Full Paper Transactions, SMiRT 28
Toronto, Canada, August 10-15, 2025

INVESTIGATING THE SULFATE ATTACK OF CONCRETE USING EXPLAINABLE ARTIFICIAL INTELLIGENCE

Abobaker Ba Ragaa[1], Fahim Al-Neshawy[2], Mohamed Noureldin[3]

[1] Doctoral Researcher, Aalto University, Espoo, Finland (abobaker.baragaa@aalto.fi)
[2] Staff Scientist, Aalto University, Espoo, Finland (fahim.alneshawy@aalto.fi)
[3] Assistant Professor, Aalto University, Espoo, Finland (mohamed.noureldin@aalto.fi)

ABSTRACT

In this study, explainable Artificial Intelligence (XAI) techniques are used to investigate the behaviour of certain features on sulfate attack progression through concrete in submerged environments, exposed to sulfate ions. The paper utilizes various XAI methods to evaluate the effects of the used features. An extreme gradient boosting (XGBoost) model was used on a synthetic dataset, which contains features related to the concrete mix, as well as the conditions of exposure. The methods used are Partial Dependence Plots (PDP), Shapley Additive Explanations (SHAP), local SHAP and Local Interpretable Model-Agnostic Explanations (LIME). The results indicate that for the used dataset, the features that affected the sulfate progression the most were related to the exposure conditions of the specimen. Features associated with the mix design did not have as strong of a correlation with the given sulfate content in comparison.

INTRODUCTION

In the Finnish nuclear waste repositories, concrete is used in the engineered barriers for final disposal in low and intermediate level waste repositories (LILW). The concrete is designed for a long service life, and it should be able to withstand its environment during this period. In the final disposal stage, the caverns are filled gradually with saline groundwater, and the concrete is exposed to chloride and sulfate attacks in this stage.

Cementitious phases can react with the sulfate ions present in the groundwater, which can alter the behaviour and morphology, resulting in swelling, cracking and spalling of the concrete. Sulfate attacks are common in concrete in contact with water, whether that is groundwater or seawater. The sulfate ions (magnesium sulfate for instance) can interact with cementitious phases, producing ettringite and gypsum, which have a larger volume than the original phases, leading to internal stresses. In the Finnish LILW repositories, elongated contact of the concrete with groundwater in the post-closure phase can result in sulfate degradation, thus, this study is done to utilize Machine Learning (ML) algorithms to investigate the effect of different features on the progression of sulfate attack on submerged concrete.

Explainable Artificial Intelligence (XAI) is an emerging field, studying the behaviour of Artificial Intelligence (AI) and ML models, in a way that attempts to holistically explain their decision-making rationale in an understandable way, making the black-box AI models more transparent. XAI methods are starting to be utilized in recent years for the analysis of different concrete properties, such as the assessment of concrete strength and durability behaviour (Guzmán-Torres et al., 2024; Mohammadi Golafshani et al., 2023; Tran et al., 2022). In this present study, several XAI techniques are employed to explain the behaviour of sulfate attacks on concrete. The techniques include global explanations, and local explanations for specific datapoints.

XAI-BASED FRAMEWORK

In this work, a synthetic dataset of sulfate attack in submerged environment is used (Ba Ragaa et al., 2025), which is based on real datasets obtained from the literature. The synthetic dataset consists of 7000 datapoints, which was generated using Generative Adversarial Network (GAN) based on 172 real datapoints. The features of the data include the water to binder (w/b) ratio, the fly ash content, the concentration of sulfate during storage, the storage period, and the measurement depth. The output is the sulfate content at the measurement depth. With this data, an Extreme Gradient Boosting (XGBoost) model is used to predict the sulfate ingress into the concrete (Chen & Guestrin, 2016; Wang et al., 2019). XAI techniques are then employed quantify and assess the impact of the input features on the output of the model. These include assessment at the scale of the global model, as well as local dependencies at specific range or a specific point. The framework of the model is highlighted in
Figure *1*.

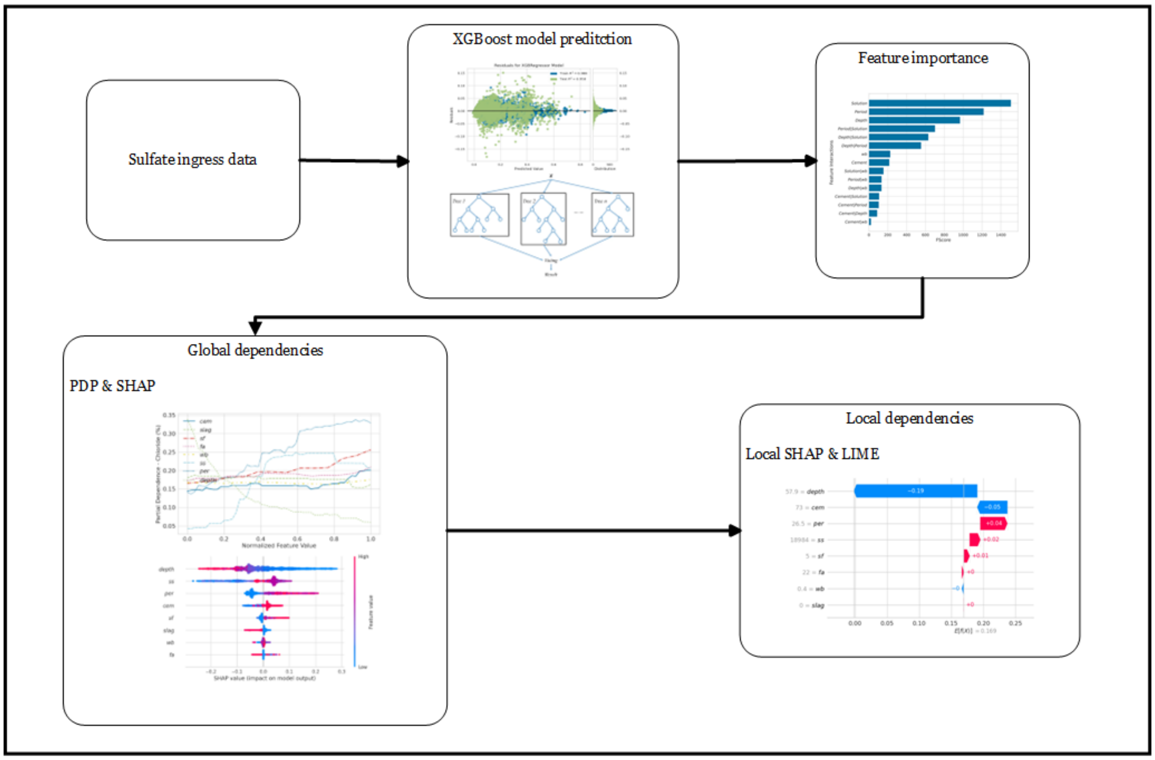

Figure 1: The proposed framework of the study.

RESULTS AND DISCUSSION

XGBoost model

Extreme Gradient Boosting (XGBoost) is used to model the sulfate content in the used dataset. *Table 1* and *Figure 2* highlight the performance of the used model on the dataset, showing the predictions and the residuals of the model in both the training and testing sets. The model achieves coefficient of determination (R^2) score of 0.956, a Root Mean Squared Error (RMSE) of 0.112, and a Mean Absolute Error (MAE) of 0.068 on the testing set.

Table 1: Accuracy metrics of the used XGBoost model.

Result	**Test**	**Training**	**Whole**
R^2	0.956	0.991	0.979
RMSE	0.112	0.049	0.077
MAE	0.068	0.032	0.044

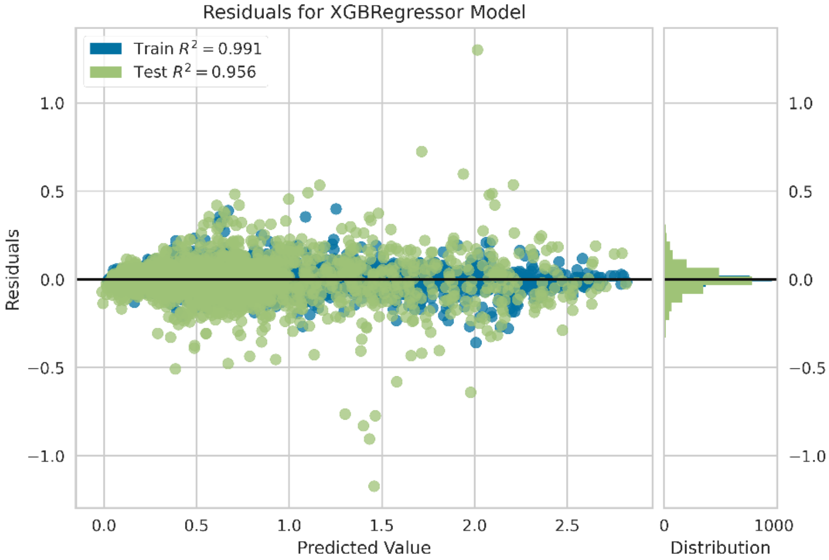

Figure 2: Residuals of the XGBoost model.

Feature importance

Figure 3 shows that features related to the concrete's exposure conditions, such as the intensity of the solution, the exposure period, and the depth of the measurement have the highest impact to the model output. Generally, features related to the concrete (such as the cement and fly ash content, and the w/b ratio) do not have a significant impact to the output.

Global explanations

In this work, XAI techniques are used to interpret the model behaviour in both the global and local scale. In the global explanations, the effect of the features on the output of the model is explained through the whole model. Two techniques are utilized for this purpose, Partial Dependence Plots (PDP) (Greenwell et al., 2018) and Shapley additive explanations (SHAP) (Lundberg & Lee, 2017; Shapley, 1952).

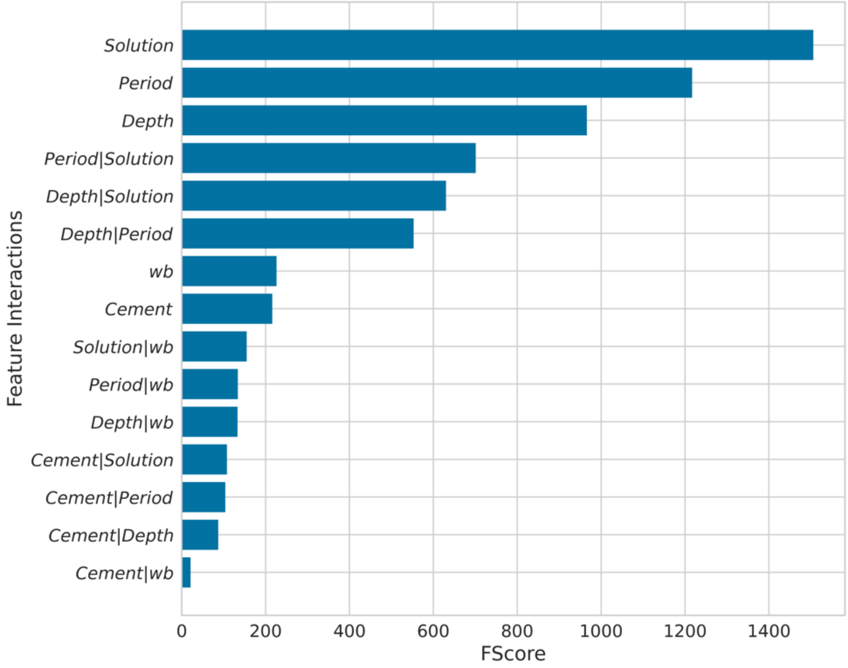

Figure 3: Feature importance and interaction.

Figure 4 shows the PDP for all input features. In PDP, each feature is modified independently of the other features, and the changes in the output are recorded to quantify the influence of each feature on the model output. In this figure, the stable horizontal line indicates that the output of the model does not change with the variation of the feature. This behaviour can be clearly seen in some features, especially w/b and fly ash content. Meanwhile, lines with higher variation indicate the influence of the corresponding feature on the output of the model. The most significant feature being the depth of measurement, where the sulfate content decreases sharply with the increase of the depth. A similar trend can be seen with the storage solution, where the increase of the sulfate content in the exposure solution increases the corresponding sulfate content.

Figure 5 Shows the results of the SHAP explainer. The features are ranked based on their influence on the output of the model. The SHAP results generally agree with the results obtained from the PDP plot, showing that the features related to the mix parameters have less of an impact on the output of the model in compared to features related to exposure conditions.

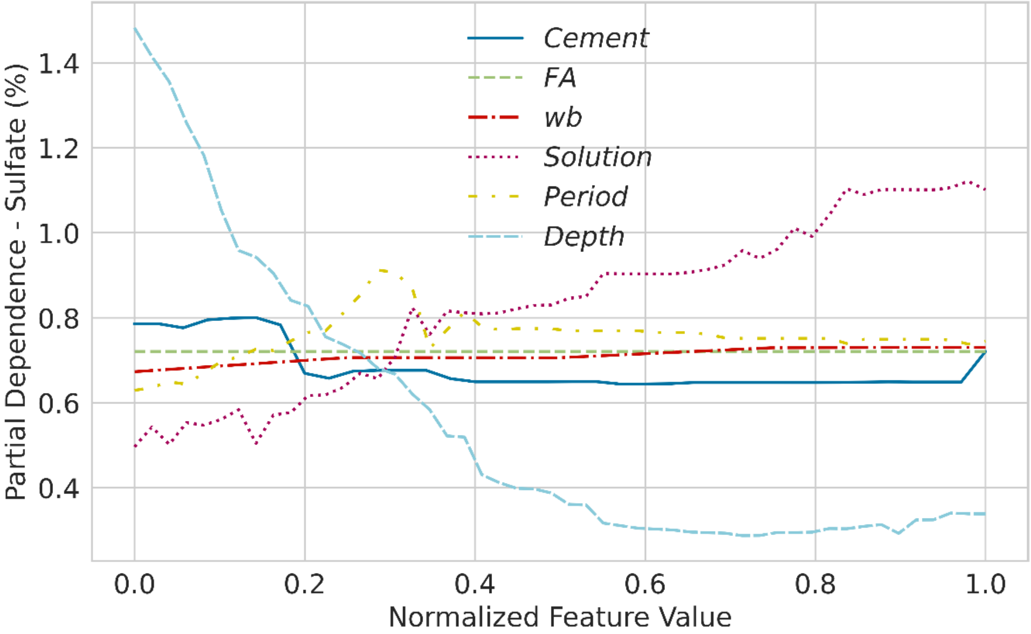

Figure 4: PDP for the input features on the model output.

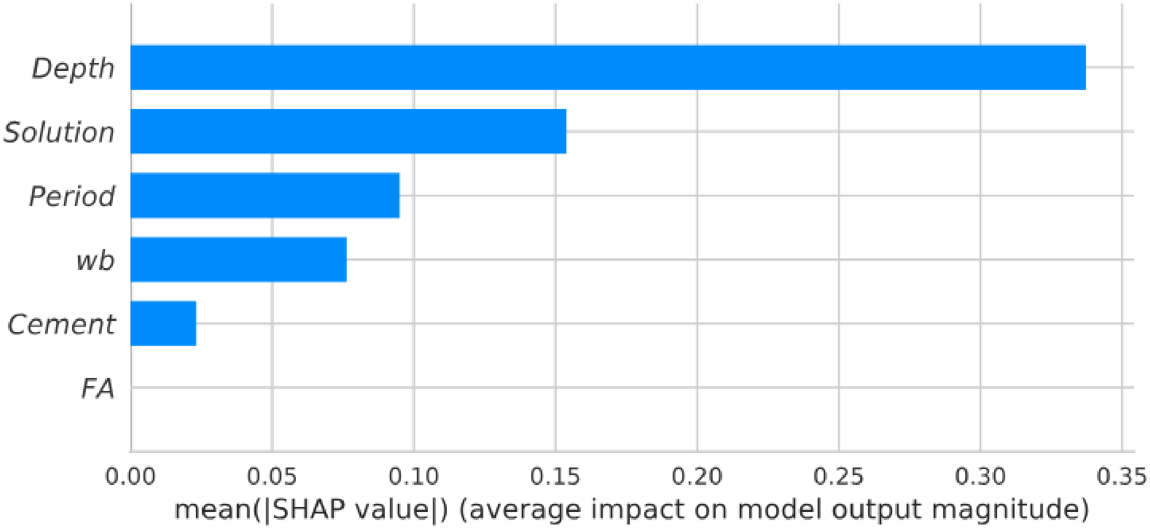

Figure 5: SHAP summary plot.

Local explanations

With local explainability, instead of looking at the behaviour through the whole dataset, certain regions, or in this case specific datapoints are taken, and they are explained individually. This can give an outlook on the model behaviour. In this work, we look at four different examples in the used dataset, which are shown in *Table 2*. The datapoints are chosen to highlight some of the variability in the model behaviour with different data instances. 2 local explainability methods are utilized, local SHAP, and Local Interpretable Model-Agnostic Explanations (LIME) (Ribeiro et al., 2016).

Table 2: The datapoints used in the local XAI analysis.

Sample	Cement	Fly Ash	w/b	Solution	Duration	Depth	Sulfate
1	100	0	0.52	67150	0.3	11	0.89
2	70	30	0.49	38050	0.05	6.2	0.63
3	100	0	0.52	70340	0.1	21	0.27
4	100	0	0.49	148360	1	7.1	1.02

Figure 6 shows the local SHAP results for the chosen datapoints. In this XAI technique, a baseline value (0.726 here) is used, and the specific feature values for the chosen datapoint are used to quantify each feature's contribution to the final model value. In the first instance, the model predicts a sulfate content of 0.925, and the plot shows how each feature contributes to raising the baseline value to the model prediction.

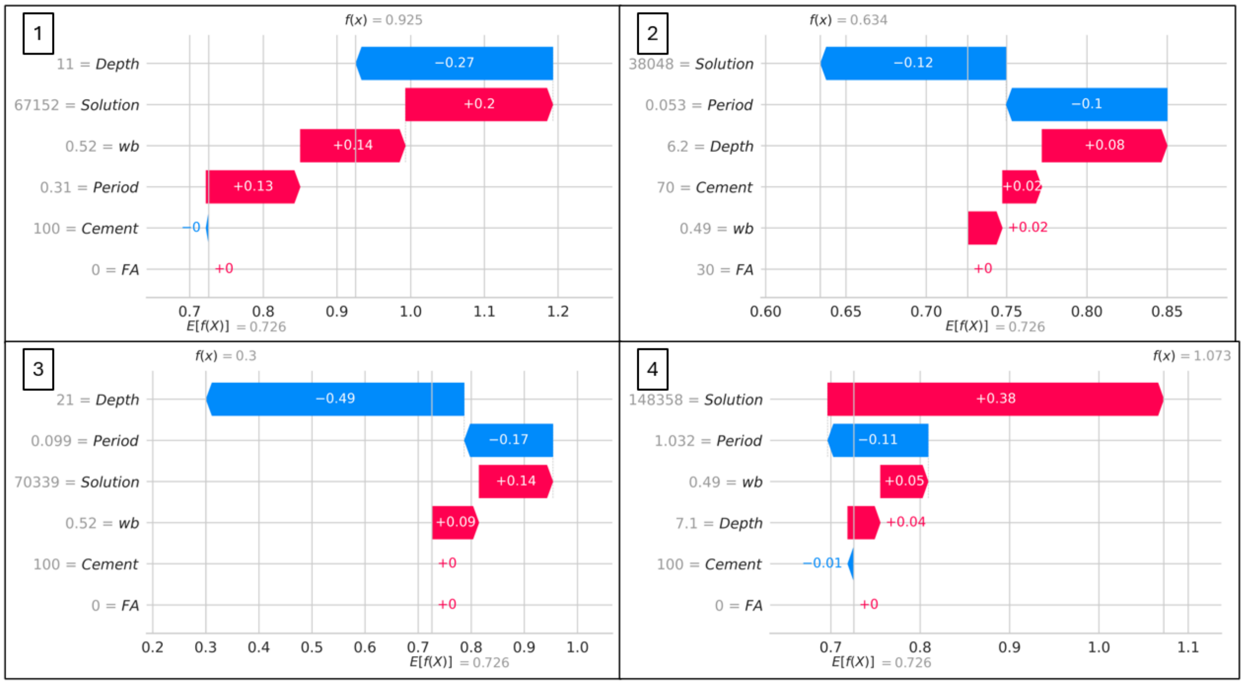

Figure 6: Local SHAP plot for the specified samples.

Figure 7 shows the LIME results of the chosen datapoints. In this instance, the LIME technique shows the impact of each feature without relying on a baseline value. LIME also shows the influence of the features in specific ranges, which is written on the left side of each plot. For example, when looking exclusively at the depth in the 4 datapoints used, the behaviour of the model changes

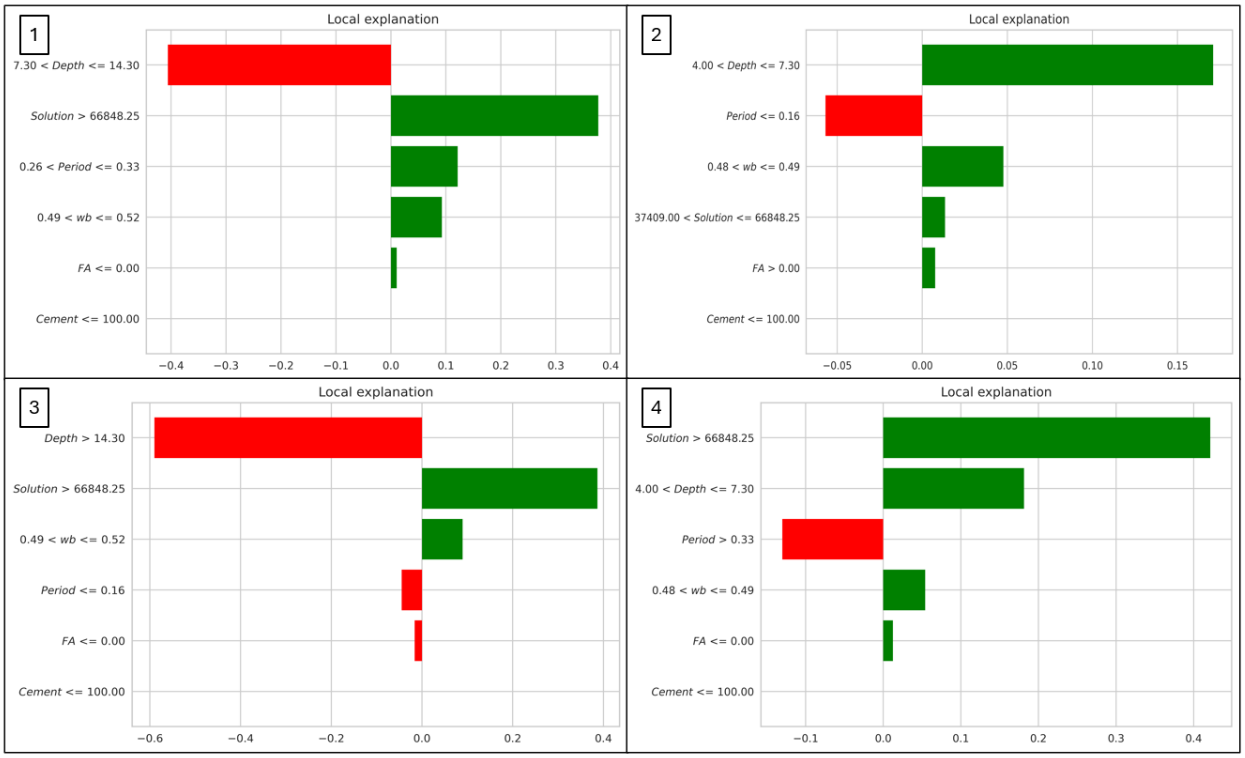

Figure 7: LIME plots for the specified samples.

CONCLUSION

In this work, the explainable artificial intelligence (XAI) techniques have been utilized to assess the sulfate attack behaviour on submerged concrete. This work was carried out to aid in understanding sulfate attacks on concrete in low and intermediate level nuclear waste repositories. The work included predicting the sulfate content in the concrete exposed to groundwater using XGBoost, followed by explaining the model choices (post-hoc) by evaluating the effect of the different utilized features in the model, in both the global and the local scale. The study employed various techniques for assessment, such as using PDP, SHAP, and LIME. Generally, it was discovered that the most important features were related to the exposure conditions of the concrete, and the duration of the exposure, rather than features related to the concrete mixture.

REFERENCES

Ba Ragaa, A., Al-Neshawy, F., & Noureldin, M. (2025). AI-based framework for concrete durability assessment using generative adversarial networks and bayesian neural networks. *Construction and Building Materials*, *471*, 140722. https://doi.org/10.1016/J.CONBUILDMAT.2025.140722

Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, *13-17-August-2016*, 785–794. https://doi.org/10.1145/2939672.2939785

Greenwell, B. M., Boehmke, B. C., & Mccarthy, A. J. (2018). *A Simple and Effective Model-Based Variable Importance Measure*.

Guzmán-Torres, J. A., Domínguez-Mota, F. J., Tinoco-Guerrero, G., Tinoco-Ruíz, J. G., & Alonso-Guzmán, E. M. (2024). Extreme fine-tuning and explainable AI model for non-destructive prediction of concrete compressive strength, the case of ConcreteXAI dataset. *Advances in Engineering Software*, *192*, 103630. https://doi.org/10.1016/J.ADVENGSOFT.2024.103630

Lundberg, S. M., & Lee, S. I. (2017). A Unified Approach to Interpreting Model Predictions. *Advances in Neural Information Processing Systems*, *2017-December*, 4766–4775. https://arxiv.org/abs/1705.07874v2

Mohammadi Golafshani, E., Kashani, A., Behnood, A., & Kim, T. (2023). Modeling the chloride migration of recycled aggregate concrete using ensemble learners for sustainable building construction. *Journal of Cleaner Production*, *407*, 136968. https://doi.org/10.1016/J.JCLEPRO.2023.136968

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). “Why Should I Trust You?”: Explaining the Predictions of Any Classifier. *NAACL-HLT 2016 - 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Proceedings of the Demonstrations Session*, 97–101. https://doi.org/10.18653/v1/n16-3020

Shapley, L. S. (1952). A Value for N-Person Games. *A Value for N-Person Games*. https://doi.org/10.7249/P0295

Tran, V. Q., Giap, V. L., Vu, D. P., George, R. C., & Ho, L. S. (2022). Application of machine learning technique for predicting and evaluating chloride ingress in concrete. *Frontiers of Structural and Civil Engineering*, *16*(9), 1153–1169. https://doi.org/10.1007/S11709-022-0830-4/METRICS

Wang, Y., Pan, Z., Zheng, J., Qian, L., & Li, M. (2019). A hybrid ensemble method for pulsar candidate classification. *Astrophysics and Space Science*, *364*(8), 1–13. https://doi.org/10.1007/S10509-019-3602-4/FIGURES/11