ORIGINAL ARTICLE

Open Access

Revolutionizing engineered cementitious composite materials (ECC): the impact of XGBoost-SHAP analysis on polyvinyl alcohol (PVA) based ECC predictions

Md Nasir Uddin[1,3]* , Al-Amin[1] and Shameem Hossain[2]

Abstract

This study integrates previous experimental data and employs machine learning (ML) methods, including Random Forest (RF), Support Vector Machine (SVM), Artificial Neural Network (ANN), and eXtreme Gradient Boosting (XGBoost), to predict the compressive strength (CS) and tensile strength (TS) of engineered cementitious composites (ECC). XGBoost emerged as the superior model among the four ML models, providing an interpretable and highly accurate predictive framework. To optimize the model performance, hyperparameter tuning using a fivefold cross-validation approach with the data divided into 80% training and 20% testing subsets. The Shapley Additive Explanations (SHAP) algorithm was also employed to reveal the impact of important features, such as the water/binder ratio, fly ash content, and water reducer dosage, on the model's predictions and their interrelationships. The XGBoost demonstrates the most exemplary performance, as reflected in the R^2 values of 0.92 and 0.97 for CS and TS testing, respectively. The SHAP analysis provided insights into the impact of individual features on CS and TS, shedding light on how specific characteristics influence the predictive accuracy of these properties. This highly accurate prediction model uncovers insights into correlated features, aids in creating new mix designs of ECC, and supports global efforts toward a low-carbon future in the construction industry by reducing carbon emissions.

Keywords Engineered Cementitious Composite, PVA-fiber, Random Forest, Artificial Neural Network, Support Vector Machine, Extreme Gradient Boosting (XGBoost), Shapley Additive Explanations

摘要

本研究整合了以往实验数据，采用随机森林算法(RF)，支持向量机算法(SVM)，人工神经网络(ANN)和极限梯度提升算法(XGBoost)等多种机器学习(ML)方法，预测了工程水泥，基复合材料(ECC)的抗压强度(CS)和抗拉强度(TS)。 在四种机器学习模型中，XGBoost表现最为出色，提供了一个解释性强且高度准确的预测框架。 为优化预测模型性能，采用五折交叉验证方法进行超参数调优，并将数据集划分为 80% 的训练子集和 20% 的测试子集。 此外，还采用夏普利加法解释算法 (SHAP) 算法揭示了水胶比，粉煤灰含量和减水剂用量等重要特征对模型预测及其内在关系的影响。 XGBoost在预测CS和TS方面表现 最佳 其R²值分别为0.92和0.97。 SHAP分析揭示了单个特征对CS和TS的影响及其对预测准确性的贡献。 本研究提出的高精确

*Correspondence:
Md Nasir Uddin
nasiruddin@tongji.edu.cn; mdnasiruddin@txstate.edu
Full list of author information is available at the end of the article

预测模型不仅加深了对ECC相关特征的理解，还为ECC新型配合比设计提供了指导，从而有助于减少ECC碳排放，推动全球建筑业迈向低碳未来。

关键词 工程水泥基复合材料(ECC), PVA 纤维, 随机森林(RF), 人工神经网络(ANN), 支持向量机(SVM), 极限梯度提升(XGBoost), 夏普利加法解释算法(SHAP)

1 Introduction

The demand for concrete has drastically increased among all construction materials. Concrete production is needed at a large scale with the development in the construction industry and cities' urbanization. The main constituents of conventional concrete are cement, fine aggregates, and coarse aggregates. Traditional concrete has shown excellent behavior under compression compared to tension. The performance of concrete under tension can be improved, and various methods have been successfully applied, e.g., steel bars or several types of fiber reinforcements that are widely used. High-strength concrete demand and consumption have significantly increased based on construction industry applications. Existing literature has reported that the brittleness of concrete increased with the enhancement in concrete behavior under compression, limiting its structural applications. However, structural applications can be supported using cementitious materials based on high ductile performance, e.g., ECC [1–7].

Research has been conducted on the performance enhancement of (ECC) under compression and tension loading with the addition of different fibers, e.g., polyvinyl alcohol (PVA) [6, 8], basalt fiber [9, 10], polyethylene (PE) [11, 12], polypropylene (PP) [13–15], natural fiber [16] and steel fibers [17–19], hybrid fibers [17, 20]. The formulation of ECC predominantly entails the incorporation of constituents (e.g., cement, fly ash, silica fume, ground granulated blast furnace slag (GGBFS), metacoline, and micro sand). They not only enhance the mechanical performance of concrete but also significantly contribute to green and low-carbon development in civil engineering [21, 22], however, higher amount of industrial waste products reduces the strength. Optimizing proper amount is a key and a challage to develop sutable binders. By reducing the carbon footprint associated with cement production, these sustainable practices align with global efforts toward environmentally conscious construction. The further inclusion of various fiber contents enhances ECC's mechanical prowess, albeit with an attendant increase in complexity concerning the accurate prediction of its CS and TS. The endeavor to accurately predict the CS and TS of ECC is not merely a theoretical preoccupation but a substantive requisite that underscores the operational integrity and longevity of structures fabricated from ECC. This in turn could

lead to a substantial decrease in experimental workload and cost. The material components of green PVA- ECC are highly complex and play a crucial role (e.g., reducing carbon footprint, improve strength, ductility, toughness. The need for innovative approaches to support the green and low-carbon development of civil engineering is highlighted by the challenges of accurate modeling mechanical characteristics based on mixture design parameters through conventional regression studies [23, 24].

However, the heterogeneity engendered by the diverse fiber contents, coupled with the interactions among the primary constituents, often morphs the predictive landscape into a complex and challenging endeavor [25–28]. Previous research has reported that the CS of ECC has been evaluated with experiments based on water/binder ratios (W/B), sand-binder ratios (S/B), particle size grading, age of samples, type of cement, and shape of samples [1, 4, 7]. Supplementary cementitious materials (SCMs) (e.g., fly ash, ground granulated blast furnace slag, rice husk ash (RHA), metakaolin, volcanic ash, silica fume, calcined clays, recycled glass and natural pozzolans) are commonly used as a replacement of the portland cement in concrete. The use of SCMs in concrete, replacing 5–20% of clinker, enhances long-term mechanical properties, durability and CO_2 emissions. However, high volume clinker replacements may cause early age performance losses, prompting research to balance sustainability and performance by optimizing the mixture. Several research has been condicted on SCMs hydration [29, 30], workability [31, 32], strength [33, 34], long-term durability [35]. Making ultra high performance concrete (UHPC), fillers like GGBFS, nano-silica (NS), RHA showed promising results at 28-day compressive strengths of 100 MPa and 150 MPa were achieved with RHA mean particle sizes of $8\sim\mu m$ and $3.6\sim\mu m$, respectively [36–38]. Achieving optimal particle packing and reactivity requires extensive trial and error, making it time-consuming and resource-intensive. ML can identify new potential SCMs and fillers by analyzing vast amounts of data from literature and experiments, accelerating the discovery of effective materials for ECC or UHPC.

Examining the durability of concrete materials entails significant costs and consumes a considerable amount of time due to the intricate nature of curing and construction procedures [39]. Researchers have increasingly turned to alternative techniques to construct

prediction models that can shed light on the mechanical performance of concrete materials. This shift is driven by a desire to achieve cost-effectiveness and temporal efficiency in their research endeavours. However, the conventional methods of numerical analysis, which frequently involve complex non-linear equations, provide a significant obstacle when it comes to achieving accurate predictions [40]. As a result, there has been a growing need to accurately anticipate the performance of concrete using traditional procedures. This has led to a recognition of the complexity involved in this endeavour, prompting a shift towards the utilization of modern soft computing techniques.

In recent years, machine learning (ML) based prediction algorithms have gained significant popularity in developing prediction models for CS and TS's behavior of concrete materials by simultaneously reducing the time and cost. Naser et al. [41] described how ML can applied in the structural members, optimizing concrete mixtures for constraction site application. Tapeh et al. [42] extensively reviewed AI, ML and DL applications in structural sections, earthquake, wind, and fire engineering, and successfully showed potential future in those sectors. Such as fuzzy logic (FL) [43–45] [46], particle swarm optimization (PSO) [47, 47] artificial neural network (ANN) [48–50], genetic algorithm (GA) [51, 52, 52, 53] gene expression programming (GEP) [52, 54], random forest (RF) [55, 56], are commonly used in the concrete field. For instance, Khashman et al. [57] have developed an ANN-based prediction model to forecast the concrete's CS. Table 1 shows several ML algorithms that have previously been used to predict the CS and TS for the HPC and several cementitious composite materials. There is a paucity of literature on the subject matter concerning fiber-based concrete focusing on projecting the mechanical properties of ECC predominantly reinforced with

PVA and steel fibers. [58–60]. Uddin et al. [28] used PVA fiber to predict the CS of the ECC by ANN; however, an extensive study has not been conducted.

Random forests (RF) [61] are applied to forecast the strength of concrete materials, and RF is an advanced ML algorithm and is very effective in solving non-linear regression and classification problems [62]. Chen et al. [63] reported that XGBoost is more efficient and newly developed than other ML algorithms. XGBoost is highly recommended to avoid overfitting problems in civil engineering, e.g., predicting shear strength [64], interface shear strength [65], dynamic modules [66], CS, and creep [67, 68]. Furthermore, it has been successfully applied to predict, e.g., seismic drift [69], concrete strength [70], shear capacity of FRP-RC wall and flat slabs [71–73], buckling analysis for steel beam [74]. To a limited extent, the SHAP method has been used on risk factors [75], failure modes of RC walls [76], and concrete strength [70], which preciously illustrates how the existing research on features' importance and their global and local correlate each other and solve the complex problem. The latest ML-based algorithms have been used widely in earlier research to predict concrete properties for varied materials. Mahjoubi et al. [77] used different plastic fibers to predict and optimize the mixture for strain-hardening cementitious composites (SHCC) based on ML. However, according to the author's best knowledge, have not been found enough studies to predict the CS and TS behavior of PVA-ECC materials based on XGBoost and SHAP algorithms.

This paper develops a systematic ML-based prediction models program to only predict the CS and TS of ECC by considering the PVA fiber. For training and testing, data from 81 mixed samples were collected. The XGBoost method constructs a high-accuracy predictive model across all ML models. Random selections of 80%

Table 1 The Summary of ML-based studies to predict the CS and TS behavior of concrete

Model	Concrete	Samples	R^2	Output	Ref
ANN	HPC	727	0.961	CS, TS	[78]
Neural expert system (NEX)	HPC	864	0.883	CS	[79]
Fuzzy polynomial neural networks (FPN)	HPC	458	0.852	CS	[80]
Multiple additive regression tree (MART)	SC	1030	0.890	CS	[53]
M5P model tree (MSP)	SCC	300	0.844	CS	[81]
Random forest (RF)	HSC	357	0.892	CS	[62]
XGBoost	HPC	190	0.993	CS, TS	[82]
Random forest (RF)	HPC	190	0.984	CS, TS	[82]
AdaBoost	HPC	190	0.868	CS, TS	[82]
Random forest (RF)	MSC	754	0.981	CS	[70]
Random forest (RF)	MSC	754	0.962	TS	[70]

CS compressive strength, *TS* tensile strength, *HPC* high performance concrete, *SC* slag concrete, *SSC* self compacting concrete

and 20% are used for this purpose. The SHAP algorithm is employed to explain the model's key features and their complex correlations. The suggested model's prediction outcomes are compared against earlier research and chosen best ML algorithm.

2 Machine learning algorithm

2.1 Material and database creation

A thorough dataset was collected to develop machine learning (ML) models for predictive analysis. This dataset consisted of 820 experimental data points related to polyvinyl alcohol engineered composite concrete (PVA-ECC), which were obtained from various literature sources [83–91] based on only considering PVA fiber with F and C type fly ash. Table 2 elucidates the statistical attributes of eleven input parameters predicated on the selected experimental data gleaned from the aforementioned sources.

One of the pivotal facets of data science and machine learning paradigm is data visualization, which lends a nuanced understanding of the underlying data distribution and inherent patterns. The statistical distribution of each numerical parameter, extricated from the database, is vividly portrayed in Fig. 1. This figure delineates the histograms corresponding to the CS and TS of PVA-ECC along with other essential input variables. As exemplified in Table 2, the CS oscillates between 21.30 to 75.20, while the TS exhibits a range between 3.08 and 5.82. Other input variables are articulated in terms of mean, standard deviation, median, minimum, and maximum to foster a comprehensive understanding of the data distribution.

Figures 2 and 3 display heatmaps that visually represent the relationship between various input features and the CS and TS, respectively. Each feature is presented in a separate subplot within the heatmaps. In Fig. 2, it is observed that CS (in MPa) exhibits a highly positive

correlation with Cement (in kg/m^3), Nominal strength (NS) (in MPa), Aspect Ratio (AR), Water-to-Sand ratio (W/S) with correlation coefficients of 0.80, 0.49, 0.31, and 0.20 respectively. On the other hand, Fly ash (FA) (in kg/m^3), Sand (kg/m^3), and High-Range Water Reducer (HRWR) (in kg/m^3) illustrate a negative impact on CS. The RA of the fiber in ECC is very important because of it helps to imporove crack control and durability, workability, fiber distribution and orientation and improve the permeability. Optimiziing the RA could improve the meachanical performace of PVA-ECC.

Furthermore, Sand content and HRWR demonstrate positive correlations with coefficients of 0.31 and 0.76, respectively, while NS and Young's Modulus (YM) (in GPa) exhibit positive correlations with coefficients of 0.59 and 0.83, respectively in Fig. 3. The size of the cube and cylender has been considered using the Nevilar approach and concerted the strength into the cylinder [92]. The dog bon test has been considered according to the apan Society of Civil Engineers (JSCE) [93].

Figure 4 presents a comprehensive representation of the machine learning model, illustrating the data distribution, the approaches employed for predicting CS and TS, and the entire workflow. The study utilized the XGBoost and SHAP modelling methodologies, implemented through the Python packages 'xgboost' and 'shap' correspondingly. The analytical processes were conducted using a Python programming environment, which ensured a methodical and replicable analysis.

The current methodological framework highlights the meticulous processes entailed in data curation, visualization, and analytical modeling. These procedures play a pivotal role in enabling an exhaustive predictive analysis of the mechanical strength of PVA-ECC. The utilization of advanced machine learning models and data visualization techniques allows for a comprehensive

Table 2 Statistical variables of the different data used in the database for the ECC mixture

| Parameters	Mean	Median	Min	Max	StD	Skewness	Kurtosis
Cement(kg/m^3)	427.59	393.00	190.00	618.00	128.72	-0.26	-1.17
Fly ash(kg/m^3)	797.69	804.50	649.00	1063.00	123.81	0.51	-0.76
W/B	0.26	0.27	0.20	0.30	0.02	-1.26	6.07
Sand(kg/m^3)	438.40	447.00	338.00	494.00	36.07	-1.75	2.86
W/S	0.73	0.73	0.55	0.92	0.08	0.68	1.91
HRWR(kg/m^3)	4.47	4.95	1.60	10.20	2.56	0.62	-0.69
AR	215.86	205.13	205.13	307.69	32.27	2.73	5.60
NS (MPa)	1609.88	1604.94	1600.00	1620.00	9.97	0.03	-2.05
YM (GPa)	42.49	42.80	40.00	42.80	0.93	-2.52	4.47
CS (MPa)	46.07	47.75	21.30	75.20	13.33	-0.29	-0.69
TS (MPa)	3.58	4.03	0.00	5.81	1.59	-1.16	0.80
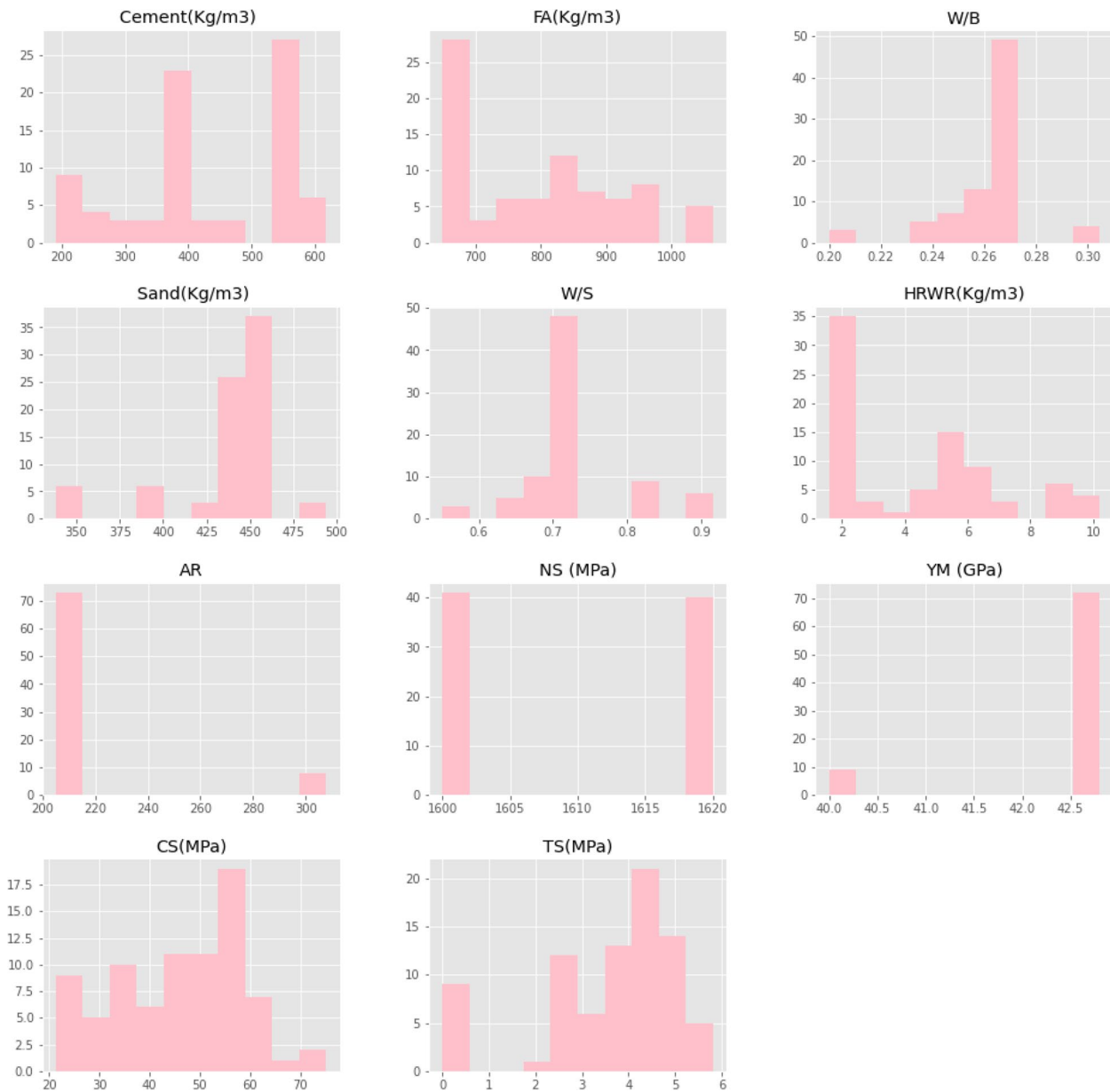
Fig. 1 Histograms of the parameters: **a** Cement (kg/m³); **b** FA (kg/m³); **c** W/B, **d** sand (kg/m³); **e** W/S; **f** HRWF (kg/m.³); **g** AR; **h** NS (MPa); **i** YM (GPa); **j** CS (MPa); **k** TS (MPa)

comprehension of the intricate relationships existing between input parameters and the mechanical properties of ECC. Consequently, this contributes significantly to the burgeoning knowledge base in the realm of predictive analysis for engineered composite materials.

2.2 Random Forest (RF)

The RF algorithm is a type of ensemble learning technique commonly employed for classification and regression tasks. It achieves this by creating a multitude of decision trees throughout the training process. The process of aggregating predictions from multiple trees enhances the model's generalizability and robustness. The bagging technique mitigates the issue of overfitting by introducing a randomization component in selecting data for each tree. Incorporating randomization at every split in decision trees improves the model's resilience against outliers and noise. Missing data can be addressed using imputation techniques or weighted splitting methods.
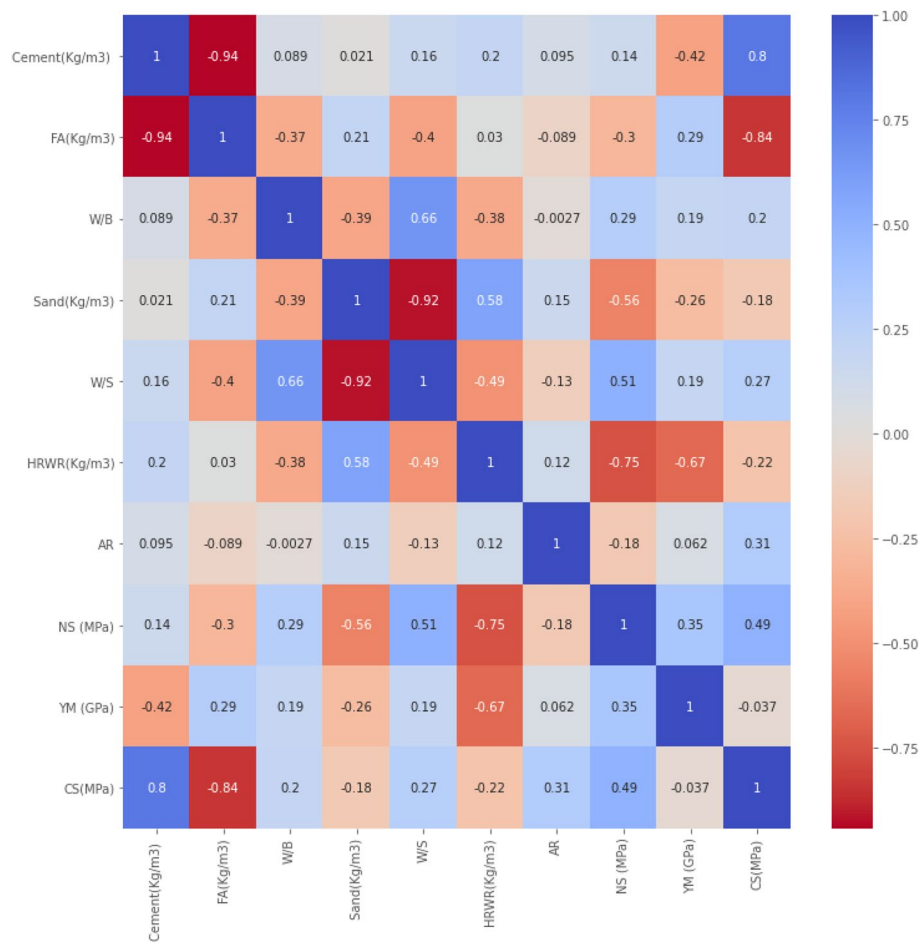
Fig. 2 Heatplot of the CS: **a** Cement (kg/m³); **b** FA (kg/m³); **c** W/B, **d** sand (kg/m³); **e** W/S; **f** HRWF (kg/m.³); **g** AR; **h** NS (MPa); **i** YM (GPa); **j** CS (MPa)

The Out-of-Bag (OOB) Error Estimation method calculates the average prediction error for each training sample by utilizing trees that were not trained on that particular sample [94].

2.3 Support Vector Machine (SVM)

The SVM is a supervised learning technique that is predominantly employed for tasks such as classification, regression, and outlier detection. The algorithm identifies a hyperplane that achieves maximum separation between classes inside the feature space. Additionally, it utilizes the kernel trick to effectively handle both linear and non-linear data by transferring it to a space with greater dimensions. SVM exhibit resilience when confronted with high-dimensional data and provide a range of kernel functions, including linear, polynomial, and radial basis function (RBF). The essential hyperparameters encompass the regularization parameter (C) and parameters relevant to the kernel. The classifier is intrinsically designed for binary classification, however, it can accommodate multiclass classification by employing approaches such as one-vs-rest or one-vs-one [95]. SVM exhibits versatility and robustness as a machine learning technique, as they can perform regression tasks (known as Support Vector Regression, SVR) and detect outliers. SVM can be low dimensional to the high binational problem using kernel function, solving a non-linear problem using a linear method [96].

2.4 Artificial Neural Network (ANNs)

ANNs are employed in regression analysis to make predictions of continuous values by utilizing input data. The neural network architecture consists of three layers: input, hidden, and output. The output layer is comprised of a solitary neuron that utilizes a linear activation function to predict continuous values. The training process is the manipulation of weights and biases through the utilization of backpropagation to decrease error. This error is typically quantified using metrics such as Mean Squared Error (MSE) or Mean Absolute Error (MAE). Hidden layers utilize activation functions such as Rectified Linear Units (ReLU) in order to capture and represent
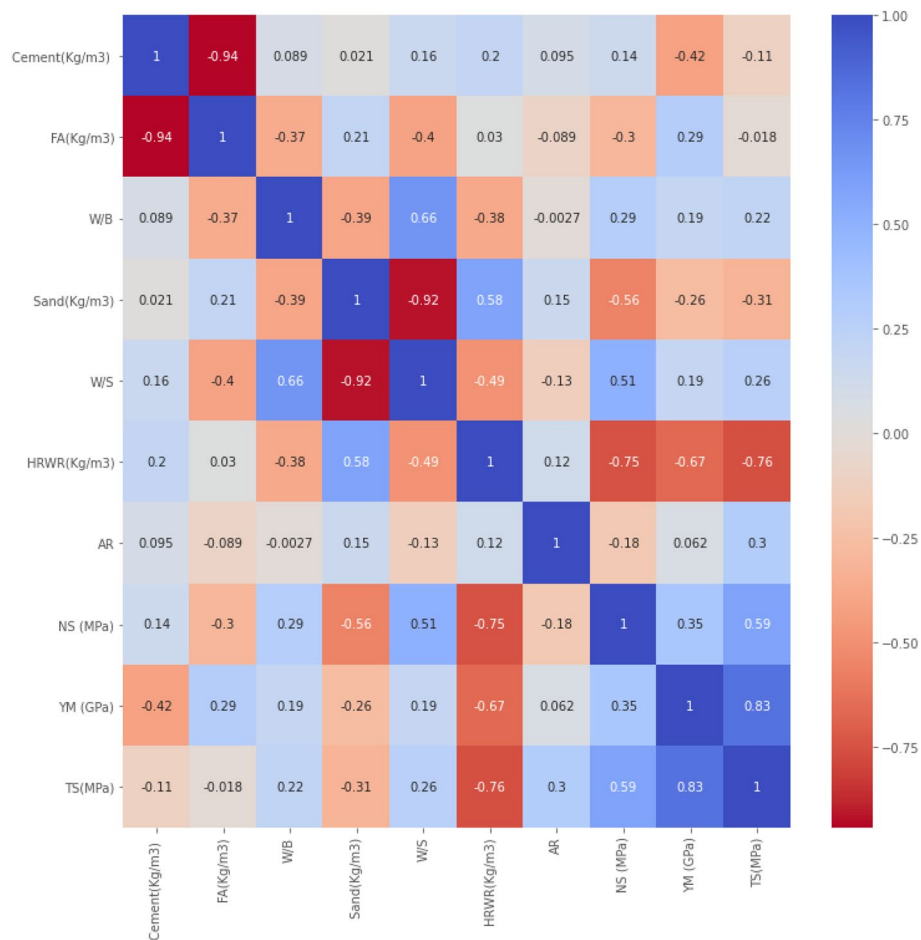
Fig. 3 Heatplot of the TS: **a** Cement (kg/m³); **b** FA (kg/m³); **c** W/B, **d** sand (kg/m³); **e** W/S; **f** HRWF (kg/m.³); **g** AR; **h** NS (MPa); **i** YM (GPa); **j** TS (MPa)
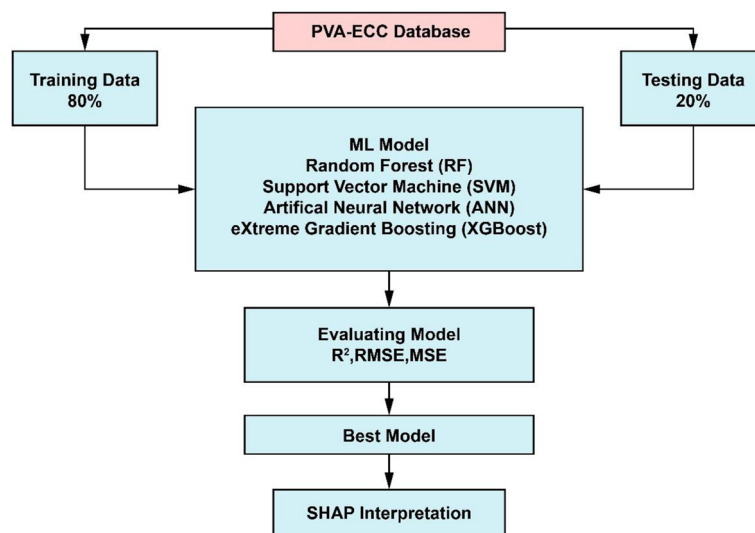

Fig. 4 Flow chart for the ML model for predicting CS

Uddin *et al. Low-carbon Materials and Green Construction*   (2024) 2:11

Page 8 of 23

non-linear relationships. The primary hyperparameters encompass the quantity of hidden layers, neurons, and the learning rate ANNs find use in several domains such as finance, engineering, and other areas where the prediction of continuous values is crucial. The performance of ANNs in these domains is typically assessed using evaluation metrics such as R-squared, mean squared error (MSE), or mean absolute error (MAE).

The neural network architecture includes input, weight, a function, activation function, and outputs Fig. 5(a) shows the neural anatomy with demonstration, Fig. 5(b) describes the neural network architecture with eleven input parameters and ten hidden layers.

2.5 Extreme Gradient Boosting (XGBoost)

A predictive ML model may be trained with the available database using the supervised family of machine learning techniques. There are several algorithms in this category; however, XGBoost, a recently created ML-based regression technique, is used to build the model because of its success in other regression-like applications [97, 98]. XGBoost is a more advanced variant of the ensemble learning method gradient boosted decision tree (GBDT) that improves loss function aspects [99]. The mathematical fundamentals will be briefly discussed in Sect. 2.

2.5.1 Shapley Additive Explanations (SHAP)

To make the model more exact, input and output features have a significant impact. It is also extremely complicated and complex. Model interpretability is understood by itself and explained to interpret the model. Lundberg and Lee introduced [100] SHapley Additive exPlanation

(SHAP) to evaluate ML model predictions using Shapley values. The average of all feature permutations' marginal contributions, showing the impact of that feature on the generated output, [99] is the shapley value for each feature. The SHAP explanation model $g(a^*)$ is a linear addition of input features expressed as follows [101].

$$g\left(a^*\right) = \phi_0 + \sum_{i=1}^{N} \phi_0 a_i \tag{1}$$

where the $a^* \in \{0, 1\}^N$, N is the number of input features ϕ_0 models output with inputs and ϕ_i used as feature attribution values [98].

$$\phi_i = \sum_{S \in N} \frac{|S|! - (N - |S| - 1)!}{N!} [f(S \cup \{x_i\}) - f(S)] \tag{2}$$

where x_i used as a feature's value. S and N used as a subset of the features and features numbers. $f(S \bigcup \{x_i\})$ and $f(S)$ are the prediction of the model trained with and without features.

3 Result and discussion

3.1 Simulation results of machine learning models

The hyperparameter technique was addoped using the gridsearch CV using fivefold cross validation for the RF, SVM, ANN and XGBoost in the model because of acchiving higher accuracy for the testing dataset. In this study, the model uses z-normalization to standardize the continuous input parameters. Naser et al. [102] explained how a wide range of performance fitness and error metrics (PFEMs) and there importance of metrics that are
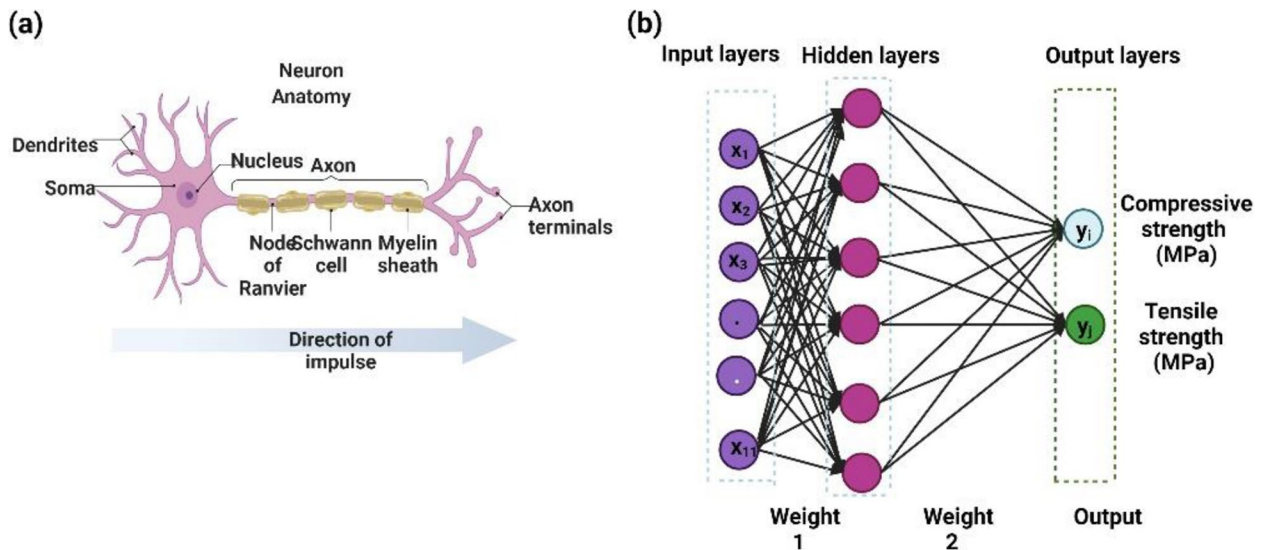

Fig. 5 a Neural anatomy, **b** The neural network architecture

commonly used in evaluating ML models, especially in regression and classification tasks within engineering and science applications. The error matrices in the paper has followed [103], because both MSE and RMSE are sensitive to larger errors because they square the prediction errors before averaging them. Many alternative metrics (e.g., Mean Absolute Percentage Error (MAPE)), might provide similar insights but do not fundamentally change the understanding of model performance [104, 105].

3.2 Predicting Random forest (RF)

Figures 6(a) and (b) depict the actual and predicted results for the CS and TS using the RF model. To optimize the model max_depth, n_estimators, min_samples_split, min_samples_leaf and max_features are 10, 150,10,15, 'auto' for the CS and 4, 150, 10, 10, auto, respectively for the TS as chosen after the GridSearchCV has been used. The RF model yields an R^2 value, MAE, MSE, and RMSE of 0.95, 1.50, 0.017, and 2.57, respectively, for the training data, and 0.87, 3.08, 22.26, and 4.71, for the testing data

in terms of CS. The R^2 value of RF is comparable to that of the ANN, as shown in Table 3. However, the XGBoost model outperforms RF in terms of CS predictions.

Regarding the prediction of TS, the RF model achieves an R^2 value, MAE, MSE, and RMSE of 0.94, 0.13, 0.03, and 0.19, respectively, for the training data, and 0.78, 0.37, 0.21, and 0.46, respectively, for the testing data. In this case, RF performs better than the ANN model but is still less accurate than XGBoost.

3.3 Predicting Support Vector Machine (SVM)

The SVM is widely utilized for regression and classification tasks in machine learning. In this study, SVM is employed to predict the CS and TS of PVA- ECC after 28 days. In order to optimize the SVM model, the 'rbf' kernel was selected and tested against the 'linear' and 'poly' kernels. For the CS, the values of gamma and C were set to 0.5 and 10 respectively. Similarly, for the TS, the values of gamma and C were chosen as 0.3 and 5 respectively.
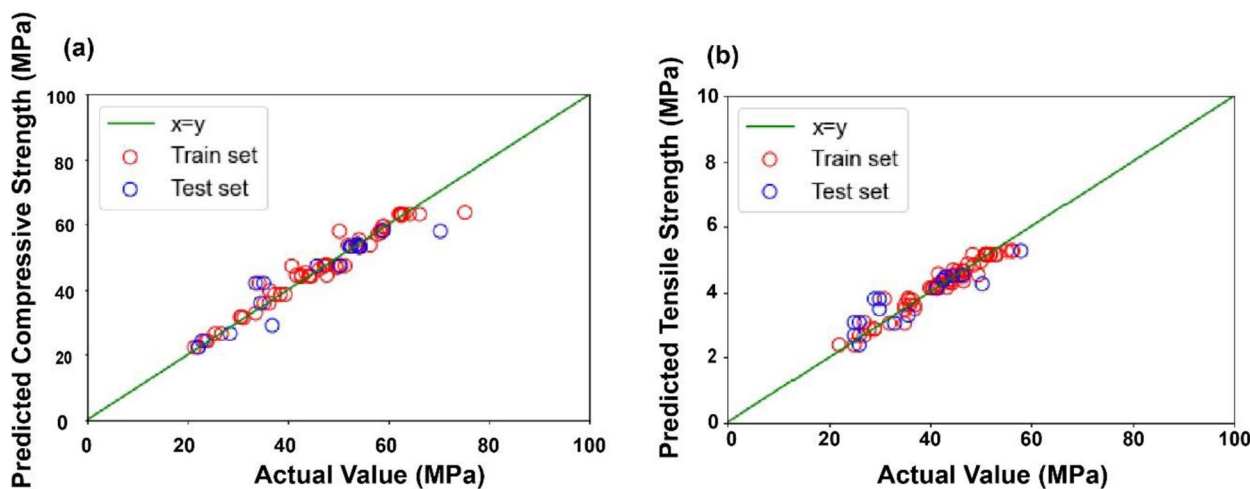

Fig. 6 a and **b** show the compressive strength and tensile strength using the RF model

Table 3 Performance evaluation of the ML model

Model	Output	Training-80%				Testing-20%			
ML	Strength	R^2	MAE (MPa)	MSE	RMSE (MPa)	R^2	MAE (MPa)	MSE	RMSE (MPa)
RF	CS	0.95	1.5	0.017	2.51	0.87	3.08	22.26	4.71
RF	TS	0.94	0.13	0.037	0.19	0.78	0.37	0.21	0.46
ANN	CS	0.98	0.87	1.9	1.38	0.89	3.98	22.89	4.10
ANN	TS	0.76	0.29	0.16	0.4	0.59	0.45	0.4	0.63
SVM	CS	0.89	3.13	17.2	4.15	0.88	3.37	21.11	4.59
SVM	TS	0.90	2.92	16.27	4.03	0.79	4.03	28.50	5.33
XGBoost	CS	0.99	0.8	1.33	1.15	0.92	2.29	14.57	3.81
XGBoost	TS	0.98	0.08	0.01	0.10	0.97	0.17	0.11	0.26

Figure 7(a) and (b) depict the performance metrics, namely R², MAE, MSE, and RMSE, for the training and testing phases. Notably, the SVM model exhibits lower values for these metrics compared to the other machine learning models when predicting CS. However, for the TS, the performance of the SVM model appears to be similar to the ANN model, indicating that the RF and XGBoost models yield more precise predictions than SVM when considering the dataset.

3.4 Predicting Artificial Neural Network (ANNs)

Table 3 highlights the accuracy of various ML models in predicting the CS. The results indicate that the XGBoost model outperforms the RF and SVM models in terms of accuracy. The RMSE values for the XGBoost model remain consistent between the training and testing sets, which suggests its superior accuracy compared to RF and SVM in predicting the CS.

However, it should be noted that the ANNs model does not accurately predict the TS due to overfitting. Thus, based on the findings presented in Table 3, it can be concluded that the XGBoost model is the most prominent and leading among all ML models utilized in this research. Figure 8(a) and (b) illustrate the use of the MLPRegressor to predict the CS and TS within the Python script. The hyperparameters selected for optimization in this case include a hidden_layer_sizes of (150), learning_rate_init of 0.001, learning_rate set to 'constant', activation function as 'relu', solver as 'adam', alpha value equal to 0.0002, batch_size set to 'auto', power_t of 0.6, and maximum iterations capped at 500.

3.5 Predicting Extreme Gradient Boosting (XGBoost)

The optimization of the XGBoost model was meticulously undertaken using GridSearchCV, which is implemented with a fivefold cross-validation strategy. The
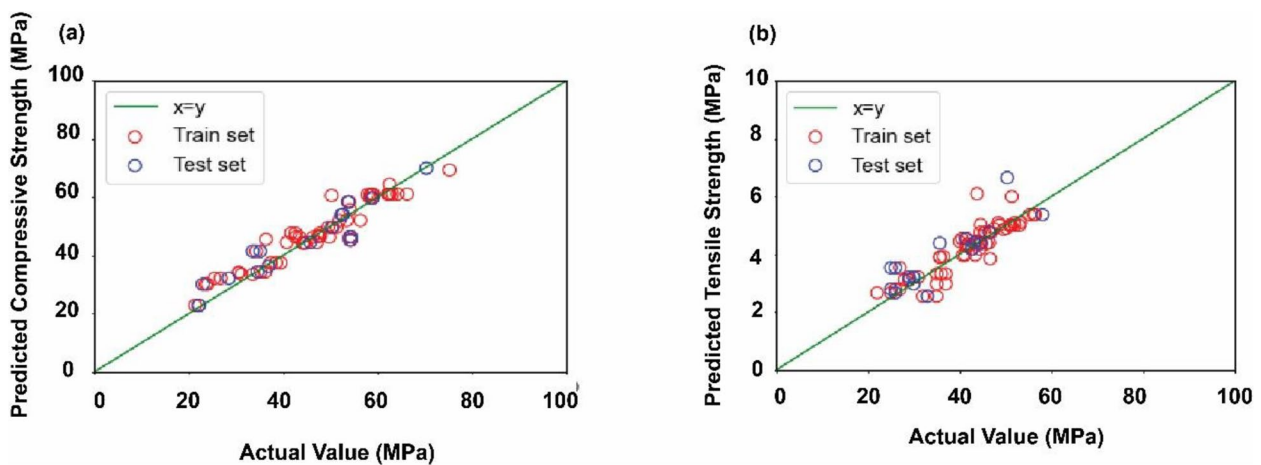

Fig. 7 **a** and **b** show the compressive strength and tensile strength using the SVM model
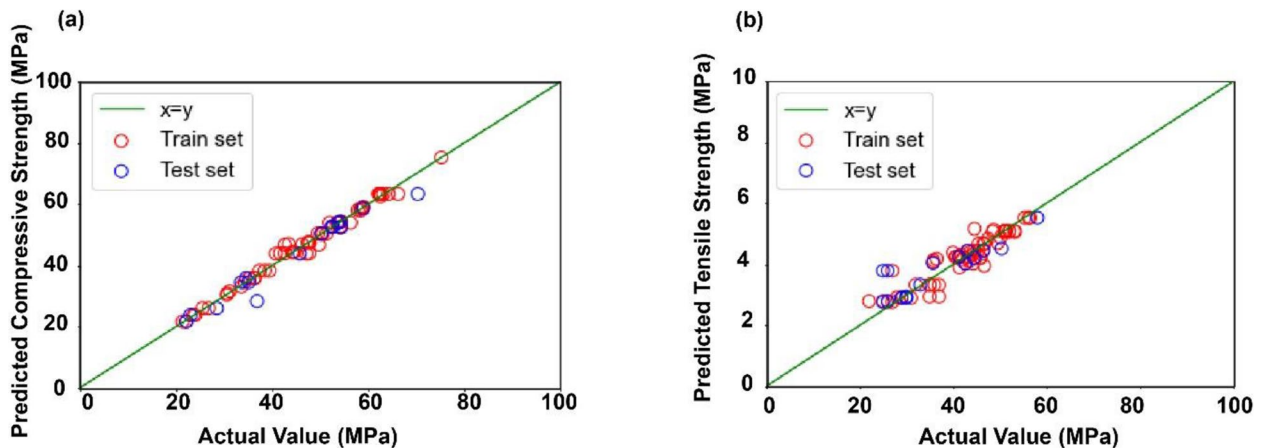

Fig. 8 **a** and **b** show the compressive strength and tensile strength using the ANN model

hyperparameters max_depth, learning_rate, and n_estimators were fine-tuned to values of 5, 0.01, and 250, respectively and 5,0.03, and 500 was chosen for the CS and TS through cross-validation, thereby accentuating the model's high-performance attributes as evidenced by the RMSE. Other parameters, such as min_child_weight=3 and reg_lambda=0.2, were meticulously set for both model. The evaluative phase deployed the testing set to ascertain the XGBoost model's empirical accuracy for predicting the CS and TS. Figure 9(a) and (b) exhibit the predicted CS and TS values attained through the XGBoost model.

Table 3 encapsulates the evaluative metrics showcasing that for CS in the training model, the R^2, MAE, MSE, and RMSE values were 0.99, 0.80, 1.33, and 1.15, respectively. For the testing model, these values morphed to 0.92, 2.29, 14.57, and 3.81, respectively. Similarly, for TS in the training model, the corresponding values were 0.98, 0.08, 0.01, and 0.10, respectively, and for the testing model, they were 0.97, 0.17, 0.11, and 0.26, respectively. Figure 8 elucidates the x=y line employed for delineating the training and testing sets within the model. The data postulates that among the cadre of machine learning models explored, XGBoost emerges as the most advanced and highly accurate model, thereby signifying its paramount potential in the predictive analysis of ECC's mechanical strength.

This discourse highlights the imperative role of optimized machine learning models, particularly XGBoost, in navigating the complex landscape of predicting mechanical strength properties of ECC, thus fostering a nexus of computational robustness and predictive precision.

Table 4 Models prediction of testing set.presents the experimental data along with the corresponding data predicted by the model to evaluate their accuracy. The errors for the CS are found to be lower for the XGBoost model compared to the SVM model. Similarly, for the TS, the error is 0.01 for the XGBoost model and 0.12 for the SVM model.

Figures 10 and 11, present a comparative analysis of the prediction performance of the machine learning (ML) models for CS and TS on both the training and testing datasets. Figure 10 exhibits that XGBoost closely aligns with the experimental datasets, demonstrating superior performance compared to SVM, in both the training and testing phases for CS. On the other hand, Fig. 11 illustrates that XGBoost yields nearly identical outcomes for both the training and testing datasets in TS. However, upon closer examination of the testing results, it becomes evident that XGBoost better aligns with the experimental datasets than SVM.

In summary, the figures visually represent the predictive capabilities of ML models for CS and TS. It is apparent that XGBoost consistently outperforms SVM, particularly when considering the testing datasets.

3.6 SHAP interpretation

A comprehensive framework for feature importance analysis (e.g., SHAP Summary Plots,) in ML is offered by SHAP, which gives specific insights into the individual and combined contributions of each feature to a model's predictions for predicting mechanical strength [106, 107]. In ML applications, this type of analysis is crucial for verifying model behavior, guaranteeing equity, and fostering transparency and feaures relations and their influnces in the strength prediction Naser et al. [108].

3.7 Generating SHAP summary plots

Figure 12 presents the mean SHAP value, which represents the average impact on the magnitude of the model
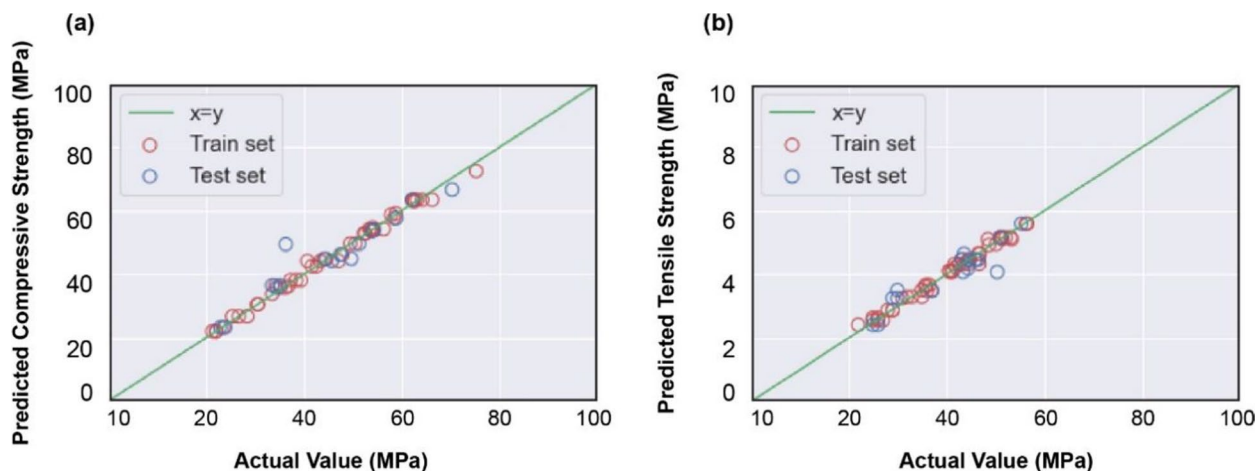

Fig. 9 **a** and **b** show the CS and TS using the XGBoost model

Table 4 Models prediction of the testing set

| No. Series	CS (MPa) Experimtal	XGBoost	SVM	TS (MPa) Experimtal	XGBoost	SVM
1.00	50.20	50.00	50.40	5.81	5.57	5.01
2.00	62.50	63.40	62.00	5.13	5.13	4.46
3.00	56.30	52.65	52.00	0.00	0.03	0.01
4.00	51.30	50.00	50.40	5.53	5.57	5.01
5.00	62.50	65.83	67.14	5.14	5.20	4.69
6.00	47.10	43.20	40.80	0.00	-0.02	-0.96
7.00	49.70	43.26	46.81	4.44	4.24	4.82
8.00	26.70	26.90	31.40	2.70	2.63	2.81
9.00	46.20	43.44	45.74	4.14	4.22	4.68
10.00	59.00	59.76	62.02	4.46	4.99	4.48
11.00	54.10	54.25	44.18	4.27	4.32	4.77
12.00	75.20	66.36	63.11	4.38	4.81	5.58
13.00	36.20	34.85	33.87	3.70	3.14	2.78
StD	12.20	11.95	11.22	1.88	1.91	2.03
Error		1.77	2.09		0.01	0.12
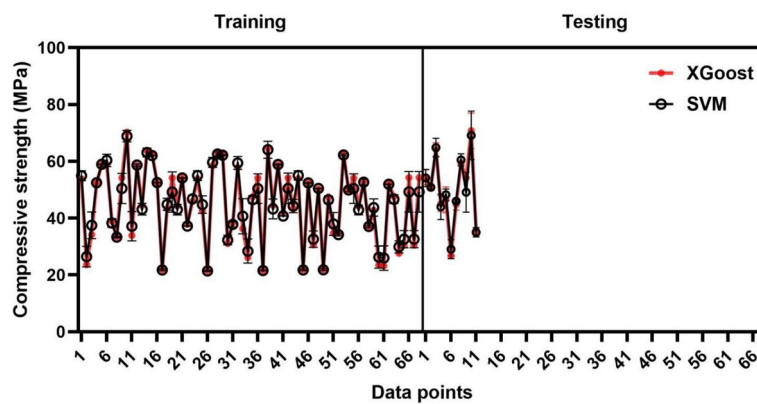
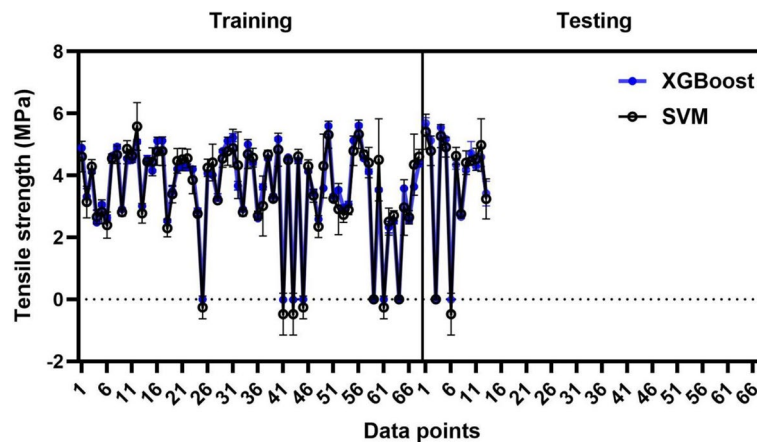**Fig. 10** The CS prediction performances of XGBoost and SVM models

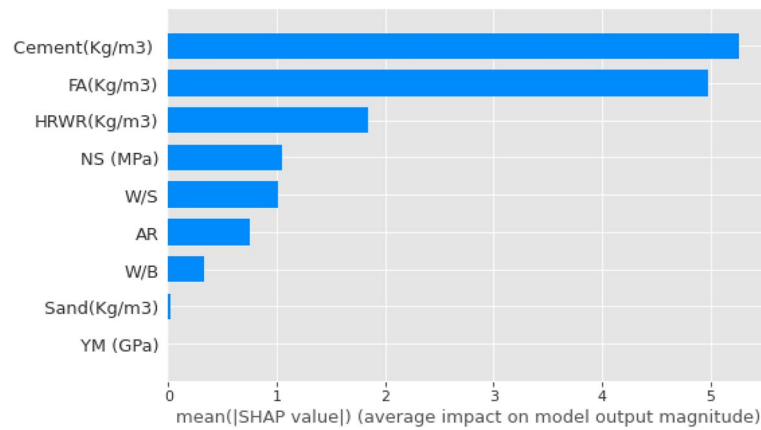**Fig. 11** The TS prediction performances of XGBoost and SVM models

Fig. 12 SHAP's value (average impact on output) for CS

output for the compressive strength (CS). Cement is a crucial parameter that contributes mainly to CS, as evidenced by its highest SHAP value. Conversely, sand has the lowest SHAP value, indicating its minimal contribution to CS. Further, FA is another dominant parameter affecting CS, while the remaining ingredients (HRWR, NS, W/S, AR, W/B) have comparatively less effect on CS.

Although Fig. 12 provides an overview of the significance of different features, it does not describe the relationship between the features and whether the outputs are influenced positively or negatively. To address this limitation, the SHAP summary plot is employed to determine the correlation between the parameters and CS (Fig. 12). Each point on the summary plot includes all the details of a shapley value corresponding to the features. By using SHAP, the global mean can be compared against the model's output to determine a specific

reason, and the features are ranked according to their importance. Additionally, a dotted line is used to indicate the negative (left) and positive (right) effects of the features, thereby illustrating how they impact both sides. The feature values are plotted as dotted lines on the x-axis and y-axis in the model to emphasize their importance.

Figure 13 demonstrates that FA is highly correlated with the CS of PVA-ECC. The dotted line for FA turns towards the negative (left) side, indicating that an increase in FA significantly reduces the CS of PVA-ECC. On the other hand, cement, NS, and AR advance towards the right side, positively influencing CS, as shown in Fig. 13. In contrast, the addition of HRWR in the mixture leads to a decrease in the CS of PVA-ECC, while adding or removing W/B, sand, and YM does not have any significant influence on CS.

Fig. 13 SHAP's summary plot for the XGBoost model for CS

The SHAP summary plot for the splitting tensile strength (TS) of the PVA-ECC mixture, as depicted in Figs. 14 and 15, presents the impact on the model output. Notably, the feature with the highest SHAP value in the plot is NS, indicating its significant influence on TS, although its impact on CS is not considerable. Similarly, the other parameters show similar effects on TS, implying that the substances that result in higher SHAP values for TS exhibit opposite trends for CS. Specifically, Figs. 14 and 15 reveal that NS, YM, and AR positively affect TS, while HRWR, cement, and W/B have a negative impact on TS, signifying that an increase in these substances will decrease TS.

Furthermore, Fig. 15 demonstrates that sand has both positive and negative influences on the resulting TS of the concrete material. It is worth noting that existing literature has shown both negative and positive impacts

on the TS of concrete [109–112]. Moreover, the nominal strength and length of PVA fiber have been found to significantly positively impact the development of TS in concrete materials [113–115].

3.8 SHAP feature instances

The SHAP instance values for the compressive strength (CS) and tensile strength (TS) of PVA-ECC are illustrated in Figs. 16 and 17, respectively. As discussed earlier, FA and Cement significantly impact CS, while NS and HRWR greatly influence TS. The effects of the other parameters are less noticeable. Figure 16 presents the instances where FA positively impacts CS in datasets 40 to 80, whereas it negatively impacts datasets 1 to 40. Similar trends can be observed for cement.

On the other hand, Fig. 17 shows that the impact of HRWR is positive in datasets 40 to 80, but it has a higher
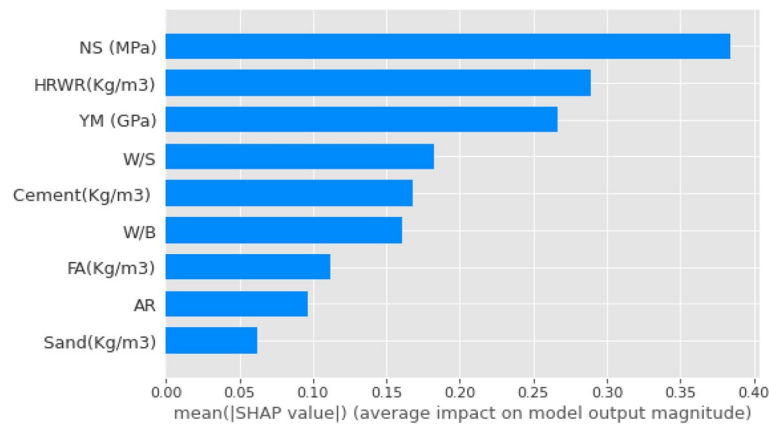

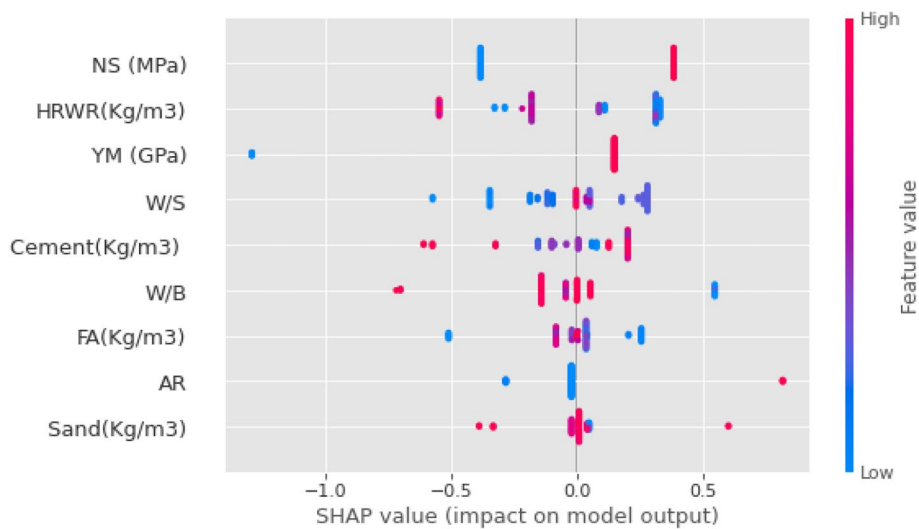**Fig. 14** SHAP's value (average impact on output) for TS

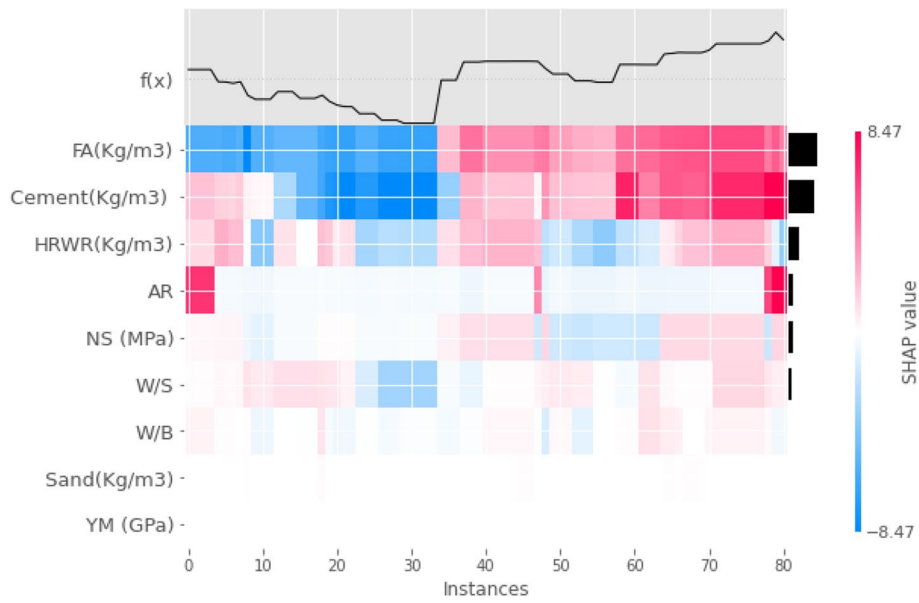**Fig. 15** SHAP's summary plot for the XGBoost model for TS
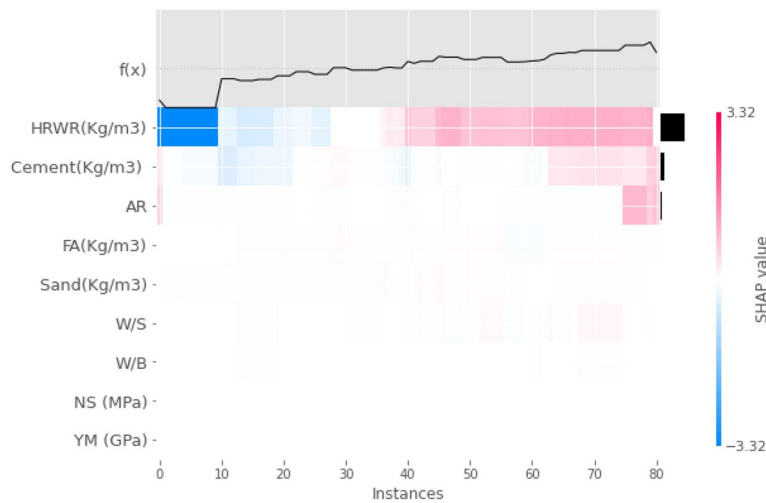
Fig. 16 SHAP's instances for the CS

Fig. 17 SHAP's instances for the TS

negative intensity in datasets 1 to 10. AR exhibits higher intensities in datasets 75 to 80. These findings help us understand the specific data points significantly influencing the model's accuracy. In future investigations, focusing on these data points or mixtures may help improve the strength of the mixtures.

3.9 Interaction values

The SHAP interaction values, also known as pairwise SHAP values, provide insights into the interaction between pairs of features in a model's prediction. These values help us understand how one feature's presence

or absence can influence another feature's effect. In the interaction value plot, red indicates high values for both features, while blue represents low values for both features. The right side of the dots represents a positive impact, while the left side represents a negative effect on the feature's value.

Figure 18 illustrates the high correlation between cement and the CS of PVA-ECC. The dotted line for cement is approaching the positive (right) side, indicating that an increase in cement content significantly increases the CS of PVA-ECC. Conversely, FA and HRWR are approaching the left side, suggesting a negative impact on
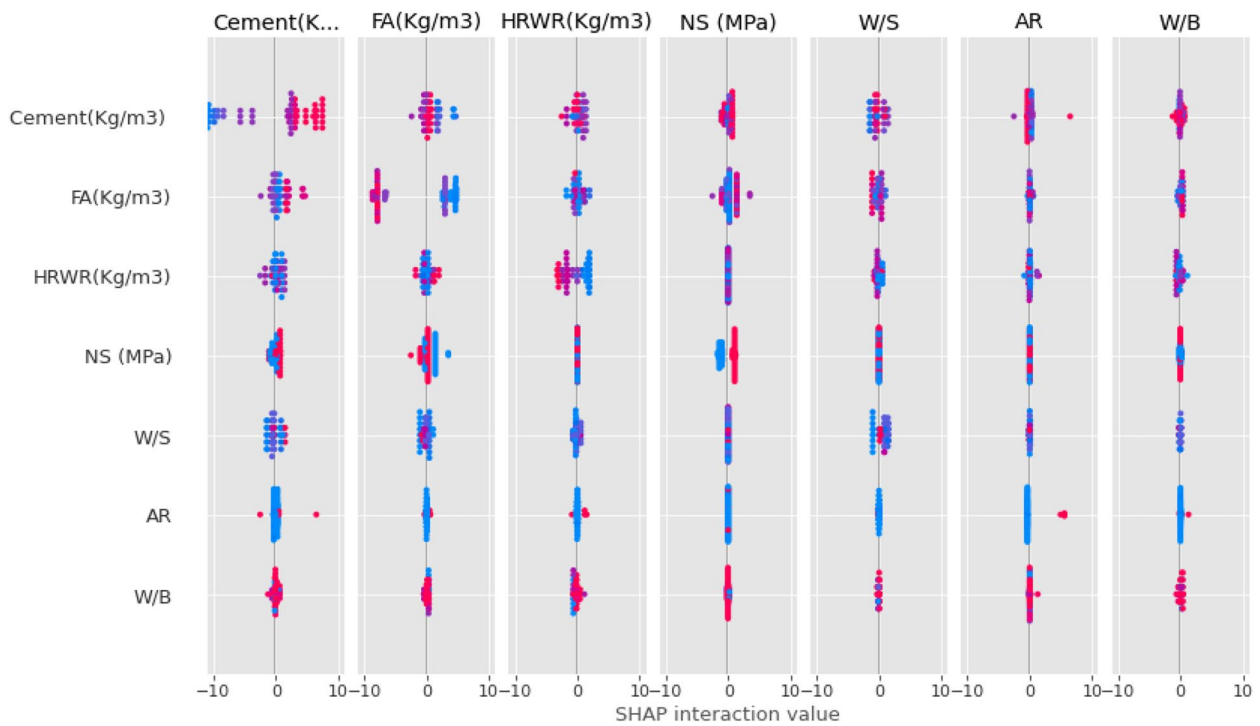
Fig. 18 SHAP interaction value for CS

CS, as depicted in Fig. 18, while increasing the NS and W/B content in the mixture will lead to an increase in the CS of PVA-ECC.

The interaction values for several features used to predict the TS are shown in Fig. 19. The SHAP interaction value indicates that the increase of NS content significantly increases TS while opposite trends are shown by HRWR and W/B turning to the negative (left) side. The interaction value of YM turns to the right side a little, which means an increase in YM will raise TS, and the changes of other parameters (YM, W/S, FA) do not affect TS as the SHAP interaction values remain in the middle ranges. Cement has high and low effects, indicating that it positively impacts TS, as shown in Fig. 19.
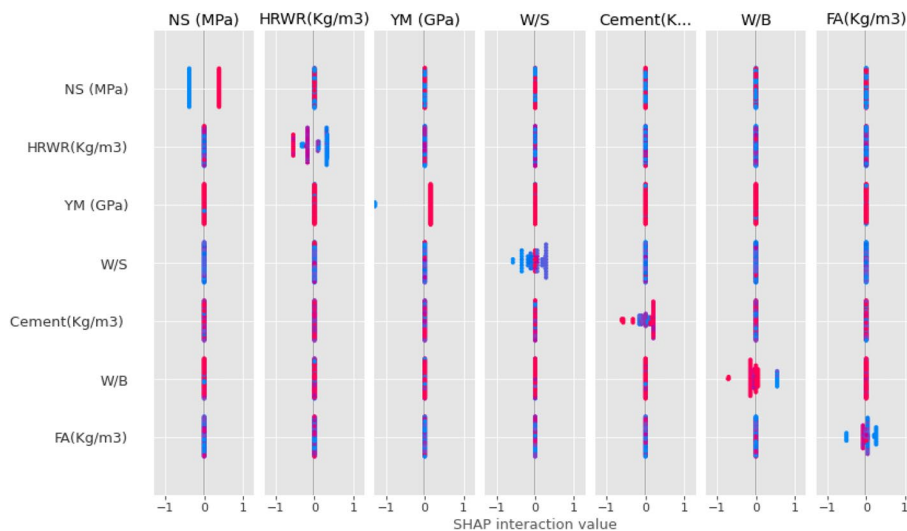

Fig. 19 SHAP interaction value for TS

3.10 Dependent plot

A SHAP dependence plot is utilized to examine the relationship between the SHAP value of features on the y-axis and their corresponding data on the x-axis. These plots are constructed to provide a clearer understanding of SHAP interpretations, as demonstrated in Fig. 20. The dependence plot illustrates how two attributes interact with each other to influence the predicted outcome of the SHAP model.

In Fig. 20(a), it is evident that the CS significantly correlates with the W/S within the range of 0.70 to 0.75, particularly when the FA is between 900 to 1000 kg/m^3 representing the higher CS. Conversely, in Fig. 20(b), a negative linear correlation is observed between the W/B ratio and the cement at a ratio of 0.265. Figure 20(c) and (d) show a negative linear correlation with the CS.

Furthermore, Fig. 21(b) reveals that an FA of 700 kg/m^3 with an HRWR of 7 kg/m^3 has a positive impact.

Conversely, increasing the HRWR to 8 or 9 kg/m^3 leads to a lower correlation with an FA of 700 kg/m^3.

In Fig. 21(d), it is shown that with an AR of 200, the cement content exhibits a strong correlation when it is less than 230 kg/m^3. In contrast, with an AR of 300, the cement content within the range of 400 to 500 kg/m^3 demonstrates a higher correlation. However, increasing the cement content beyond this range results in a decrease in the TS.

4 Conclusions

This paper demonstrates an interpretable ML model for predicting PVA-ECC CS and TS. For training and testing, data from 81 mixed samples were collected. The XGBoost method constructed a high-accuracy predictive model across all ML models. Random selections of 80% and 20% is used for this purpose. The hyperparameter optimization has been used using GridSearchCV. The SHAP algorithm is then employed to explain the model's notable
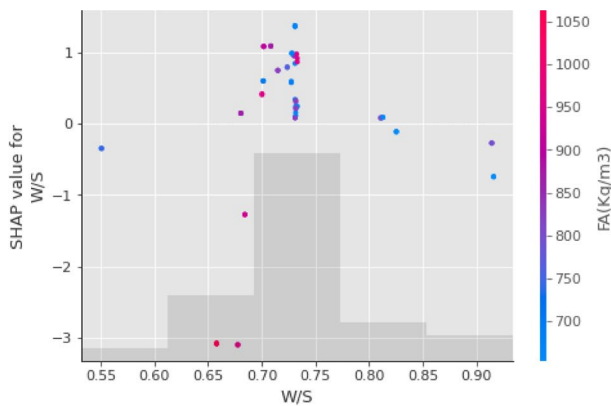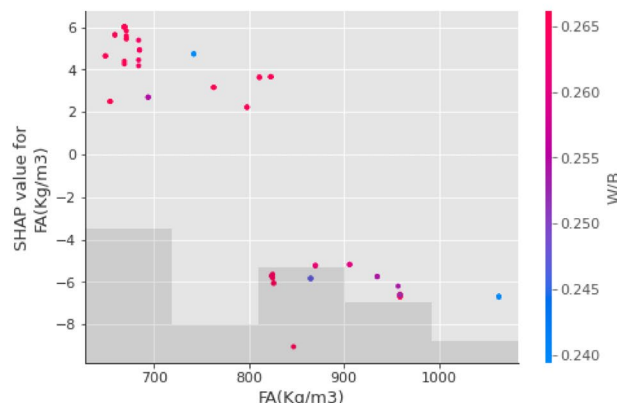

(a) Correlation between W/S & FA
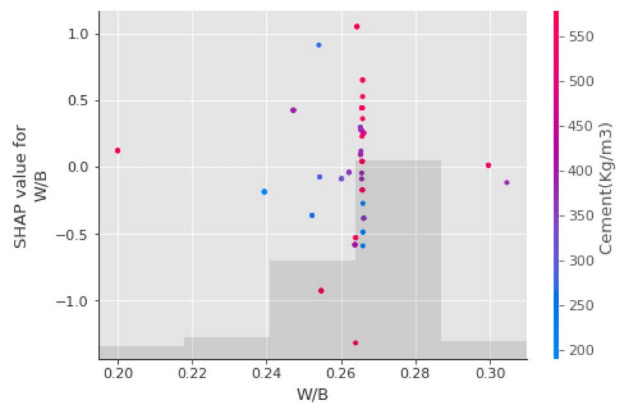
(b) Correlation between W/B & cement
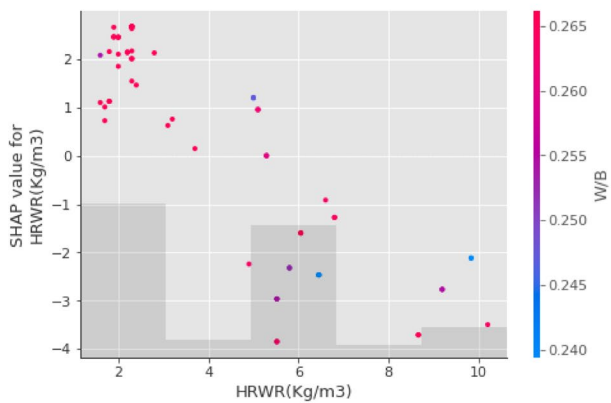
(c) Correlation between HRWR and W/B

(d) Correlation between FA and W/B

Fig. 20 The SHAP dependence plot for the CS

(a) Correlation between W/S and HRWR

(b) Correlation between FA and HRWR

(c) Correlation between W/B and HRWR
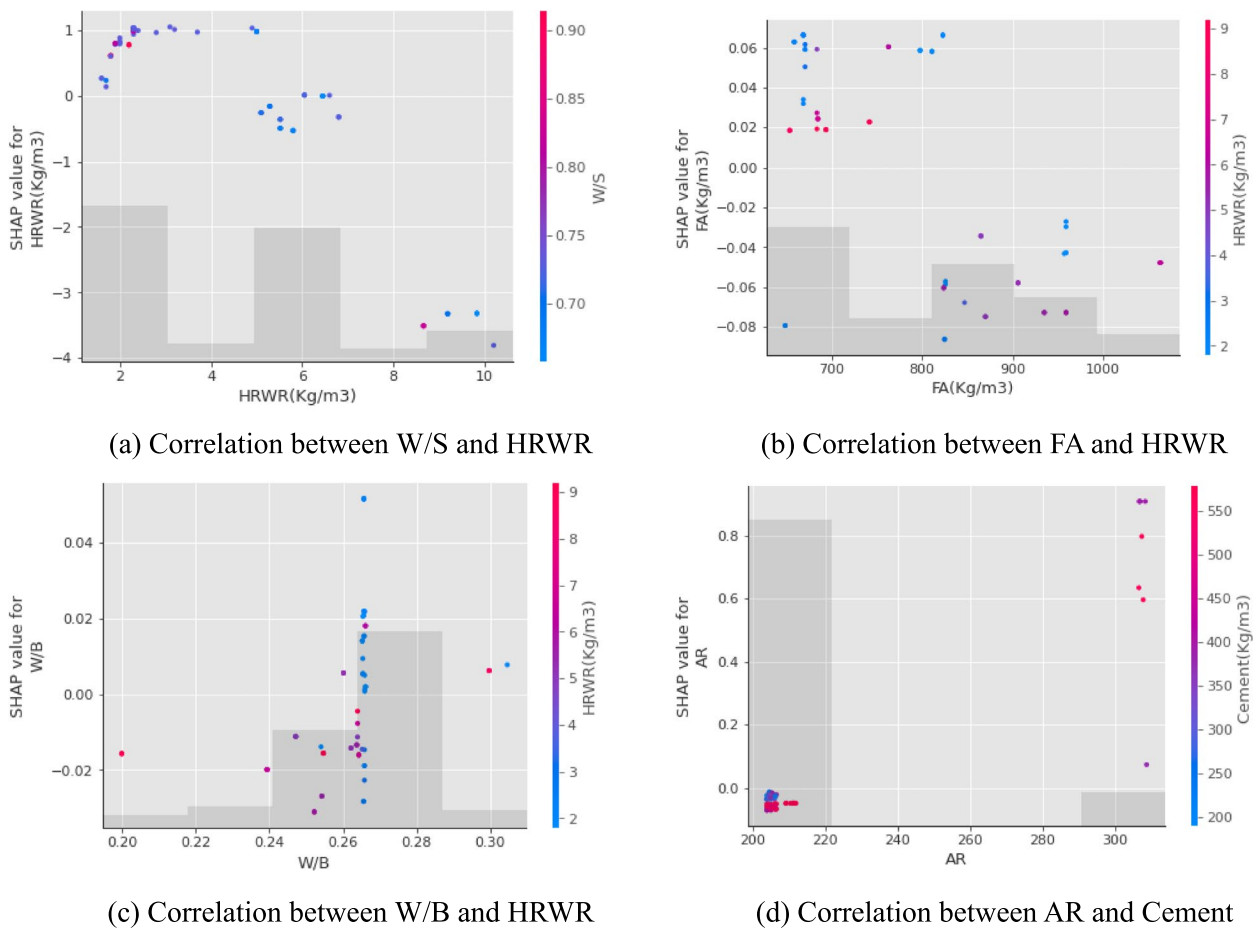

(d) Correlation between AR and Cement

Fig. 21 The SHAP dependence plot for theTS

features and their complex correlations. The suggested model's prediction outcomes are compared against earlier research and chosen best ML algorithm. The following conclusions can draw from the present study:

1. The established XGBoost demonstrated extremely high accuracy in predicting the CS and TS of PVA-ECC concrete. The performance for R^2 shows the CS and TS of 0.99 and 0.98.
2. The XGBoost model is compared to various machine learning models such as RF, SVM, and ANN. It discovered that ML methods (XGBoost, RF) beat other methods (ANN and SVM), with XGBoost achieving the best overall performance.
3. The results of SHAP interpretation shows that cement and FA are the crucial parameter that mainly affects CS, as evidenced by their highest SHAP value. The increment of cement significantly increases CS while opposite trends (reduces CS with the increase of FA) are found for FA. Although changes in NS do not influence CS, it significantly increases TS if NS

is increased. Other parameters that positively affect CS have adverse effects on TS. The SHAP features instances suggests that FA and cement positively affect CS when the instance value exceeds 35. For HRWR, the impact is average on CS and significant for TS in the mentioned instances ranges. Similar trends are also found in SHAP interaction analysis for cement and FA in cases of CS, while for TS, HRWR is found to have negative impacts, although NS has positive impacts

4. SHAP dependency plots demonstrate that the correlation between FA and W/S can positively (increase) affect CS in a particular range (W/S is 0.70–0.75, FA is 900–100 kg/m^3). In contrast, other correlations show negative trends for CS. FA and HRWR negatively impact TS when FA is 700 kg/m^3 and HRWR is 7 kg/m^3.
5. The influence of cement, and NS on TS gives a positive trend, which shows that increasing those features in a certain limitation can give high performance. The sand size is not specified in the database; it shows

both positive and negative trends on TS. A specific database for same-size sand can give an accurate prediction for improving the TS using XGBoost SHAP. Therefore, more research should need to specify the sand's influence on TS, It will need more involvement in the mechanical performance of fiber-based concrete in the future.

6. SHAP values findings revealed how a characteristic affects CS and TS prediction straightforwardly and precisely. By setting up meaningful value ranges for different attributes, this data can be used to build superior CS and TS models. It is also under consideration that further improvements are required to make a perfect database; adding more data related to PVA-ECC can help accurate prediction.

Finally it can concluded that, the XGBoost model can quickly calculate concrete's CS and TS in various mixed proportions, and can be utilized to see if the designed mix proportion can meet the target strength requirements. It can help engineers undertake approval analysis during the building stage and safety analysis during the service stage by forecasting the needed CS and TS at 28 days. Moreover, the model can also help to deeply understand the mechanical interaction inside the mixture to find out their intercorrelation, which helps to make a new mixture design with a better understanding. The SHAP value is extraordinarily complex, and new features need to explore in different fiber-reinforced concrete. This research helps understand the features' importance, adopt the optimized mixture properties, and provide high strength and plays a crucial role in promoting green and low-carbon development, aligning with global efforts towards sustainability and environmental conservation. These aspects will be studied based on the work presented in this paper shortly in fiber-based 3D concrete, where the mixtures are costly and complicated.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1007/s44242-024-00042-w.

Supplementary Material 1.

Authors' contributions
All authors participated in the conceptualization and design of the study. The tasks of material preparation, data collection, and analysis were carried out by Md Nasir Uddin The initial draft of the manuscript was composed by Al-amin, and Shameem Hossain. All authors providing comments on earlier versions of the manuscript. All authors have reviewed and approved the final version of the manuscript.

Author details
[1]Ingram School of Engineering, Texas State University, San Marcos, TX 78666, USA. [2]Department of Bridge Engineering, College of Civil Engineering, Tongji University, Shanghai, China. [3]Department of Disaster Mitigation for Structures, Collage of Civil Engineering, Tongji University, Siping Road 1239, Shanghai 200092, China.

References
1. Singh, S. B., Munjal, P., & Thammishetti, N. (2015). Role of water/cement ratio on strength development of cement mortar. *Journal of Building Engineering, 4*, 94–100. https://doi.org/10.1016/j.jobe.2015.09.003
2. Molero, M., Segura, I., Izquierdo, M. A. G., Fuente, J. V., & Anaya, J. J. (2009). Sand/cement ratio evaluation on mortar using neural networks and ultrasonic transmission inspection. *Ultrasonics, 49*, 231–237. https://doi.org/10.1016/j.ultras.2008.08.006
3. Mahdinia, S., Eskandari-Naddaf, H., & Shadnia, R. (2017). Effect of Main factors on fracture mode of mortar, a graphical study. *Civil Engineering Journal, 3*(10), 897-90310. https://doi.org/10.28991/cej-030923
4. Haach, V. G., Vasconcelos, G., & Loureno, P. B. (2011). Influence of aggregates grading and water/cement ratio in workability and hardened properties of mortars. *Construction Building Materials, 25*, 2980–2987. https://doi.org/10.1016/j.conbuildmat.2010.11.011
5. Mukharjee, B. B., & Barai, S. V. (2014). Assessment of the influence of Nano-Silica on the behavior of mortar using factorial design of experiments. *Construction and Building Materials, 68*, 416–425. https://doi.org/10.1016/j.conbuildmat.2014.06.074
6. Kan, L. L., Shi, R. X., & Zhu, J. (2019). Effect of fineness and calcium content of fly ash on the mechanical properties of Engineered Cementitious Composites (ECC). *Construction and Building Materials, 209*, 476–484. https://doi.org/10.1016/j.conbuildmat.2019.03.129
7. Mahdinia, S., Eskandari-Naddaf, H., & Shadnia, R. (2019). Effect of cement strength class on the prediction of compressive strength of cement mortar using GEP method. *Construction and Building Materials, 198*, 27–41. https://doi.org/10.1016/j.conbuildmat.2018.11.265
8. Yu, J., & Leung, C. K. Y. (2017). Strength improvement of strain-hardening cementitious composites with ultrahigh-volume fly ash. *Journal of Materials in Civil Engineering, 29*, 05017003. https://doi.org/10.1061/(asce)mt.1943-5533.0001987
9. Yuan, F., Pan, J., & Leung, C. K. Y. (2013). Flexural behaviors of ECC and concrete/ECC composite beams reinforced with basalt fiber-reinforced polymer. *Journal of Composites for Construction, 17*, 591–602. https://doi.org/10.1061/(asce)cc.1943-5614.0000381
10. Xu, M., Song, S., Feng, L., Zhou, J., Li, H., & Li, V. C. (2021). Development of basalt fiber engineered cementitious composites and its mechanical properties. *Construction and Building Materials, 266*, 121173. https://doi.org/10.1016/j.conbuildmat.2020.121173
11. Yu, K., Yu, J., & Lu, Z. (2018). Mechanical characteristics of ultra high performance strain hardening cementitious composites. In

Strain-hardening cement-based composites: SHCC4 4 (pp. 230–237). Springer Netherlands. https://doi.org/10.1007/978-94-024-1194-2_27

12. Do Yun, H. (2013). Flexural behavior and crack-damage mitigation of plain concrete beam with a strain-hardening cement composite (SHCC) layer at tensile region. *Composites Part B Engineering, 45,* 377–387. https://doi.org/10.1016/j.compositesb.2012.05.053

13. ChiaHwan, Y., & JianBo, H. (2014). The mechanical behavior of fiber reinforced PP ECC beams under reverse cyclic loading. Advances in Materials Science and Engineering, *2014*(1), 159790. https://doi.org/10.1155/2014/159790

14. Zawawi, N. A., Oh, C. L., Lee, S. W., Zain, M. R. M., & Yahya, N. A. (2021). Mechanical properties of engineered cementitious composite (ECC): An overview. In *ICSDEMS 2019: Proceedings of the International Conference on Sustainable Design, Engineering, Management and Sciences* (pp. 259–264). Springer Singapore. https://doi.org/10.1007/978-981-15-3765-3_26

15. Zhu, H., Yu, K., & Li, V. C. (2021). Sprayable engineered cementitious composites (ECC) using calcined clay limestone cement (LC3) and PP fiber. *Cement and Concrete Composites, 115,* 103868. https://doi.org/10.1016/j.cemconcomp.2020.103868

16. Mousavinejad, S. H. G., & Alemi, M. P. (2022). Micro-structural and mechanical properties of PVA fiber reinforced engineered cementitious composite incorporating natural and artificial pozzolanic materials under different temperatures. *Construction and Building Materials, 346,* 128180.

17. Wang, Q., Yi, Y., Ma, G., & Luo, H. (2019). Hybrid effects of steel fibers, basalt fibers and calcium sulfate on mechanical performance of PVA-ECC containing high-volume fly ash. *Cement and Concrete Composites, 97,* 357–368. https://doi.org/10.1016/j.cemconcomp.2019.01.009

18. Fischer, G., & Li, V. C. (2002). Influence of matrix ductility on tension-stiffening behavior of steel reinforced engineered cementitious composites (ECC). *ACI Structural Journal, 99,* 104–111. https://doi.org/10.14359/11041

19. Kang, M. C., Yoo, D. Y., & Gupta, R. (2021). Machine learning-based prediction for compressive and flexural strengths of steel fiber-reinforced concrete. *Construction and Building Materials, 266,* 121117. https://doi.org/10.1016/j.conbuildmat.2020.121117

20. Singh, M., Saini, B., & Chalak, H. (2021). Evaluation of cost-effective hybrid fiber reinforced ECC. *Sadhana - Academic Proc Engineering and Science, 46,* 109. https://doi.org/10.1007/s12046-021-01637-w

21. Hossain, S., Uddin, M. N., & Yan, K. (2024). Prediction of the mechanical performance of polyethylene fiber-based engineered cementitious composite (PE-ECC). *Low-Carbon Material Green Construction, 2*(1), 9. https://doi.org/10.1007/s44242-024-00040-y

22. Haque, M. A., Dai, J. G., & Zhao, X. L. (2024). Magnesium cements and their carbonation curing: a state-of-the-art review. *Low-carbon Materials and Green Construction, 2*(1), 2. https://doi.org/10.1007/s44242-023-00033-3

23. Uddin, M. N., Shanmugasundaram, N., Praveenkumar, S., & Li, L. Z. (2024). Prediction of compressive strength and tensile strain of engineered cementitious composite using machine learning. *International Journal of Mechanics and Materials in Design*, 1–46. https://doi.org/10.1007/s10999-023-09695-0

24. Shoji, D., He, Z., Zhang, D., & Li, V. C. (2022). The greening of engineered cementitious composites (ECC): A review. *Construction and Building Materials, 327,* 126701. https://doi.org/10.1016/j.conbuildmat.2022.126701

25. Apostolopoulou, M., Armaghani, D. J., Bakolas, A., Douvika, M. G., Moropoulou, A., & Asteris, P. G. (2019). Compressive strength of natural hydraulic lime mortars using soft computing techniques. *Procedia Structural Integrity, 17,* 914–923. https://doi.org/10.1016/j.prostr.2019.08.122

26. Apostolopoulou, M., Asteris, P. G., Armaghani, D. J., Douvika, M. G., Lourenço, P. B., Cavaleri, L., Bakolas, A., & Moropoulou, A. (2020). Mapping and holistic design of natural hydraulic lime mortars. *Cement and Concrete Research, 136,* 106167. https://doi.org/10.1016/j.cemconres.2020.106167

27. Asteris, P. G., & Mokos, V. G. (2020). Concrete compressive strength using artificial neural networks. *Neural Computing and Applications, 32,* 11807–11826. https://doi.org/10.1007/s00521-019-04663-2

28. Uddin, M. N., Li, L. Z., Ahmed, A., & Almajhali, K. Y. M. (2022). Prediction of PVA fiber effect in Engineered Composite cement (ECC) by Artificial neural Network (ANN). *Materials Today: Proceedings, 65,* 537–542. https://doi.org/10.1016/j.matpr.2022.03.088

29. Snellings, R., Mertens, G., & Elsen, J. (2012). Supplementary cementitious materials. *Reviews in mineralogy and geochemistry, 74*(1), 211–278. https://doi.org/10.2138/rmg.2012.74.6

30. Oey, T., Kumar, A., Bullard, J. W., Neithalath, N., & Sant, G. (2013). The filler effect: The influence of filler content and surface area on cementitious reaction rates. *Journal of the American Ceramic Society, 96,* 1978–1990. https://doi.org/10.1111/jace.12264

31. Quercia, G., Lazaro, A., Geus, J. W., & Brouwers, H. J. H. (2013). Characterization of morphology and texture of several amorphous nano-silica particles used in concrete. *Cement and Concrete Composites, 44,* 77–92. https://doi.org/10.1016/j.cemconcomp.2013.05.006

32. Kong, D., Su, Y., Du, X., Yang, Y., Wei, S., & Shah, S. P. (2013). Influence of nano-silica agglomeration on fresh properties of cement pastes. *Construction and Building Materials, 43,* 557–562. https://doi.org/10.1016/j.conbuildmat.2013.02.066

33. Güneyisi, E., Gesoǧlu, M., Karaoǧlu, S., & Mermerdaş, K. (2012). Strength, permeability and shrinkage cracking of silica fume and metakaolin concretes. *Construction and Building Materials, 34,* 120–130. https://doi.org/10.1016/j.conbuildmat.2012.02.017

34. Siddique, R., & Kadri, E.-H. (2011). Effect of metakaolin and foundry sand on the near surface characteristics of concrete. *Construction and Building Materials, 25,* 3257–3266.

35. San Nicolas, R., Cyr, M., & Escadeillas, G. (2014). Performance-based approach to durability of concrete containing flash-calcined metakaolin as cement replacement. *Construction and Building Materials, 55,* 313–322. https://doi.org/10.1016/j.conbuildmat.2014.01.063

36. Van Tuan, N., Ye, G., Van Breugel, K., Fraaij, A. L. A., & Bui, D. D. (2011). The study of using rice husk ash to produce ultra high performance concrete. *Construction and Building Materials, 25,* 2030–2035. https://doi.org/10.1016/j.conbuildmat.2010.11.046

37. Uddin, M. N., Li, L. Z., Deng, B. Y., & Ye, J. (2023). Interpretable XGBoost–SHAP machine learning technique to predict the compressive strength of environment-friendly rice husk ash concrete. *Innovative Infrastructure Solutions, 8*(5), 147.

38. Yu, R., Spiesz, P., & Brouwers, H. J. H. (2015). Development of an eco-friendly Ultra-High Performance Concrete (UHPC) with efficient cement and mineral admixtures uses. *Cement and Concrete Composites, 55,* 383–394. https://doi.org/10.1016/j.cemconcomp.2014.09.024

39. Madadi, A., Eskandari-Naddaf, H., & Gharouni-Nik, M. (2017). Lightweight ferrocement matrix compressive behavior: experiments versus finite element analysis. *Arabian Journal for Science and Engineering, 42,* 4001–4013. https://doi.org/10.1007/s13369-017-2557-4

40. Korouzhdeh, T., Eskandari-Naddaf, H., & Gharouni-Nik, M. (2017). An improved ant colony model for cost optimization of composite beams. *Applied Artificial Intelligence, 31,* 44–63. https://doi.org/10.1080/08839514.2017.1296681

41. Naser, M. Z. (2023). Machine learning for civil and environmental engineers: *A practical approach to data-driven analysis, explainability, and causality*. John Wiley & Sons. https://www.wiley.com/en-us/Machine+Learning+for+Civil+and+Environmental+Engineers%3A+A+Practical+Approach+to+Data+Driven+Analysis%2C+Explainability%2C+and+Causality-p-9781119897613.

42. Tapeh, A. T. G., & Naser, M. Z. (2023). Artificial intelligence, machine learning, and deep learning in structural engineering: a scientometrics review of trends and best practices. *Archives of Computational Methods in Engineering, 30*(1), 115–159. https://doi.org/10.1007/s11831-022-09793-w

43. Ali, S., Kenari, J., & Mosalmannejad, H. (2013). A Committee Machine with Different Adaptive Neuro Fuzzy Inference System for Water Saturation Prediction *Universal Journal of Petroleum Sciences, 1,* 30–38.

44. Khademi, F., Akbari, M., Jamal, S. M., & Nikoo, M. (2017). Multiple linear regression, artificial neural network, and fuzzy logic prediction of 28 days compressive strength of concrete. *Frontiers of Structural and Civil Engineering, 11,* 90–99. https://doi.org/10.1007/s11709-016-0363-9

45. Özcan, F., Atiş, C. D., Karahan, O., Uncuoǧlu, E., & Tanyildizi, H. (2009). Comparison of artificial neural network and fuzzy logic models for prediction of long-term compressive strength of silica fume concrete.

Advances in Engineering Software, 40, 856–863. https://doi.org/10.1016/j.advengsoft.2009.01.005

46. Benemaran, R. S., & Esmaeili-Falak, M. (2020). Optimization of cost and mechanical properties of concrete with admixtures using MARS and PSO. *Computers and Concrete, 26*, 309–316. https://doi.org/10.12989/cac.2020.26.4.309

47. Shariati, M., Mafipour, M. S., Mehrabi, P., Bahadori, A., Zandi, Y., Salih, M. N. A., Nguyen, H., Dou, J., Song, X., & Poi-Ngian, S. (2019). Application of a hybrid artificial neural network-particle swarm optimization (ANN-PSO) model in behavior prediction of channel shear connectors embedded in normal and high-strength concrete. *Applied sciences, 9*(24), 5534. https://doi.org/10.3390/app9245534

48. Eskandari-Naddaf, H., & Kazemi, R. (2017). ANN prediction of cement mortar compressive strength, influence of cement strength class. *Construction and Building Materials, 138*, 1–11. https://doi.org/10.1016/j.conbuildmat.2017.01.132

49. Uddin, M. N., Li, L.-Z., Khan, R. K. M., Shahriar, F., & Sob, L. W. T. (2021). Axial capacity prediction of concrete-filled steel tubular short members using multiple linear regression and artificial neural network. *Materials Science Forum*, *1047*, 220–226. https://doi.org/10.4028/www.scientific.net/MSF.1047.220

50. Topçu, I. B., & Saridemir, M. (2008). Prediction of rubberized mortar properties using artificial neural network and fuzzy logic. *Journal of Materials Processing Technology, 199*, 108–118. https://doi.org/10.1016/j.jmatprotec.2007.08.042

51. Zhang, Y., Aslani, F., & Lehane, B. (2021). Compressive strength of rubberized concrete: Regression and GA-BPNN approaches using ultrasonic pulse velocity. *Construction and Building Materials, 307*, 124951. https://doi.org/10.1016/j.conbuildmat.2021.124951

52. Cruzado, M. P., & Camino, J. L. (2018, November). GA-Based Optimization of Normal Weight Concrete Compressive Strength with Maranta Arundinacea. In *2018 IEEE 10th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management (HNICEM)* (pp. 1–3). IEEE. https://doi.org/10.1109/HNICEM.2018.8666245.

53. Chou, J. S., Chiu, C. K., Farfoura, M., & Al-Taharwa, I. (2011). Optimizing the prediction accuracy of concrete compressive strength based on a comparison of data-mining techniques. *Journal of Computing in Civil Engineering, 25*, 242–253. https://doi.org/10.1061/(asce)cp.1943-5487.0000088

54. Shahmansouri, A. A., Bengar, H. A., & Ghanbari, S. (2020). Compressive strength prediction of eco-efficient GGBS-based geopolymer concrete using GEP method. *Journal of Building Engineering, 31*, 101326. https://doi.org/10.1016/j.jobe.2020.101326

55. Gomaa, E., Han, T., ElGawady, M., Huang, J., & Kumar, A. (2021). Machine learning to predict properties of fresh and hardened alkali-activated concrete. *Cement and Concrete Composites, 115*, 103863. https://doi.org/10.1016/j.cemconcomp.2020.103863

56. Chen, H., Deng, T., Du, T., Chen, B., Skibniewski, M. J., & Zhang, L. (2022). An RF and LSSVM–NSGA-II method for the multi-objective optimization of high-performance concrete durability. *Cement and Concrete Composites, 129*, 104446. https://doi.org/10.1016/j.cemconcomp.2022.104446

57. Khashman, A., & Akpinar, P. (2017). Non-destructive prediction of concrete compressive strength using neural networks. *Procedia Computer Science, 108*, 2358–2362. https://doi.org/10.1016/j.procs.2017.05.039

58. Hossain, K. M. A., Anwar, M. S., & Samani, S. G. (2018). Regression and artificial neural network models for strength properties of engineered cementitious composites. *Neural Computing and Applications, 29*, 631–645. https://doi.org/10.1007/s00521-016-2602-3

59. Altayeb, M., Wang, X., & Musa, T. H. (2021). An ensemble method for predicting the mechanical properties of strain hardening cementitious composites. *Construction Building and Materials, 286*, 122807. https://doi.org/10.1016/j.conbuildmat.2021.122807

60. Shi, L., Lin, S. T. K., Lu, Y., Ye, L., & Zhang, Y. X. (2018). Artificial neural network based mechanical and electrical property prediction of engineered cementitious composites. *Construction and Building Materials, 174*, 667–674. https://doi.org/10.1016/j.conbuildmat.2018.04.127

61. Louppe, G. (2014). Understanding random forests: From theory to practice. *arXiv preprint arXiv: 1407.7502.* http://arxiv.org/abs/1407.7502.

62. Zhang, J., Ma, G., Huang, Y., Aslani, F., & Nener, B. (2019). Modelling uniaxial compressive strength of lightweight self-compacting concrete using random forest regression. *Construction and Building Materials, 210*, 713–719. https://doi.org/10.1016/j.conbuildmat.2019.03.189

63. Chen, T., & Guestrin, C. (2016, August). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (pp. 785–794). https://doi.org/10.1145/2939672.2939785

64. Rahman, J., Ahmed, K. S., Khan, N. I., Islam, K., & Mangalathu, S. (2021). Data-driven shear strength prediction of steel fiber reinforced concrete beams using machine learning approach. *Engineering Structures, 233*, 111743. https://doi.org/10.1016/j.engstruct.2020.111743

65. Xu, J. G., Chen, S. Z., Xu, W. J., & Shen, Z. S. (2021). Concrete-to-concrete interface shear strength prediction based on explainable extreme gradient boosting approach. *Construction and Building Materials, 308*, 125088. https://doi.org/10.1016/j.conbuildmat.2021.125088

66. Ali, Y., Hussain, F., Irfan, M., & Buller, A. S. (2021). An eXtreme Gradient Boosting model for predicting dynamic modulus of asphalt concrete mixtures. *Construction and Building Materials, 295*, 123642. https://doi.org/10.1016/j.conbuildmat.2021.123642

67. Nguyen-Sy, T., Wakim, J., To, Q. D., Vu, M. N., Nguyen, T. D., & Nguyen, T. T. (2020). Predicting the compressive strength of concrete from its compositions and age using the extreme gradient boosting method. *Construction and Building Materials, 260*, 119757. https://doi.org/10.1016/j.conbuildmat.2020.119757

68. Liang, M., Chang, Z., Wan, Z., Gan, Y., Schlangen, E., & Šavija, B. (2022). Interpretable Ensemble-Machine-Learning models for predicting creep behavior of concrete. *Cement and Concrete Composites, 125*, 104295. https://doi.org/10.1016/j.cemconcomp.2021.104295

69. Nguyen, H. D., Dao, N. D., & Shin, M. (2021). Prediction of seismic drift responses of planar steel moment frames using artificial neural network and extreme gradient boosting. *Engineering Structures, 242,* 112518. https://doi.org/10.1016/j.engstruct.2021.112518

70. Lyngdoh, G. A., Zaki, M., Krishnan, N. A., & Das, S. (2022). Prediction of concrete strengths enabled by missing data imputation and interpretable machine learning. *Cement and Concrete Composites, 128,* 104414. https://doi.org/10.1016/j.cemconcomp.2022.104414

71. Wakjira, T. G., Al-Hamrani, A., Ebead, U., & Alnahhal, W. (2022). Shear capacity prediction of FRP-RC beams using single and ensenble ExPlainable Machine learning models. *Composite Structures, 287,* 115381. https://doi.org/10.1016/j.compstruct.2022.115381

72. Mangalathu, S., Jang, H., Hwang, S. H., & Jeon, J. S. (2020). Data-driven machine-learning-based seismic failure mode identification of reinforced concrete shear walls. *Engineering Structures, 208,* 110331. https://doi.org/10.1016/j.engstruct.2020.110331

73. Zhang, H., Cheng, X., Li, Y., & Du, X. (2022). Prediction of failure modes, strength, and deformation capacity of RC shear walls through machine learning. *Journal of Building Engineering, 50*, 104145. https://doi.org/10.1016/j.jobe.2022.104145

74. Degtyarev, V. V., & Tsavdaridis, K. D. (2022). Buckling and ultimate load prediction models for perforated steel beams using machine learning algorithms. *Journal of Building Engineering, 51*, 104316.

75. Wen, X., Xie, Y., Wu, L., & Jiang, L. (2021). Quantifying and comparing the effects of key risk factors on various types of roadway segment crashes with LightGBM and SHAP. *Accident Analysis & Prevention, 159*, 106261. https://doi.org/10.1016/j.aap.2021.106261

76. Mangalathu, S., Hwang, S. H., & Jeon, J. S. (2020). Failure mode and effects analysis of RC members based on machine-learning-based SHapley Additive exPlanations (SHAP) approach. *Engineering Structures, 219*, 110927. https://doi.org/10.1016/j.engstruct.2020.110927

77. Mahjoubi, S., Barhemat, R., Guo, P., Meng, W., & Bao, Y. (2021). Prediction and multi-objective optimization of mechanical, economical, and environmental properties for strain-hardening cementitious composites (SHCC) based on automated machine learning and metaheuristic algorithms. *Journal of Cleaner Production, 329*, 129665. https://doi.org/10.1016/j.jclepro.2021.129665

78. Bui, D. K., Nguyen, T., Chou, J. S., Nguyen-Xuan, H., & Ngo, T. D. (2018). A modified firefly algorithm-artificial neural network expert system for predicting compressive and tensile strength of high-performance concrete. *Construction and Building Materials, 180*, 320–333. https://doi.org/10.1016/j.conbuildmat.2018.05.201

79. Gupta, R., Kewalramani, M. A., & Goel, A. (2006). Prediction of Concrete Strength Using Neural-Expert System. *Journal of Materials in Civil*

Engineering,18, 462–466. https://doi.org/10.1061/(asce)0899-1561(2006)18:3(462)

80. Zarandi, M. F., Türksen, I. B., Sobhani, J., & Ramezanianpour, A. A. (2008). Fuzzy polynomial neural networks for approximation of the compressive strength of concrete. *Applied Soft Computing, 8*(1), 488–498. https://doi.org/10.1016/j.asoc.2007.02.010

81. Naderpour, H., Rafiean, A. H., & Fakharian, P. (2018). Compressive strength prediction of environmentally friendly concrete using artificial neural networks. *Journal of building engineering, 16*, 213–219. https://doi.org/10.1016/j.jobe.2018.01.007

82. Li, Q. F., & Song, Z. M. (2022). High-performance concrete strength prediction based on ensemble learning. *Construction and Building Materials, 324*, 126694. https://doi.org/10.1016/j.conbuildmat.2022.126694

83. Yang, E. H., Yang, Y., & Li, V. C. (2007). Use of high volumes of fly ash to improve ECC mechanical properties and material greenness. *ACI materials journal, 104*(6), 620. https://doi.org/10.14359/18966

84. Şahmaran, M., Lachemi, M., Hossain, K. M., & Li, V. C. (2009). Internal curing of engineered cementitious composites for prevention of early age autogenous shrinkage cracking. *Cement and concrete research, 39*(10), 893–901. https://doi.org/10.1016/j.cemconres.2009.07.006

85. Şahmaran, M., & Li, V. C. (2009). Durability properties of micro-cracked ECC containing high volumes fly ash. *Cement and Concrete Research, 39*(11), 1033–1043. https://doi.org/10.1016/j.cemconres.2009.07.009

86. Tosun-Felekoğlu, K., Felekoğlu, B., Ranade, R., Lee, B. Y., & Li, V. C. (2014). The role of flaw size and fiber distribution on tensile ductility of PVA-ECC. *Composites Part B: Engineering, 56*, 536–545. https://doi.org/10.1016/j.compositesb.2013.08.089

87. Liu, H., Zhang, Q., Li, V., Su, H., & Gu, C. (2017). Durability study on engineered cementitious composites (ECC) under sulfate and chloride environment. *Construction and Building Materials, 133*, 171–181. https://doi.org/10.1016/j.conbuildmat.2016.12.074

88. Chung, K. L., Ghannam, M., & Zhang, C. (2018). Effect of Specimen Shapes on Compressive Strength of Engineered Cementitious Composites (ECCs) with Different Values of Water-to-Binder Ratio and PVA Fiber. *Arabian Journal for Science and Engineering, 43*, 1825–1837. https://doi.org/10.1007/s13369-017-2776-8

89. Şahmaran, M., Özbay, E., Yücel, H. E., Lachemi, M., & Li, V. C. (2011). Effect of Fly Ash and PVA Fiber on Microstructural Damage and Residual Properties of Engineered Cementitious Composites Exposed to High Temperatures. *Journal of Materials in Civil Engineering, 23*, 1735–1745. https://doi.org/10.1061/(asce)mt.1943-5533.0000335

90. Erdem, T. K. (2014). Specimen size effect on the residual properties of engineered cementitious composites subjected to high temperatures. *Cement and Concrete Composites, 45*, 1–8. https://doi.org/10.1016/j.cemconcomp.2013.09.019

91. Sahmaran, M., Lachemi, M., Hossain, K. M., Ranade, R., & Li, V. C. (2009). Influence of aggregate type and size on ductility and mechanical properties of engineered cementitious composites. *ACI Materials Journal, 106*(3), 308.

92. Neville, A. M. (2006). *Concrete: Neville's insights and issues.* Thomas Telford. https://doi.org/10.1680/cniai.34686

93. Rokugo, K. (Ed.). (2008). *Recommendations for design and construction of high performance fiber reinforced cement composites with multiple fine cracks (HPFRCC).* Japan Society of Civil Engineers, Concrete Committee. http://www.jsce.or.jp/committee/concrete/e/index.html.

94. Gholami, R., & Fakhari, N. (2017). Support vector machine: principles, parameters, and applications. In *Handbook of neural computation* (pp. 515–535). Academic Press. https://doi.org/10.1016/B978-0-12-811318-9.00027-2.

95. Chou, J. S., Tsai, C. F., Pham, A. D., & Lu, Y. H. (2014). Machine learning in concrete strength simulations: Multi-nation data analytics. *Construction and Building Materials, 73*, 771–780. https://doi.org/10.1016/j.conbuildmat.2014.09.054

96. Ren, Q., Li, M., Zhang, M., Shen, Y., & Si, W. (2019). Prediction of ultimate axial capacity of square concrete-filled steel tubular short columns using a hybrid intelligent algorithm. *Applied Sciences, 9*(14), 2802. https://doi.org/10.3390/app9142802

97. Chen, T., & Guestrin, C. (2016, August). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (pp. 785–794). https://doi.org/10.1145/2939672.2939785.

98. Bakouregui, A. S., Mohamed, H. M., Yahia, A., & Benmokrane, B. (2021). Explainable extreme gradient boosting tree-based prediction of load-carrying capacity of FRP-RC columns. *Engineering Structures, 245*, 112836. https://doi.org/10.1016/j.engstruct.2021.112836

99. Rzychoń, M., Żogała, A., & Rog, L. (2022). SHAP-based interpretation of an XGBoost model in the prediction of grindability of coals and their blends. *International Journal of Coal Preparation and Utilization, 42*(11), 3348–3368. https://doi.org/10.1080/19392699.2021.1959324

100. Lundberg, S. M., Erion, G. G., & Lee, S. I. (2018). Consistent individualized feature attribution for tree ensembles. *arXiv preprint arXiv:1802.03888.* http://arxiv.org/abs/1802.03888.

101. Feng, D.-C., Wang, W.-J., Mangalathu, S., & Taciroglu, E. (2021). Interpretable XGBoost-SHAP Machine-Learning Model for Shear Strength Prediction of Squat RC Walls. *Journal of the Structural Engineering. American Society of Civil Engineers, 147*, 04021173. https://doi.org/10.1061/(asce)st.1943-541x.0003115

102. Naser, M. Z., & Alavi, A. H. (2023). Error metrics and performance fitness indicators for artificial intelligence and machine learning in engineering and sciences. *Architecture, Structures and Construction, 3*(4), 499–517. https://doi.org/10.1007/s44150-021-00015-8

103. Uddin, M. N., Mahamoudou, F., Deng, B. Y., Musa, M. M. E., & Sob, L. W. T. (2023). Prediction of rheological parameters of 3D printed polypropylene fiber-reinforced concrete (3DP-PPRC) by machine learning. *Materials Today: Proceedings.* https://doi.org/10.1016/j.matpr.2023.03.191. https://www.sciencedirect.com/science/article/pii/S2214785323012361

104. Jackson, E. K., Roberts, W., Nelsen, B., Williams, G. P., Nelson, E. J., & Ames, D. P. (2019). Introductory overview: Error metrics for hydrologic modelling – A review of common practices and an open source library to facilitate use and adoption. *Environmental Modelling and Software, 119*, 32–48. https://doi.org/10.1016/j.envsoft.2019.05.001

105. Uddin, M. N., Ye, J., Haque, M. A., Yu, K., & Li, L. (2024). A novel compressive strength estimation approach for 3D printed fiber-reinforced concrete: integrating machine learning and gene expression programming. *Multiscale and Multidisciplinary Modeling, Experiments and Design,* 1–22.https://doi.org/10.1007/s41939-024-00439-x

106. Haque, M. A., Chen, B., Kashem, A., Qureshi, T., & Ahmed, A. A. M. (2023). Hybrid intelligence models for compressive strength prediction of MPC composites and parametric analysis with SHAP algorithm. *Materials Today Communications, 35,* 105547. https://doi.org/10.1016/j.mtcomm.2023.105547

107. Zheng, X., Xie, Y., Yang, X., Amin, M. N., Nazar, S., Khan, S. A., Althoey, F., & Deifalla, A. F. (2023). A data-driven approach to predict the compressive strength of alkali-activated materials and correlation of influencing parameters using SHapley Additive exPlanations (SHAP) analysis. *Journal of Materials Research and Technology, 25*, 4074–4093. https://doi.org/10.1016/j.jmrt.2023.06.207

108. Naser, M. Z. (2021). An engineer's guide to eXplainable Artificial Intelligence and Interpretable Machine Learning: Navigating causality, forced goodness, and the false perception of inference. *Automation in Construction, 129*, 103821. https://doi.org/10.1016/j.autcon.2021.103821

109. Huang, B. T., Yu, J., Wu, J. Q., Dai, J. G., & Leung, C. K. (2020). Seawater sea-sand Engineered Cementitious Composites (SS-ECC) for marine and coastal applications. *Composites Communications, 20*, 100353. https://doi.org/10.1016/j.coco.2020.04.019

110. Li, Y., Guan, X., Zhang, C., & Liu, T. (2020). Development of High-Strength and High-Ductility ECC with Saturated Multiple Cracking Based on the Flaw Effect of Coarse River Sand. *Journal of Materials in Civil Engineering, 32*, 04020317. https://doi.org/10.1061/(asce)mt.1943-5533.0003405

111. Sahmaran, M., Lachemi, M., Hossain, K. M., Ranade, R., & Li, V. C. (2009). Influence of aggregate type and size on ductility and mechanical properties of engineered cementitious composites. *ACI Materials Journal, 106*(3), 308. https://doi.org/10.14359/56556

112. Li, Y., Li, J., Yang, E. H., & Guan, X. (2022). Mechanism study of crack propagation in river sand Engineered Cementitious Composites (ECC). *Cement and Concrete Composites, 128*, 104434. https://doi.org/10.1016/j.cemconcomp.2022.104434

113. Kumar, R., Hashmi, S. A. R., Nimanpure, S., & Naik, A. (2017). Enhanced dynamic mechanical properties of kenaf epoxy composites. *Adv. Mater. Proceeding,* 749–757. https://doi.org/10.5185/amp.2017/981

Uddin *et al. Low-carbon Materials and Green Construction*   *(2024) 2:11*

Page 23 of 23

114. Kumar, S., Prasad, L., Patel, V. K., Kumar, V., Kumar, A., Yadav, A., & Winczek, J. (2021). Physical and mechanical properties of natural leaf fiber-reinforced epoxy polyester composites. *Polymers (Basel), 13*, 1369. https://doi.org/10.3390/polym13091369

115. Udaya Kiran, C., Ramachandra Reddy, G., Dabade, B. M., & Rajesham, S. (2007). Tensile properties of sun hemp, banana and sisal fiber reinforced polyester composites. *Journal of Reinforced Plastics and Composites, 26*(10), 1043–1050. https://doi.org/10.1177/0731684407079423

Publisher's Note