USA
[5]MIT-IBM Watson AI Lab, Cambridge, MA, USA
[6]MIT Climate and Sustainability Consortium, Massachusetts Institute of Technology, Cambridge, MA, USA
duanyf99@mit.edu, kg51@rice.edu, kseverson@microsoft.com, chenjie@us.ibm.com, jgregory@mit.edu, elsao@mit.edu

Abstract

Concrete production accounts for about 8-9% of global anthropogenic CO_2 emissions, and decarbonizing the concrete industry constitutes a major component of global decarbonization efforts. Developing machine learning (ML) models for concrete property prediction has emerged as a promising strategy to optimize concrete mix designs, aiming to minimize CO_2 emissions while meeting engineering requirements. However, many existing efforts face limitations due to the lack of scientific consistency and validation on industrial datasets, hindering the explainability and applicability of resulting ML models. To address this limitation, this work proposes to incorporate domain knowledge into a novel amortized Gaussian process (AGP) architecture to predict concrete strength of highly complex industry-adopted concrete mixes. This domain-informed AGP model not only achieves high predictive performance (with R^2 values of 0.884) but also significantly enhances model explainability, capturing a wide range of intricate details of domain knowledge that other models (including random forest (RF), multi-layer perceptron (MLP), and standard GP) fail to grasp. This study highlights the power of domain-informed AGP model in concrete strength prediction, paving the way for reliable mix design of more sustainable concrete.

Introduction

Concrete industry contributes an estimated 8-9% of global anthropogenic CO_2 emissions (Ellis et al. 2020), making it one of the largest industrial sources of greenhouse gas emissions. The majority of concrete emissions arise from the production of ordinary Portland cement (OPC), the most widely used binders in concrete. Using supplementary cementitious materials (SCMs, mostly industrial byproducts), e.g., slag, fly ash, and silica fume, to partially replace OPC is one of the most promising strategies to decarbonize concrete industry (IEA 2018). However, the use of SCMs adds complexity to the mix design of concrete, which is already a highly complex composite material. To enable concrete mix design for better performance and sustainability, it is necessary to develop reliable models for predicting important concrete properties, like strength, based on mix information.

XAI4Sci: Explainable machine learning for sciences, AAAI-24 (xai4sci.github.io)

Past experimental studies have revealed some semi-quantitative relations between strength and time, and certain mix features including the water-to-cementitious (W/CM) ratio (Yeh 2006) (Chidiac, Moutassem, and Mahmoodzadeh 2013). However, these simple empirical functions fail to account for the intricate connections among a diverse set of influential factors, resulting in poor predictions for mix designs beyond the experimental data used for developing them. In contrast, modern machine learning (ML) methods excel at capturing complex inter-correlations among numerous features, showing promise in predicting concrete properties. Although past ML studies have made valuable contributions to this area, they mostly rely on small lab datasets (generally < 1000 data points) (Young et al. 2019) (Chaabene, Flah, and Nehdi 2020) (DeRousseau et al. 2019) (Nunez et al. 2021) (Li et al. 2022), and lack the incorporation of domain knowledge and yield suboptimal performance ($R^2 <$ 0.6) on industrial datasets. These limitations impede the explainability and applicability of the resulting ML models, thereby hampering the confidence of the inherently conservative concrete industry in these black-box ML models.

To leverage ML for the sustainable transformation of the concrete industry, the challenge of model explainability needs attention. In this study, we explored the use of domain-informed amortized Gaussian process (AGP) models to enhance the explainability of ML-based concrete strength prediction, using a large and highly complex industrial dataset. We focus this work on the modeling of strength development instead of other concrete properties, because strength is the most important and commonly reported engineering property for concrete mixes (DeRousseau et al. 2019).

Methods

Amortized Gaussian Process (AGP)

To capture the complex relationship between mixture design and compressive strength, as well as to generate predictive uncertainty and allow for non-uniform time sampling, we select Gaussian process (GP) (Rasmussen and Williams 2006) as the model class for our application. Instead of concatenating time and mixture constituents as a single input vector for the GP, we propose the use of amortized Gaussian processes (AGP). The AGP, described below, allows for the parameterization of a function which describes strength as a function of time. This univariate relationship can be informed by domain knowledge, as relationships between compressive strength and time are well studied.

The AGP is similar to a standard GP in that it describes distributions over functions and is fully specified by mean and covariance functions, $m(\mathbf{x})$ and $k(\mathbf{x}, \mathbf{x}')$, respectively. The primary difference is that the AGP introduces functions which output the hyperparameters of the GP, thus amortizing the typical cost of learning. Each compressive strength trajectory is modeled as $y(t) = f(t) + \epsilon(t)$ where y is compressive strength, f is a GP, ϵ is observation noise, and t is time. The specific mean and covariance functions are

$$m(t; \theta) = \theta_1 \log t + \theta_2$$

$$k(t, t'; \theta) = \theta_3 \exp\left(-\frac{(\log t - \log t')^2}{2(\theta_4)^2}\right)$$

$$\theta_i = \mathrm{MLP}_{\phi_{\mathrm{i}}}(\mathbf{z}) \quad \forall \mathrm{i}$$

where $\mathbf{z}$ is a vector of concrete constituent amounts, θ_i are the GP hyperparameters, and ϕ_i are the parameters of the i-th MLP. The choice of a log-linear mean function is motivated by domain knowledge: this empirical relation between strength and time for a given sample has been commonly used by the concrete research community (Chidiac, Moutassem, and Mahmoodzadeh 2013). The parameters, ϕ, are learned using the maximum likelihood objective applied to a population of concrete samples which is enabled by automatic differentiation packages.

AGP stands out as a suitable architecture for strength modeling primarily owing to the seamless alignment between the time-series nature of strength evolution and the inherent advantage of GP in time series prediction (Roberts et al. 2013). Furthermore, the availability of empirical functions for strength-time relation allows incorporation of domain knowledge.

We compared the performance of the domain-informed AGP model against other machine learning models, including MLP, RF, standard GP, as well as AGP without the domain-informed mean functions.

Experimental Setup

Amortized Gaussian Process (AGP): The model is trained with a learning rate of 0.001, a batch size of 732 (10% of the training data), and 100 training epochs. For all parameterization MLPs, we used a 2-layer fully connected MLP with a hidden size 10. A softplus layer is added for θ_1 to enforce positive strength development over time.

Random Forest (RF): Grid search based on cross validation is applied to select optimal hyperparameters. The model with best performance has 1000 regression trees.

Multi-layer Perceptron (MLP): Grid search based on cross validation is applied. The model with best performance has 1 hidden layer of size 500, trained with a learning rate of 0.001 with 200 epochs.

Standard Gaussian Process (GP): An RBF kernel with white noise is used. No other hyperparameters are required.

Dataset

We perform all modelling using an industrial dataset from a concrete producer, comprising information on 9151 unique mixes (with a total of 38019 strength measurements at different time points). A total of 19 mix design variables are used as input features $\mathbf{z}$. The unique mixes are randomly split into 80% training data and 20% testing data.

Model Explainability

We analyzed and compared model explainability through (1) feature importance analysis; (2) sensitivity analysis

- We applied SHAP analysis (Lundberg and Lee 2017) to examine the feature importance of the θ_2 MLP in AGP, which corresponds to the 1-day mean strength. In assessing model explainability, we focus on evaluating whether the developed ML models can identify crucial features known to impact concrete strength.
- Sensitivity analysis was performed for several important features by varying its value while keeping the others constant. This allows us to scrutinize whether the ML-predicted trends align with domain knowledge.

Results and Discussion

Model performance: The performance of the various models in predicting concrete strength are compared in Table 1, which shows that the RF model and the AGP with a domain-informed mean function (as detailed above) demonstrate superior performance, exhibiting the highest R^2 values (0.884-0.895), along with the lowest RMSE (924-881 psi) and MAPE (9.6-9.9%), followed by the MLP and standard GP. All these models are seen to significantly outperform the baseline AGP model with a simple linear mean function. The comparison among the GP models demonstrates that despite GP's task-agnostic nature, a formulation that incorporates domain knowledge can significantly improve its predictive performance. The potential of AGP as a predictive model within a specific domain can be fully unlocked by using formulations that align with domain understandings, leveraging its inherent flexibility in function specification.

Feature importance analysis: To assess the ML models' ability in capturing the impact of individual features on strength prediction, we have performed SHAP feature importance analysis, with the results for the early-age strength (1-day) shown in Figure 1. Notably, the SHAP analyis for the standard GP model demands significantly more computation time ($\times 50+$) than any other model, which presents a major drawback for practical application. Standard GP is excluded from SHAP analysis due to this reason together with its inferior performance (see Table 1). A comparison of Figure 1 with domain knowledge reveals that the domain-informed AGP generally outperforms RF and MLP in capturing the impact of individual features. For example, the AGP results reveal the large positive impact of cement quantities (i.e., cement I&II and cement II) and total HRWR (high-range water-reducing admixture), and the negative impact of W/CM (water-to-cementitious material) ratio, total AEA (air-entraining admixture), total fly ash and total slag

Table 1: Comparison of the model performance in predicting concrete strength (based on the testing set), as assessed by R^2, root mean square error (RMSE) and mean absolute percentage error (MAPE).

Model	$\mathbf{R^2}$	**RMSE, psi**	**MAPE, %**
AGP (domain-informed mean function)	0.884	924	9.9
Basline AGP (simple linear mean function)	0.668	1566	16.1
Random Forest (RF)	0.895	881	9.6
Multi-layer Perceptron (MLP)	0.877	952	9.9
Standard Gaussian Process (GP)	0.866	993	10.5

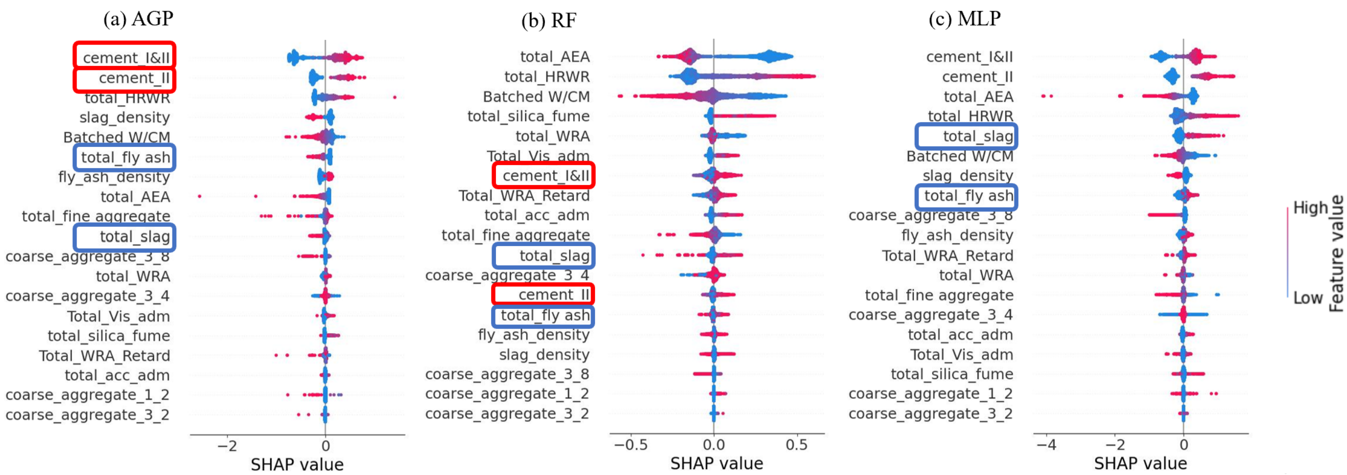

Figure 1: SHAP feature importance analysis for the (a) domain-informed Amortized Gaussian Process (AGP), (b) Random Forest (RF), and (c) Multi-layer Perceptron (MLP) for 1-day concrete strength, based on the 20% testing set. AGP shows better scientific consistency than RF and MLP, demonstrated by the highlighted features.

content on early-age strength. The scientific mechanisms behind these are well-established. For example,the large positive impact of cement content on strength is due to its role as the main reactive component binding sand and rock together to form concrete (Iyer 2020). Fly ash and slag are both less reactive than cement, therefore elevating their content results in diminished early-age strength (Poon, Lam, and Wong 1999) (Skibsted and Snellings 2019).

In contrast to AGP, the RF and GP models fail to demonstrate scientific consistency through feature importance. For RF, features known to be crucial, cement quantity and total fly ash and slag content, are ranked at moderate to lower positions on the list, and it's hard to discern from the distributions whether total fly ash and slag have a positive or negative impact. For MLP, the total slag and fly ash content are observed to positively influence 1-day strength, which contradicts established domain knowledge. Overall, the SHAP analysis suggests that the domain-informed AGP outperforms RF and MLP models in scientific consistency.

Local SHAP analysis with the AGP model is visualized for two distinctive data points representing low-strength mixes (Figure 2 (a)) and high-strength mixes (Figure 2 (b)) respectively, for local inference. Compared to the global SHAP results shown in Figure 1, the negative impact of high W/CM on early age strength is notably pronounced for the low-strength mix, with the absence of cement_II and high range water reducing admixture (HRWR). For high-strength mixes, on the other hand, the results highlight the importance of high cement content and HRWR usage along with low W/CM. Consistent with the distinctive features of different SCMs, we also see from high-ranking important features that the addition of silica fume improves strength significantly while high slag content contributes negatively. The chemical mechanisms underlying these effects are further elaborated in the following sensitivity analysis.

Sensitivity analysis: To further evaluate the explainability of the domain-informed AGP model, we have performed sensitivity analysis on several important features, namely the percentage of cement replaced by fly ash, slag, and silica fume, along with the W/CM ratio. The results are presented in Figure 2, which shows how changes in these mix design features impact the compressive strength of concrete at different ages (3-day and 28-day) while keeping the other features constant. Note that the modelling data set has a maximum replacement ratio of 40%, 60% and 10% for fly ash, slag and silica fume, respectively, meaning that predictions beyond those values are extrapolations. A comparison of the results with domain knowledge clearly reveals that the the domain-informed AGP model outperforms the RF, MLP and standard GP models. In the case of fly ash, the AGP model show that increasing its content in replacement of OPC leads to a decrease in early-age strength (3-day), aligning with expectation as fly ash is less reactive than OPC (Poon, Lam, and Wong 1999). At 28 days, the AGP model

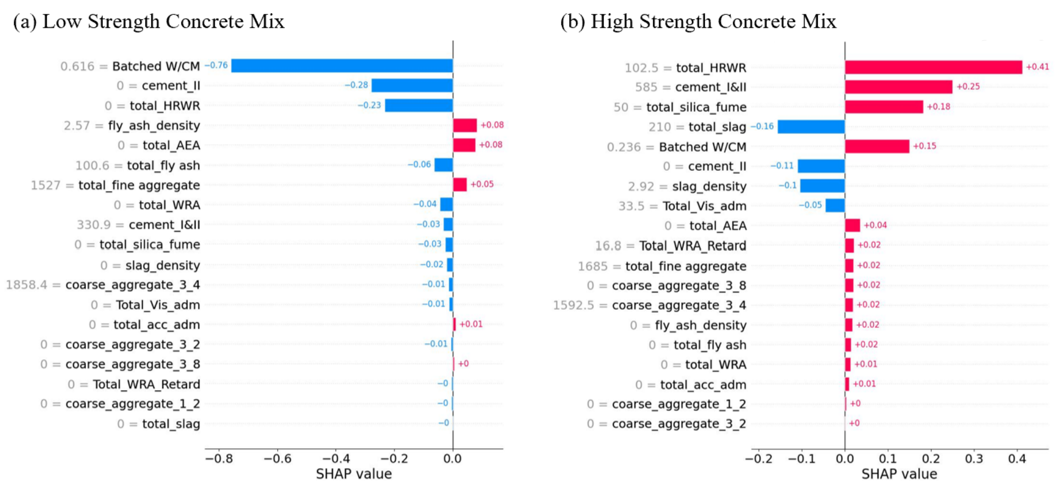

Figure 2: Local SHAP bar plots with AGP model for representative (a) low strength and (b) high strength concrete mixes.

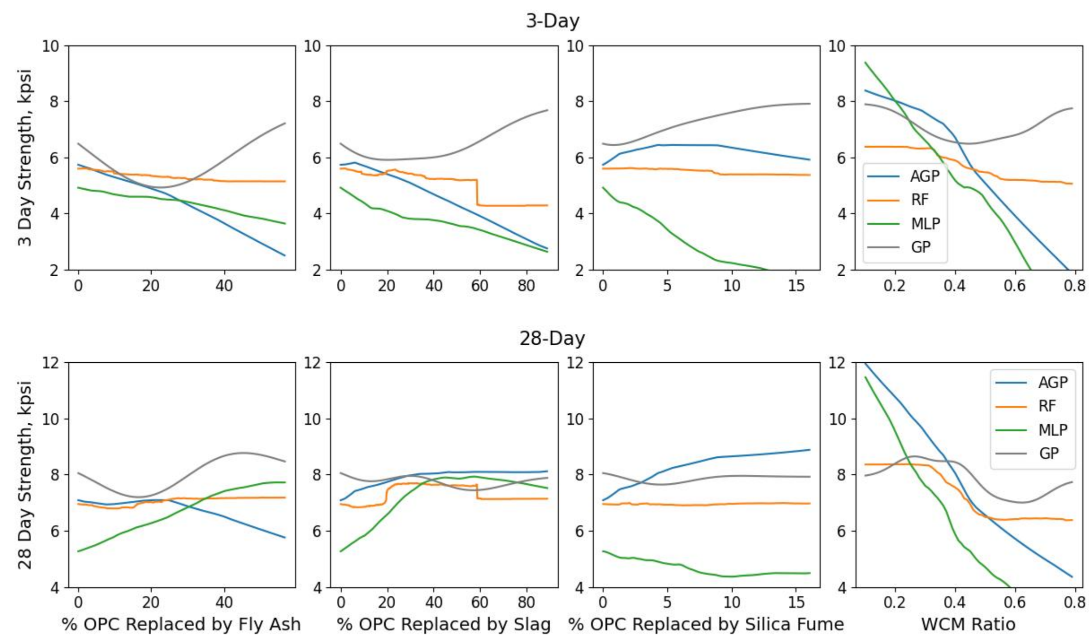

Figure 3: Sensitivity analysis for AGP, RF, MLP, and GP: impact of replacing Ordinary Portland Cement (OPC) by fly ash, slag, and silica fume, and W/CM ratio on the 3-day and 28-day strengths. AGP results align the best with domain knowledge.

reveals no strength decrease at moderate replacement ratios (e.g., $<\sim 25\%$), consistent with domain knowledge. There is a slow reaction mechanism between fly ash and portlandite ($Ca(OH)_2$), a reaction product of cement hydration, forming the strength-giving calcium-silicate hydrate (C-S-H) gel (Snellings, Mertens, and Elsen 2012). This allows the strength of the cement-fly ash blends to gradually catch up in the long term (Bendapudi and Saha 2011). However, when the content of fly ash is further increased, the 28-day strength decreases, because reducing cement content beyond a certain range leads to insufficient portlandite available to react with fly ash and form the strength-giving C-S-H gel. While the AGP model adeptly captures the trends, the other three models largely fail. Notably, the standard GP model fails to capture the negative impact of fly ash on 3-day strength at high replacement ratios above 20%. For the 28-day strength, RF, MLP and standard GP suggest that high replacement ratio above $\sim 30\%$ leads to similar or higher 28-day strength than pure OPC, in contrast to domain knowledge.

Slag replacement impacts concrete strength through similar reactions to fly ash, except that slag is more reactive than fly ash and consumes less portlandite due to its higher Ca content. As a result, a higher replacement ratio can be achieved for slag without compromising long-term strength (Skibsted and Snellings 2019). AGP shows that slag content is inversely correlated with strength at 3-day and positively

Table 2: Log-likelihood from training & testing predictions (based on probability densities).

Model	Training	Testing
AGP (domain-informed)	2.176	2.173
RF	3.378	1.525

correlated at 28-day, which aligns with domain knowledge as more C-S-H gel is formed over time. MLP captures similar overall trends, while standard GP and RF fail to do so.

Furthermore, silica fume, commonly used to improve concrete strength (Juneger and Siddique 2015) (Juneger, Snellings, and Bernal 2019), is known for its beneficial impact by strengthening the weak interface between rock aggregate and cement paste, especially at later ages (Nežerka et al. 2019). This positive impact on strength is well captured by the AGP model, but not by the other models.

In addition, only the AGP and MLP models capture the well-known inverse correlation between W/CM ratio and strength (Mehta and Monteiro 2014). RF and standard GP notably fail to extrapolate for W/CM beyond 0.3-0.5.

Overall, the sensitivity analysis demonstrates AGP's superior capability in capturing trends consistent with domain knowledge across different features. This is especially noteworthy given that the right ends of all the plots (and left end for W/CM) are model extrapolations. In contrast, the RF model, despite having the highest prediction accuracy (Table 1), fails to capture much of the domain knowledge and presents lack of smoothness or extrabolability. Despite the cross-validation performed for the RF, it inherently favors fitting to the data sample instead of capturing overall trends. In our specific context, where the data distributions of training and testing sets are close to each other, the testing performance of RF remains unaffected. However, as feature values deviate further from the majority of training data, the discontinuity poses challenges for RF to make reasonable extrapolation. This issue highlights a significant advantage of AGP for downstream uses, namely the inverse mix design for creating more sustainable concrete. AGP is well-suited to handle situations where optimized mixes are expected to lie far outside the feature value ranges of current industry-adopted mixes.

Uncertainty quantification: Finally, we compared the prediction uncertainty of the AGP and RF models. Prediction uncertainty of AGP for any input can be calculated through a distribution directly specified by the posterior mean and variance. Meanwhile, likelihood for RF is based on fitting a Gaussian distribution, where the mean and standard deviation are calculated from predictions of all regression trees. As shown in Table 2, AGP predictions on the test set has much higher likelihood than RF. Notably, AGP maintain a similar level of likelihood for both training and test sets, while RF suffers a drastic drop. This renders AGP a potentially more calibrated and reliable option than RF for predictions in domains like concrete science.

Conclusion

Development of accurate and explainable machine learning (ML) models for concrete property prediction is crucial for leveraging ML for concrete decarbonization. This study demonstrates that the architecture of amortized Gaussian process (AGP) allows the incorporation of domain-informed empirical equations. By incorporating a domain-informed mean function, the resulting AGP model is shown to exhibit greater performance than the baseline AGP (based on simple linear mean function), standard GP and multi-layer perceptron (MLP), second only to random forest (RF) by a small margin. Moreover, AGP provides a win-win outcome of performance and explainability, while RF fails in scientific consistency. SHAP analysis shows that the domain-informed AGP successfully identifies important features known to impact concrete strength (either positively or negatively), while the RF and MLP models fall short. Sensitivity analysis shows that the domain-informed AGP correctly captures a range of intricate details on how different supplementary cementitious materials (SCMs), and water-cementitious (W/CM) ratio influence concrete strength at both 3 days and 28 days. AGP also shows to extrapolate reasonably beyond the feature value range in the original dataset. In contrast, all the other models fall short in capturing most details. Hence, this study highlights the potential of using domain-informed AGP for enhanced explainability in ML-based concrete strength prediction for complex mixes, thereby paving the way for its future adoption by the concrete industry in the sustainable transition.

Limitations and Future Work

Despite our efforts to account for potential issues, we acknowledge the following limitations in the current work:

- Our AI models and optimization frameworks are based on concrete strength data, without considering other important engineering properties.
- The local inference for individual data points remains to be fully analyzed in further in-depth studies.
- The potential advantages of empolying more explainable models in downstream applications remain to be fully demonstrated.

For future works, we anticipate expanding the scope of our analysis by incorporating other concrete properties, including slump and air content, into both predictive model and explainability analysis. These models will be instrumental in addressing property-related constraints in inverse concrete mix design. The ability of the models to capture smoothness and extrapolate is anticipated to greatly enhance the reliability of optimized mixes. As the mixes optimized to reduce climate impacts are expected to be substantially different from the majority of current industrial data, we also plan to scrutinize the local feature importance of optimized mixes to ensure scientific consistency in our findings.

References

Bendapudi, S. C.; and Saha, P. 2011. Contribution of Fly ash to the properties of Mortar and Concrete. *International Journal of Earth Sciences and Engineering*, 4: 1017–1023.

Chaabene, W. B.; Flah, M.; and Nehdi, M. L. 2020. Machine learning prediction of mechanical properties of concrete: Critical review. *Construction and Building Materials*, 260: 119889.

Chidiac, S. E.; Moutassem, F.; and Mahmoodzadeh, F. 2013. Compressive strength model for concrete. *Magazine of Concrete Research*, 65(9): 557–572.

DeRousseau, M.; Laftchiev, E.; Kasprzyk, J. R.; Balaji, R.; and Srubar III, W. V. 2019. A Comparison of Machine Learning Methods for Predicting the Compressive Strength of Field-Placed Concrete. *Construction and Building Materials*, 228: 116661.

Ellis, L. D.; Badel, A. F.; Chiang, M. L.; and Chiang, Y.-M. 2020. Toward electrochemical synthesis of cement—An electrolyzer-based process for decarbonating CaCO3 while producing useful gas streams. *Proceedings of the National Academy of Sciences*, 117: 12584–12591.

IEA. 2018. Technology Roadmap: Low-Carbon Transition in the Cement Industry. https://www.iea.org/reports/technology-roadmap-low-carbon-transition-in-the-cement-industry. Accessed: 2023-09-20.

Iyer, N. R. 2020. An overview of cementitious construction materials. In Samui, P.; Kim, D.; Iyer, N. R.; and Chaudhary, S., eds., *New Materials in Civil Engineering*, 1–64. Oxford, UK: Butterworth-Heinemann.

Juneger, M. C.; and Siddique, R. 2015. Recent advances in understanding the role of supplementary cementitious materials in concrete. *Cement and Concrete Research*, 78: 71–80.

Juneger, M. C.; Snellings, R.; and Bernal, S. A. 2019. Supplementary cementitious materials: New sources, characterization, and performance insights. *Cement and Concrete Research*, 122: 257–273.

Li, Z.; Yoon, J.; Zhang, R.; Rajabipour, F.; III, W. V. S.; Dabo, I.; and Radlinska, A. 2022. Machine learning in concrete science: applications, challenges, and best practices. *npj Computational Materials*, 8: 1–17.

Lundberg, S. M.; and Lee, S.-I. 2017. A Unified Approach to Interpreting Model Predictions. In *31st Conference on Neural Information Processing Systems (NIPS 2017)*.

Mehta, P. K.; and Monteiro, P. J. 2014. *Concrete: microstructure, properties, and materials*. McGraw-Hill Education.

Nežerka, V.; Bílý, P.; Hrbek, V.; and Fládr, J. 2019. Impact of silica fume, fly ash, and metakaolin on the thickness and strength of the ITZ in concrete. *Cement and Concrete Composites*, 103: 256–262.

Nunez, I.; Marani, A.; Flah, M.; and Nehdi, M. L. 2021. Estimating compressive strength of modern concrete mixtures using computational intelligence: A systematic review. *Construction and Building Materials*, 310: 125279.

Poon, C. S.; Lam, L.; and Wong, Y. L. 1999. Effects of Fly Ash and Silica Fume on Interfacial Porosity of Concrete. *Journal of Materials in Civil Engineering*, 11(3): 197–205.

Rasmussen, C. E.; and Williams, C. K. 2006. *Gaussian Process for Machine Learning*. Cambridge, Mass.: MIT Press.

Roberts, S.; Osborne, M.; Ebden, M.; Reece, S.; Gibson, N.; and Aigrain, S. 2013. Gaussian process for time-series modelling. *Philosophical Transaction of Royal Society A*, 371: 20110550.

Skibsted, J.; and Snellings, R. 2019. Reactivity of supplementary cementitious materials (SCMs) in cement blends. *Cement and Concrete Research*, 124: 105799.

Snellings, R.; Mertens, G.; and Elsen, J. 2012. Supplementary cementitious materials. *Reviews in Mineralogy and Geochemistry*, 74(1): 211–278.

Yeh, I.-C. 2006. Generalization of strength versus water–cementitious ratio relationship to age. *Cement and Concrete Research*, 36(10): 1865–1873.

Young, B. A.; Hall, A.; Pilon, L.; Gupta, P.; and Sant, G. 2019. Can the compressive strength of concrete be estimated from knowledge of the mixture proportions?: New insights from statistical analysis and machine learning methods. *Cement and Concrete Research*, 115: 379–388.