Advancing Lightweight Engineered Cementitious Composites: An Interpretable Machine Learning Framework

Name of Presenter: Md Nasir Uddin
Degree and program (of presenter): PhD in MSEC
Co-Authors names: Dr. Xijun Shi
Funders if any: None

TEXAS STATE UNIVERSITY

TEXAS STATE THE GRADUATE COLLEGE

Purpose

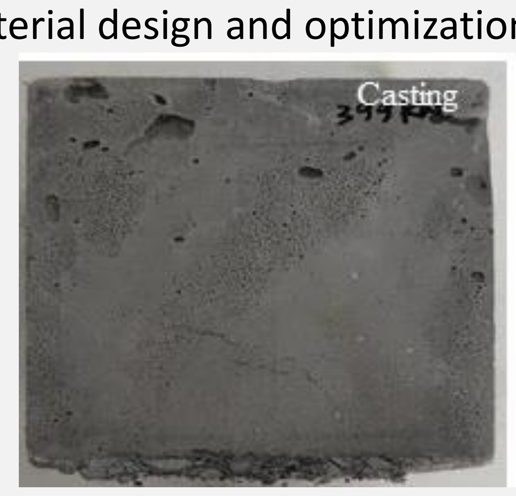
This abstract outlines a research study focused on enhancing the design and utilization of Lightweight Engineered Cementitious Composites (LWECCs) in civil engineering through the application of Machine Learning (ML) techniques. The primary purposes of this research are:

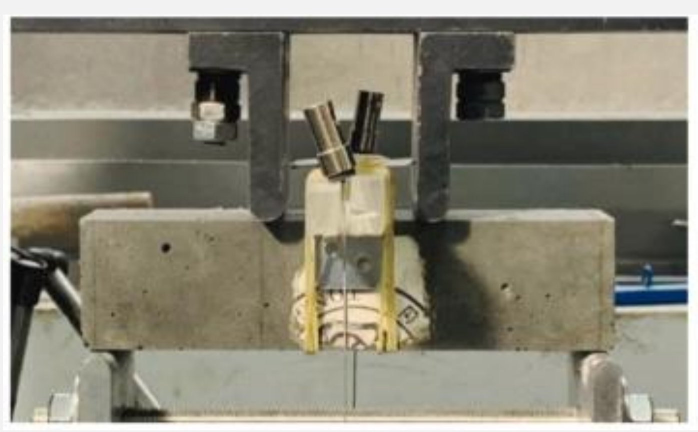
Predictive Modeling of Material Properties: Developing predictive models using advanced ML algorithms like eXtreme Gradient Boosting (XGBoost), Light Gradient Boosting Machine (LightGBM) and gene expression programming (GEP) to accurately forecast the Compressive Strength (CS) and Flexural Strength (FS) of LWECCs. These strength parameters are crucial for assessing the material's suitability for various construction applications.

Optimization of Mixture Designs: By collecting data on mixture design components and their resulting strengths from existing literature, and analyzing this data through ML models, the research aims to identify the most effective combinations of materials, specifically focusing on LWECCs reinforced with fiber

Insight into Material Properties through SHAP Analysis: Utilizing SHapley Additive exPlanations (SHAP) analysis to understand how different mixture components influence the predictive models' outcomes. This can provide valuable insights into how various ingredients in the composites contribute to their overall strength, informing better material design and engineering practices.

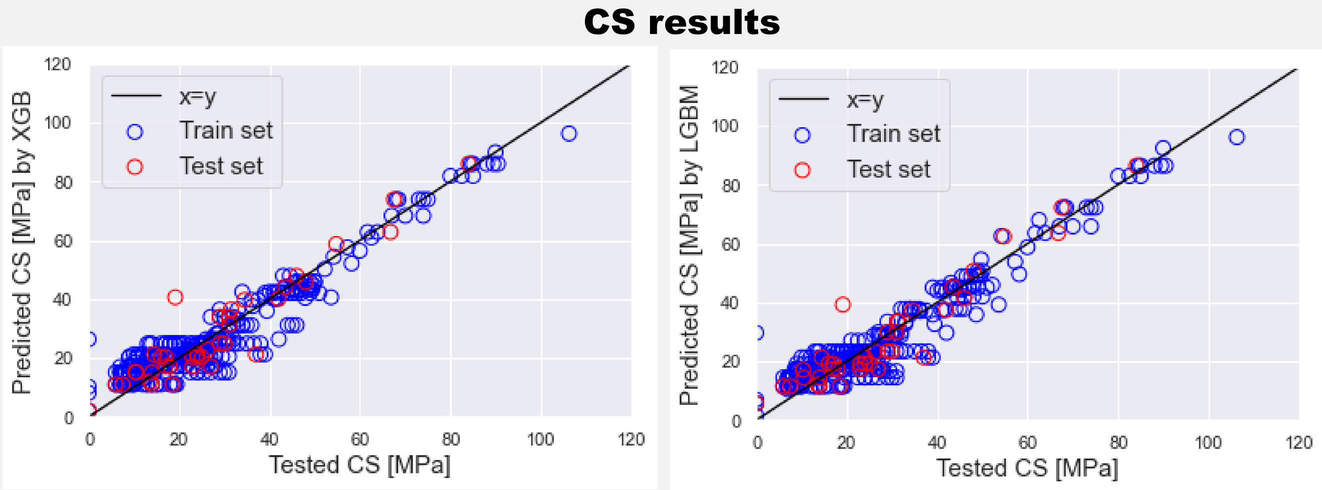
Develop Empirical equation for the material application: The overarching goal of this research is to demonstrate how integrating cutting-edge ML techniques can significantly advance material design and optimization in civil engineering – especially using GEP modeling.



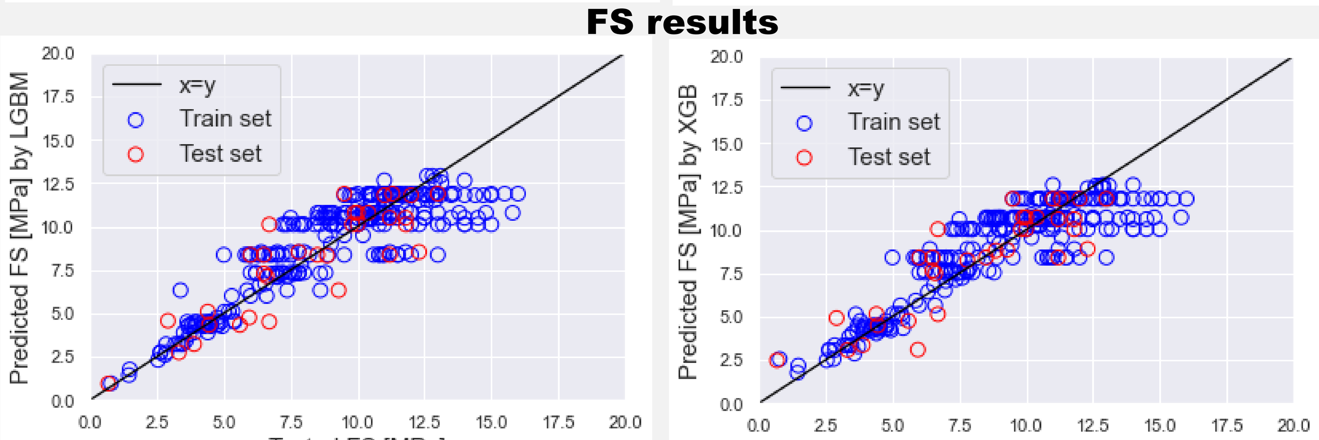


ML results

CS results



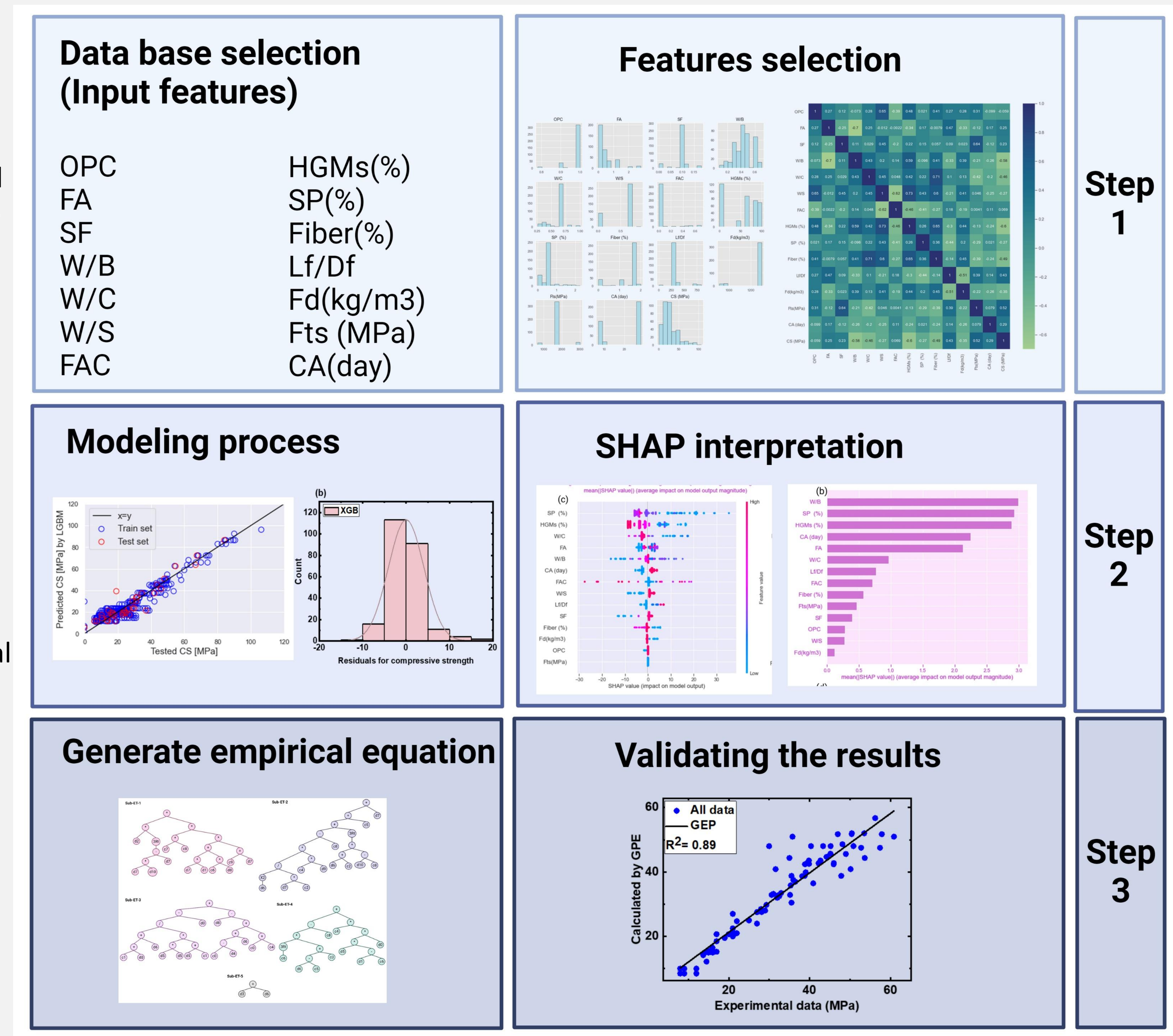
FS results



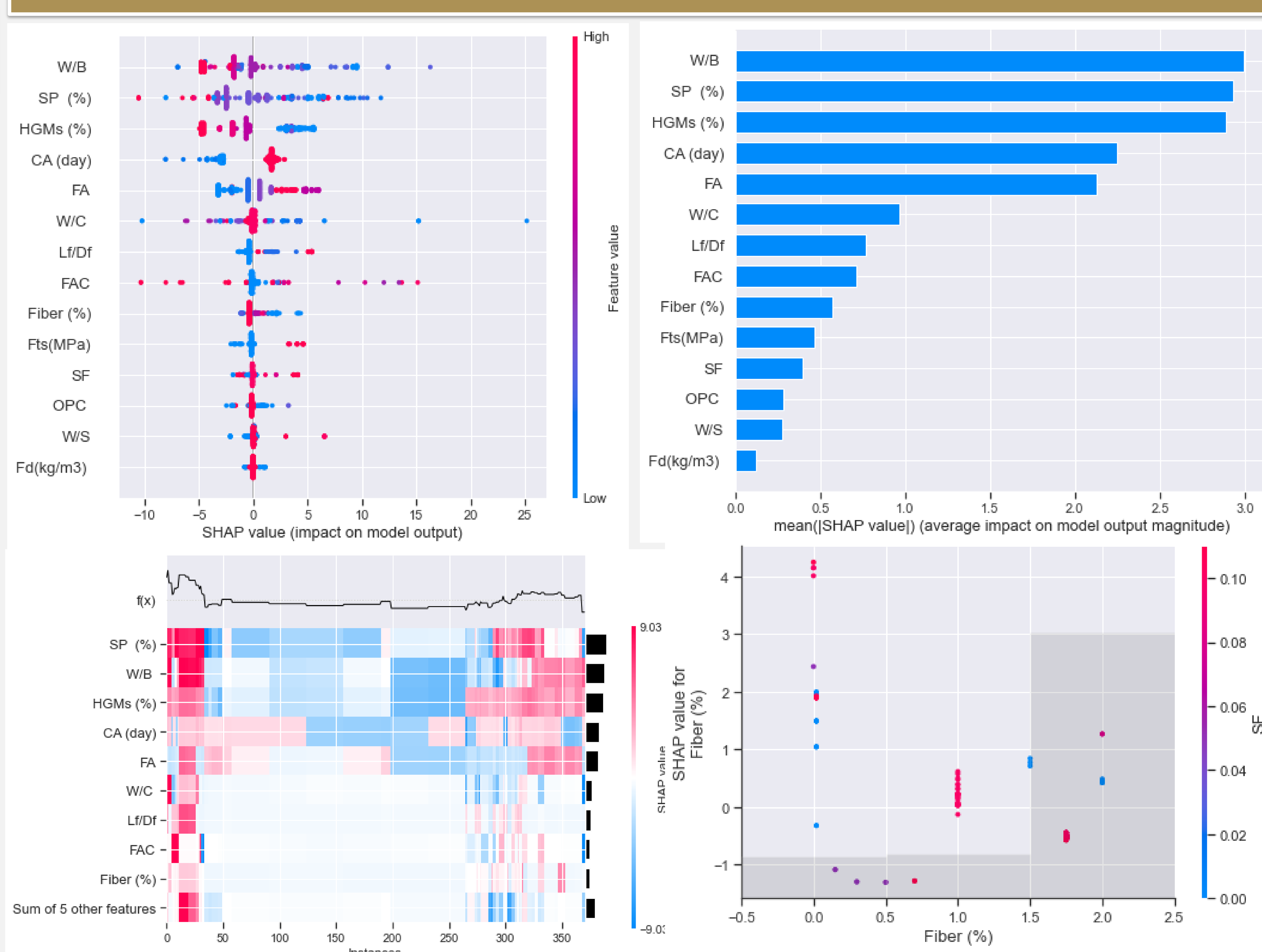
ML	Training						Testing					
CS	R2	MSLE	MAE	MSE	RMSE	MAPE	R2	MSLE	MAE	MSE	RMSE	MAPE
XGB	0.89	0.15	4.414	33.533	5.79	26.45	0.887	0.099	4.458	35.62	5.988	31.221
LGBM	0.90	0.149	4.36	32.591	5.708	25.521	0.886	0.162	4.598	35.942	5.995	31.456

XGBoost	max_depth	[1,15]	15	[1,5]	5
	n_estimators	[100, 500]	500	[100, 500]	250
	gamma	[0, 1]	0	[0, 1]	1
	learning_rate	[0.01, 0.05]	0.03	[0.01, 0.5]	0.5
	subsample	[0.6,1]	01	[0.6,1]	1

Methodology



SHAP results



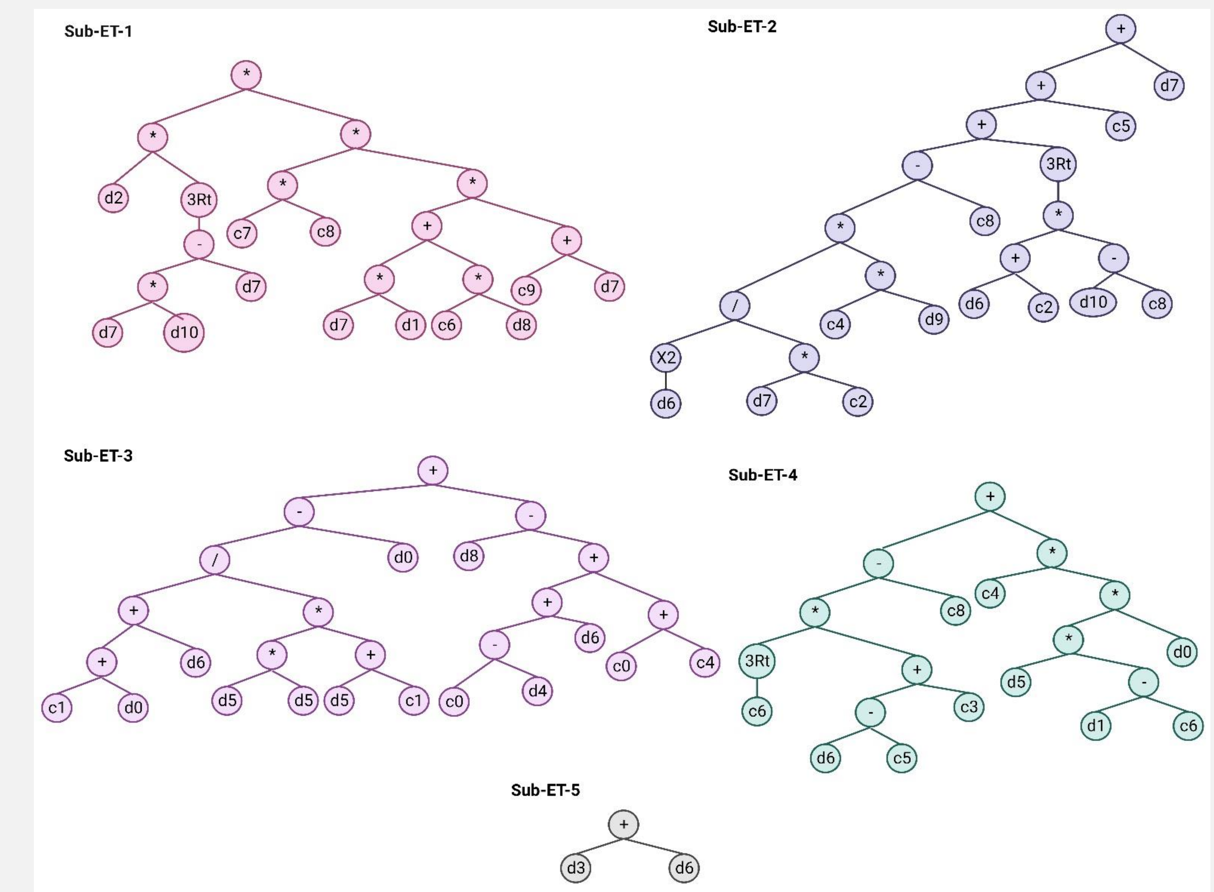
References

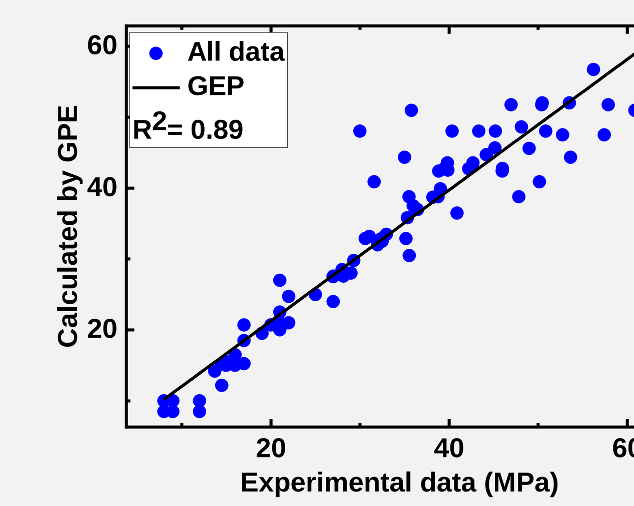
[1] C. Cakiroglu, Y. Aydın, G. Bekdaş, and Z. W. Geem, "Interpretable Predictive Modelling of Basalt Fiber Reinforced Concrete Splitting Tensile Strength Using Ensemble Machine Learning Methods and SHAP Approach," *Materials (Basel).*, vol. 16, no. 13, p. 4578, 2023, doi: 10.3390/ma16134578.

[2] P. G. Asteris, A. D. Skentou, A. Bardhan, P. Samui, and K. Pilakoutas, "Predicting concrete compressive strength using hybrid ensembling of surrogate machine learning models," *Cem. Concr. Res.*, vol. 145, 2021, doi: 10.1016/j.cemconres.2021.106449.

[3] Mousavi SM, Aminian P, Gandomi AH, Alavi AH, Bolandi H. A new predictive model for compressive strength of HPC using gene expression programming. Adv Eng Softw 2012;45:105–14. https://doi.org/10.1016/j.advengsoft.2011.09.014.

Industrial implications





Model	R	R^2	MSE	RMSE	MAE	MAPE	SMAPE
GEP	0.86	0.89	20.5	4.34	3.04	8.18	8.02

$$y = a + b + c + d + e$$

$$a = \left(d_2 \times \sqrt[3]{(d_7 \times d_{10}) - d_7}\right) \times \left((c_7 \times c_8) \times \left(\left((d_7 \times d_1) + (c_6 \times d_8)\right) \times (c_9 + d_7)\right)\right)$$

$$b = \left(\left(\left(\left(\left(\left(\frac{d_6^2}{d_7 \times c_2}\right) \times (c_4 \times d_9)\right) - c_8\right) + \left(\sqrt[3]{(d_6 + c_2) \times (d_{10} - c_8)}\right)\right) + c_5\right) + d_7\right)$$

$$c = \left(\frac{((c_1 + d_0) + d_6)}{(d_5 \times d_5) \times (d_5 + c_1)} - d_0\right) + \left(d_8 - ((c_0 - d_4) + d_6) + (c_0 + c_4)\right)$$

$$d = \left(\left(\sqrt[3]{c_6} \times ((d_6 - c_5) + c_3)\right) - c_8\right) + \left(c_4 \times (d_5 \times (d_1 - c_6)) \times d_0\right)$$

$$e = d_3 + d_6$$

Conclusion

1. The ML models XGB and LGBM exhibited high accuracy in predicting the CS of LWECC during training (R^2 = 0.89 and 0.90) and testing (R^2 = 0.88 and 0.88) respectively. This indicates that the XGB and LGBM models exhibited similar levels of accuracy with minimal error and overfitting.
2. Hyperparametric techniques, using *GridsearhCV* was utilized to optimize the ML performances.
3. SHAP results also presented for the CS here, which shows W/B ratio has higher impact in the compressive strength. It also represents if we increase the W/B, SP (%), HGMs (%) and FAC in the mixture it will decrease the compressive strength.
4. A novel empirical equation based was developed on GEP to predict the CS of LWECC. The GEP model was rigorously validated using the new dataset. This equation exhibits a high level of accuracy in predicting various outcomes, as demonstrated by a testing accuracy of (R^2=0.890).