Prediction of Compressive Strength of UHPC Mix Designs Based on Mix Components Using Machine Learning Techniques

Md Abdullah Al Sarfin – PhD student, Utah State University, Department of Civil and Environmental Engineering, Logan, UT, USA, Email: abdullah.alsarfin@usu.edu

Srishti Banerji – Assistant Professor, Utah State University, Department of Civil and Environmental Engineering, Logan, UT, USA, Email: srishti.banerji@usu.edu

Mohsen Zaker Esteghamati – Assistant Professor, Utah State University, Department of Civil and Environmental Engineering, Logan, UT, USA, Email: mohsen.zaker@usu.edu

Andrew D. Sorensen (corresponding author) – Associate Professor, Utah State University, Department of Civil and Environmental Engineering, Logan, UT, USA, Email: andrew.sorensen@usu.edu

Abstract

The high cost and limited availability of commercial ultra-high performance concrete (UHPC) products have motivated researchers and several state departments of transportation to develop non-proprietary UHPC mixes using locally available materials. Several non-proprietary mixes following different mixing and curing regimens have been developed in different geographic locations across the United States. Although these existing mixes can be utilized as starting points for developing suitable UHPC mix design for a project, fine-tuning localized mix designs is still a cost, labor, and time intensive task. A limited number of studies in the existing literature have utilized machine learning techniques to predict strength of inputted UHPC mixes. However, practitioners typically determine the concrete mix design after selecting the target performance criteria such as compressive strength, and workability for the project. To address this need, this study is a first step towards utilizing machine learning techniques for reverse-engineering UHPC mix proportions based on selected performance criteria and incorporating readily available materials. A database of 215 UHPC mixes extracted from 24 published manuscripts is compiled. This database is used to train ensemble learning algorithms such as Random Forest, Support Vector Machine (SVM), and Gradient Boosting. A comparative assessment on the most promising ML models trained with selected mix design parameters is presented. The tool developed in this study will be beneficial for practitioners to establish a locally sourced non-proprietary mix design of UHPC based on the required attributes.

Keywords: Non-proprietary mix design, Machine learning, Ultra-high performance concrete.

1. Introduction

Ultra-high performance concrete (UHPC) is a class of advanced cementitious material with superior strength and durability (Piérard et al.; Kodur et al.; Banerji et al.; Banerji and Kodur). UHPC is also known to have excellent performance under high strain rates (Thomas and Sorensen; Al Sarfin), seismic loading (Chao et al.), fatigue loading (Jia et al.). The achievement of such superior properties lies on the careful sampling and proportioning of the mixture variables, low water-to-binder ratio, effective utilization of the pozzolanic behavior of supplementary cementitious materials (SCM), and optimum packing density (de Larrard and Sedran). Current literatures show that a wide range and variety of SCM, fillers, and fibers are employed to achieve the desired properties of the material. For instance, silica fume, fly ash, slag, limestone powder, and metakaolin are some of the common pozzolanic supplements. With so many possible combinations of materials available, designing a UHPC based on locally available or economically optimized materials is of benefit to designers. Utilizing alternate data-driven methods can effectively reduce the time and associated cost in developing and optimizing non-proprietary UHPC mixes.

In recent years, researchers have been exploring the application of data-driven techniques in structural and material engineering with promising results (Mahjoubi, Barhemat, et al.; Esteghamati and Flint; Banerji). Ghafari et al. trained an artificial neural network (ANN) with 15 neurons in the hidden layers and reported improved prediction of compressive strength and slump flow compared to a statistical mix design (Ghafari, Bandarabadi, et al.). García developed a four-layer multi-layer-perceptron model for predicting compressive strength of UHPC by training the model with experimental and collected data (Abellán-García). Saleh et al. proposed a Gaussian process modeling with batch Bayesian optimization framework to infer the mixture design of UHPC (Saleh et al.). Abuodeh et al. employed Sequential Feature Selection and Neural Interpretation Diagram to identify the critical constituents by training the ANN model with a database of 110 compressive strength tests (Abuodeh et al.). Mahjoubi et al. noted the challenges for wider acceptance of machine learning methods for development and application of UHPC such as limited scope of existing machine learning models due to limited number of variables, low generalization performance and overfitting issues due to small datasets, inadequate information on the input variables, and limited studies on hyper-parameter optimization (Mahjoubi, Meng, et al.).

This study aims to contribute to the existing knowledge by creating a comprehensive database of UHPC mixtures and utilizing an integrated framework to train and tune machine learning methods for identifying the critical mixture components that predict key performance properties of UHPC using recursive feature elimination (RFE) which has not been used in other studies involving training machine-learning models for UHPC mixes.

2. Methodology

The following subsection describes the methodology to collect and preprocess the dataset, as well as for training and tuning the developed machine learning models.

2.1. Data Collection and Preparation

A dataset is assembled by collecting data from publicly available published research works (Habel et al.; Wille et al.; Allena and Newtson; Wang et al.; Hassan et al.; Graybeal; Yoo et al.; Yu et al.; Ghafari, Costa, et al.; Rangaraju and Li; Meng et al.; Soliman and Tagnit-Hamou; Khaloo et al.; Alsalman et al.; Meng and Khayat; Wu, Shi, et al.; Sadrmomtazi et al.; Wu, Khayat, et al.; Arora et al.; Mo et al.; Nguyen et al.; Shafei; Banerji and Kodur). From 24 research papers, 215 mix proportions are collected. The mixture parameters are: cement (C), water (W), silica fume (SF), fly ash (FA), limestone powder (L), slag (GBBS), metakaolin (M), quartz flour (QF), nano-silica (nS), river or regular concrete sand (S), micro-sand (MS), masonry sand (MnS), light-weight sand (LWS), quartz sand (QS), nano-$CaCO_3$ (NC), glass powder (GP), coarse aggregates (CA), water reducing admixtures (SP), and curing regime (Cure).

The data is extracted with GRABIT in graphical format, a program developed with MATLAB (Doke). Missing information is filled in by a regression-based imputation, as necessary (Young et al.). The data are normalized using z-score normalization (Wan; Raju et al.; Mahjoubi, Meng, et al.). The complied dataset is investigated for partial correlation between the features and multi-collinearity by calculating Pearson correlation coefficient, R, and variance inflation factor (VIF) (Alin).

2.2. Model Selection

Initially, seven machine-learning models, with default hyper-parameters, are trained by splitting the dataset of 215 mixes into 75% of the data for training and 25% of the data for testing. The machine learning models analyzed are least square linear regression, ridge regression, lasso regression, Support Vector machine with linear kernel, Random Forest, Decision Tree, and Gradient Boosting. Based on the performance of the models, Support Vector Machine, Random Forest, and Gradient Boosting are chosen for further tuning.

2.3. Model Optimization and Validation

Recursive Feature Elimination is employed to reduce model complexity and overfitting, to save computation cost, and to improve interpretability (Li et al.; Liu and Motoda). In this method, a

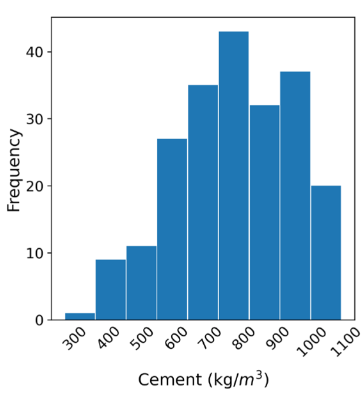


Figure 1: Frequency distribution of cement. (1 kg/m³ = 0.0624 lbs./ft³)

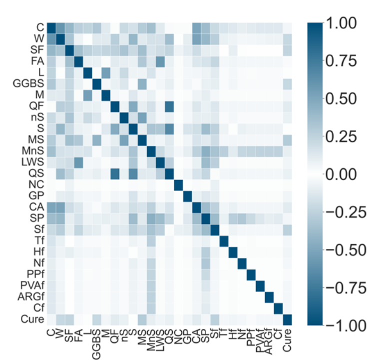


Figure 2: Correlation matrix of the features.

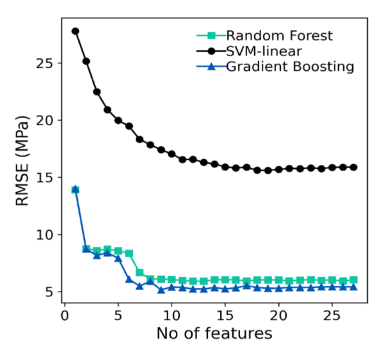


Figure 3: Feature selection through recursive feature elimination (1 MPa = 0.145 ksi)

machine learning algorithm is trained with the initial set of features and less important features are eliminated recursively based on a ranking criterion until the optimum set of features is achieved.

The selected models are optimized by hyper-parameter tuning using randomized search and grid search. Grid search performs an exhaustive search for the optimum hyper-parameters over a specified range. One-fourth of the data are held for testing the optimized models. Furthermore, the actual and predicted compressive strengths obtained for the test dataset are compared to validate the accuracy of the final models.

3. Results and Discussion

The frequency distributions of all the features are plotted to investigate the distribution of the features and to identify any potential irregularities as shown for cement in Figure 1.

Table 1: Selected features for three machine learning models

Model	Selected Features
Random Forest	C, W, SF, FA, L, QF, S, MS, MnS, GP, SP, Sf, Cure
Support Vector Machine	C, W, SF, FA, GGBS, QF, nS, MS, MnS, LWS, GP, Sf, Hf, Cure
Gradient Boosting	C, W, SF, FA, QF, S, MS, MnS, GP, SP, Sf, Cure

The correlation matrix of the normalized features is shown in Figure 2. The maximum correlation coefficient is 0.8 for quartz flour and quartz sand. The remaining correlation coefficients are below 0.7. The mean VIF is 9.08. A VIF value smaller than 10 is indicative of no multicollinearity problem in the dataset (Alin). Therefore, the compiled dataset does not show high collinearity and is used to train machine learning models without further processing.

Table 2: Performance metrics of untuned and tuned models (1 MPa = 0.145 ksi)

		Training Scores		Testing Scores	
		Before Tuning	After Tuning	Before Tuning	After Tuning
Random Forest	RMSE (MPa)	6.03	3.99	10.38	10.46
	R^2	0.96	0.98	0.82	0.82
Support Vector Machine	RMSE (MPa)	15.89	6.72	17.64	8.61
	R^2	0.7	0.95	0.49	0.88
Gradient Boosting	RMSE (MPa)	5.42	6.34	11.15	9.04
	R^2	0.97	0.95	0.8	0.87

Selection of features through the recursive feature elimination procedure is shown in Figure 3. The selected features for three machine learning models are shown in Table 1. The number of selected features for Random Forest, Support Vector machine, and Gradient Boosting are 13, 14, and 12, respectively. Among the supplementary cementitious materials (SCM), silica fume and fly ash are revealed to be the most important component of the mixtures as these two are selected by the recursive feature elimination procedure for all three models. Additionally, glass powder, a filler material, is another component selected for all the machine learning models. At least one type of sand (S, MS, MnS, and LWS), straight steel fibers, and curing is found to be other important mix components to predict the compressive strength of UHPC mixes.

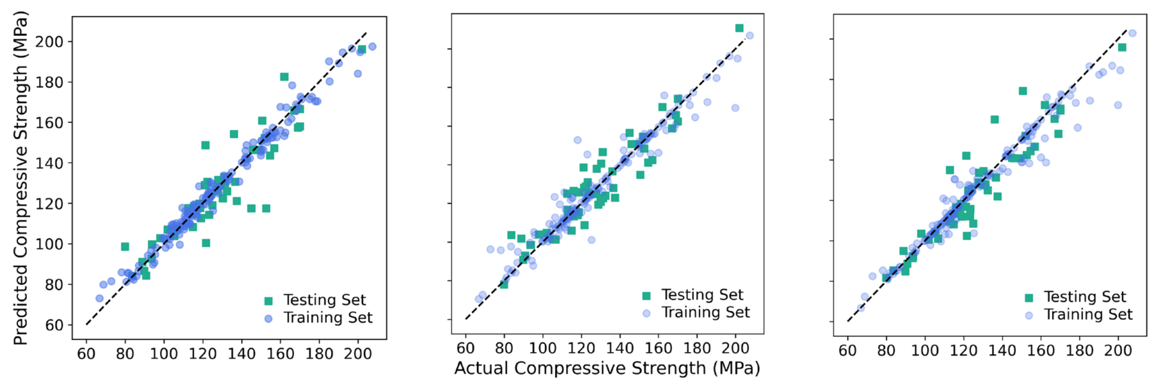
The models are retrained with the selected set of features and the optimum hyper-parameters are obtained through randomized search and grid search. The final models are evaluated by performance metrics. The performance metrics for tuned and untuned models are compared in Table 2. The performance metrics of random forest and gradient boosting methods for training set before tuning shows high accuracy, while the testing scores shows relatively low accuracy. After tuning, significant improvement in performance scores of testing dataset for Support Vector Machine and Gradient Boosting methods are observed.

For support vector machine, the scores for the training and testing sets from untuned models are less accurate than the other two models. But after tuning, the training and testing $RMSE$ are 6.67 MPa and 11.32 MPa, respectively, and the R^2 for training and testing are 0.94 and 0.85. Finalized hyperparameter are presented in Table 3.

Table 3: Finalized hyper-parameters

Random Forest	Support Vector Machine	Gradient Boosting
No of estimators = 397 Minimum samples splits = 5 Minimum samples leaf = 1 maximum features = $\sqrt{no\ of\ features}$ bootstrap = False	C = 112 Epsilon = 0.491 Kernel = Radial basis function Gamma = 0.135	Loss= Quantile regression Learning rate = 0.2 No of estimators = 300 Minimum samples splits = 10 Minimum samples leaf = 2 Alpha = 0.45

The actual and predicted compressive strengths for the models are presented in Figure 5. The plots demonstrate that all the models perform with reasonable accuracy. Overall, the Support Vector Machine and Gradient Boosting models performs better in terms of predicting the compressive strength from the testing set.



(a) (b) (c)

Figure 4: Predicted and actual compressive strengths (a) Random Forest, (b) Support Vector Machine, and (c) Gradient Boosting. The dashed line represents zero-error (1 MPa = 0.145 ksi)

4. Conclusions

By training machine learning models with the collected dataset, following conclusions are made:

- The developed dataset of mixture designs does not show any multi-collinearity issue as the mean VIF is 9.08.
- Silica fume and fly ash are found to be the most important SCMs to predict the compressive strength of the UHPC mixes. Curing and straight steel fibers are other mix components selected by the recursive feature elimination procedure for all three models.
- In terms of fitting the training and testing sets, the Support Vector Machine (R^2 for training and testing set are 0.95 and 0.88, respectively) and Gradient Boosting Machine (R^2 for

training and testing set are 0.95 and 0.87, respectively) model performs the best to predict compressive strength of the mixtures.

As such, it can be concluded that machine-learning methods are effective and robust tools that can be used to develop designer friendly and convenient mixture designs tools. This work can be further developed to reverse-engineer the machine-learning model to predict mix component quantities based on specified performance parameters.

5. References

Abellán-García, Joaquín. "Four-Layer Perceptron Approach for Strength Prediction of UHPC." *Construction and Building Materials*, vol. 256, Elsevier, 2020, p. 119465.

Abuodeh, Omar R., et al. "Assessment of Compressive Strength of Ultra-High Performance Concrete Using Deep Machine Learning Techniques." *Applied Soft Computing*, vol. 95, Elsevier, 2020, p. 106552.

Al Sarfin, Md Abdullah. *Parametric Study of Mixture Component Contributions to Compressive Strength and Impact Energy Absorption Capacity of a High Strength Cementitious Mix with No Coarse Aggregate*. Utah State University, 2019.

Alin, Aylin. "Multicollinearity." *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 2, no. 3, Wiley Online Library, 2010, pp. 370–74.

Allena, Srinivas, and Craig M. Newtson. "Ultra-High Strength Concrete Mixtures Using Local Materials." *Journal of Civil Engineering and Architecture*, vol. 5, no. 4, David Publishing Company, 1840 Industrial Drive Suite 160 Libertyville …, 2011, pp. 322–30.

Alsalman, Ali, et al. "Development of Ultra-High Performance Concrete with Locally Available Materials." *Construction and Building Materials*, vol. 133, Elsevier, 2017, pp. 135–45.

Arora, Aashay, et al. "Material Design of Economical Ultra-High Performance Concrete (UHPC) and Evaluation of Their Properties." *Cement and Concrete Composites*, vol. 104, Elsevier, 2019, p. 103346.

Banerji, Srishti, et al. "Experimental Behavior of Ultra High Performance Fiber Reinforced Concrete Beams under Fire Conditions." *Engineering Structures*, vol. 208, Elsevier, 2020, p. 110316.

---. "Using Artificial Intelligence to Derive Temperature Dependent Mechanical Properties of Ultra-High Performance Concrete." *Leveraging Artificial Intelligence in Engineering, Management, and Safety of Infrastructure*, CRC Press, 2022, pp. 403–22.

Banerji, Srishti, and Venkatesh Kodur. "Effect of Temperature on Mechanical Properties of Ultra-high Performance Concrete." *Fire and Materials*, vol. 46, no. 1, Wiley Online Library, 2022, pp. 287–301.

Chao, Shih-Ho, et al. "Seismic Behavior of Ultra-High-Performance Fiber-Reinforced Concrete Moment Frame Members." *International Interactive Symposium on Ultra-High Performance Concrete*, vol. 1, no. 1, Iowa State University Digital Press, 2016.

de Larrard, François, and Thierry Sedran. "Optimization of Ultra-High-Performance Concrete by the Use of a Packing Model." *Cement and Concrete Research*, vol. 24, no. 6, Elsevier, 1994, pp. 997–1009.

Doke, Jero. "GRABIT." *MATLAB Central File Exchange*, 2023, https://www.mathworks.com/matlabcentral/fileexchange/7173-grabit.

Esteghamati, Mohsen Zaker, and Madeleine M. Flint. "Developing Data-Driven Surrogate Models for Holistic Performance-Based Assessment of Mid-Rise RC Frame Buildings at Early Design." *Engineering Structures*, vol. 245, Elsevier, 2021, p. 112971.

Ghafari, Ehsan, Mojtaba Bandarabadi, et al. "Prediction of Fresh and Hardened State Properties of UHPC: Comparative Study of Statistical Mixture Design and an Artificial Neural Network Model." *Journal of Materials in Civil Engineering*, vol. 27, no. 11, American Society of Civil Engineers, 2015, p. 4015017.

Ghafari, Ehsan, Hugo Costa, et al. "Statistical Mixture Design Approach for Eco-Efficient UHPC." *Cement and Concrete Composites*, vol. 55, Elsevier, 2015, pp. 17–25.

Graybeal, Benjamin A. *Development of Non-Proprietary Ultra-High Performance Concrete for Use in the Highway Bridge Sector: TechBrief*. United States. Federal Highway Administration, 2013.

Habel, Katrin, et al. "Ultra-High Performance Fibre Reinforced Concrete Mix Design in Central Canada." *Canadian Journal of Civil Engineering*, vol. 35, no. 2, 2008, pp. 217–24.

Hassan, A. M. T., et al. "Experimental Test Methods to Determine the Uniaxial Tensile and Compressive Behaviour of Ultra High Performance Fibre Reinforced Concrete (UHPFRC)." *Construction and Building Materials*, vol. 37, Elsevier, 2012, pp. 874–82.

Jia, Li, et al. "Fatigue Behavior of UHPC Beams Prestressed with External CFRP Tendons." *Journal of Composites for Construction*, vol. 26, no. 5, American Society of Civil Engineers, 2022, p. 4022066.

Khaloo, A. R., et al. "A New Mixture Design Method for Ultra-High-Strength Concrete." *ACI Materials Journal*, vol. 114, no. 2, 2017.

Kodur, Venkatesh, et al. "Effect of Temperature on Thermal Properties of Ultrahigh-Performance Concrete." *Journal of Materials in Civil Engineering*, vol. 32, no. 8, American Society of Civil Engineers, 2020, p. 4020210.

Li, Jundong, et al. "Feature Selection: A Data Perspective." *ACM Computing Surveys (CSUR)*, vol. 50, no. 6, ACM New York, NY, USA, 2017, pp. 1–45.

Liu, Huan, and Hiroshi Motoda. *Feature Selection for Knowledge Discovery and Data Mining*. Vol. 454, Springer Science & Business Media, 2012.

Mahjoubi, Soroush, Rojyar Barhemat, et al. "AI-Guided Auto-Discovery of Low-Carbon Cost-Effective Ultra-High Performance Concrete (UHPC)." *Resources, Conservation and Recycling*, vol. 189, Elsevier, 2023, p. 106741.

Mahjoubi, Soroush, Weina Meng, et al. "Auto-Tune Learning Framework for Prediction of Flowability, Mechanical Properties, and Porosity of Ultra-High-Performance Concrete (UHPC)." *Applied Soft Computing*, vol. 115, Elsevier, 2022, p. 108182.

Meng, Weina, et al. "Optimization and Performance of Cost-Effective Ultra-High Performance Concrete." *Materials and Structures*, vol. 50, Springer, 2017, pp. 1–16.

Meng, Weina, and Kamal H. Khayat. "Effect of Hybrid Fibers on Fresh Properties, Mechanical Properties, and Autogenous Shrinkage of Cost-Effective UHPC." *Journal of Materials in Civil Engineering*, vol. 30, no. 4, American Society of Civil Engineers, 2018, p. 4018030.

Mo, Zongyun, et al. "Hydration and Mechanical Properties of UHPC Matrix Containing Limestone and Different Levels of Metakaolin." *Construction and Building Materials*, vol. 256, Elsevier, 2020, p. 119454.

Nguyen, Tan-Trac, et al. "Optimised Mix Design and Elastic Modulus Prediction of Ultra-High Strength Concrete." *Construction and Building Materials*, vol. 302, Elsevier, 2021, p. 124150.

Piérard, Julie, et al. "Durability Evaluation of Different Types of UHPC." *Proceedings of the RILEM-Fib-AFGC International Symposium on Ultra-High Performance Fiber-Reinforced Concrete, Marseille, France*, 2013, pp. 1–3.

Raju, V. N. Ganapathi, et al. "Study the Influence of Normalization/Transformation Process on the Accuracy of Supervised Classification." *2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT)*, IEEE, 2020, pp. 729–35.

Rangaraju, Prasad Rao, and Zhengqi Li. "Development of UHPC Using Ternary Blends of Ultra-Fine Class F Fly Ash, Meta-Kaolin and Portland Cement." *International Interactive Symposium on Ultra-High Performance Concrete*, vol. 1, no. 1, Iowa State University Digital Press, 2016.

Sadrmomtazi, Ali, et al. "Effect of Materials Proportion on Rheology and Mechanical Strength and Microstructure of Ultra-High Performance Concrete (UHPC)." *Construction and Building Materials*, vol. 187, Elsevier, 2018, pp. 1103–12.

Saleh, Eman, et al. "You Only Design Once (YODO): Gaussian Process-Batch Bayesian Optimization Framework for Mixture Design of Ultra High Performance Concrete." *Construction and Building Materials*, vol. 330, Elsevier, 2022, p. 127270.

Shafei, Behrouz. "Development of Non-Proprietary Ultra-High Performance Concrete (UHPC) for Iowa Bridges." *The Bridge*, vol. 515, 2021, pp. 294–8103.

Soliman, Nancy A., and Arezki Tagnit-Hamou. "Using Particle Packing and Statistical Approach to Optimize Eco-Efficient Ultra-High-Performance Concrete." *ACI Materials Journal*, vol. 114, no. 6, 2017.

Thomas, R. J., and Andrew D. Sorensen. "Review of Strain Rate Effects for UHPC in Tension." *Construction and Building Materials*, vol. 153, Elsevier, 2017, pp. 846–56.

Wan, Xing. "Influence of Feature Scaling on Convergence of Gradient Iterative Algorithm." *Journal of Physics: Conference Series*, vol. 1213, no. 3, IOP Publishing, 2019, p. 32021.

Wang, Chong, et al. "Preparation of Ultra-High Performance Concrete with Common Technology and Materials." *Cement and Concrete Composites*, vol. 34, no. 4, Elsevier, 2012, pp. 538–44.

Wille, Kay, et al. "Ultra-High Performance Concrete with Compressive Strength Exceeding 150 MPa (22 Ksi): A Simpler Way." *ACI Materials Journal*, vol. 108, no. 1, 2011.

Wu, Zemei, Kamal H. Khayat, et al. "Changes in Rheology and Mechanical Properties of Ultra-High Performance Concrete with Silica Fume Content." *Cement and Concrete Research*, vol. 123, Elsevier, 2019, p. 105786.

Wu, Zemei, Caijun Shi, et al. "Effect of SCM and Nano-Particles on Static and Dynamic Mechanical Properties of UHPC." *Construction and Building Materials*, vol. 182, Elsevier, 2018, pp. 118–25.

Yoo, Doo-Yeol, et al. "Effect of Fiber Content on Mechanical and Fracture Properties of Ultra High Performance Fiber Reinforced Cementitious Composites." *Composite Structures*, vol. 106, Elsevier, 2013, pp. 742–53.

Young, William, et al. "A Survey of Methodologies for the Treatment of Missing Values within Datasets: Limitations and Benefits." *Theoretical Issues in Ergonomics Science*, vol. 12, no. 1, Taylor & Francis, 2011, pp. 15–43.

Yu, R., et al. "Development of an Eco-Friendly Ultra-High Performance Concrete (UHPC) with Efficient Cement and Mineral Admixtures Uses." *Cement and Concrete Composites*, vol. 55, Elsevier, 2015, pp. 383–94.