



OPEN

## Forecasting compressive strength of concrete containing rice husk ash using various machine learning algorithms

Ala'a R. Al-Shamasneh<sup>1</sup>, Manish Kewalramani<sup>2</sup>, Arsalan Mahmoodzadeh<sup>3</sup>✉, Abdulaziz Alghamdi<sup>4</sup>, Jasim Alnahas<sup>5</sup>, Nejib Ghazouani<sup>6</sup> & Mohammed Sulaiman<sup>7</sup>

Incorporating rice husk ash (RHA) into concrete improves the structure's compressive strength (CS) and durability and aids sustainability by lowering carbon emissions. This paper evaluates the application of twelve machine learning (ML) algorithms to predict the CS of concrete containing RHA. The dataset used to train, test, and validate the model comprised 500 experimental samples and 30 data points sourced externally. Through stepwise regression, seven input features were chosen: water-to-binder ratio (W/B), cement (C), superplasticizer (SP), water (W), RHA, coarse aggregate (CA), and fine aggregate (FA). Among the evaluated models, support vector regression (SVR), Gaussian process regression (GPR), and null-space SVR (NuSVR) models emerged as the best performing, each attaining  $R^2$  values over 0.93. DTR performed weakest, with  $R^2$  values below 0.53, illustrating the importance of algoRhythm selection. The study's results reaffirm the importance of RHA content and the W/B ratio as the two major determinants of the CS increase. To assist practitioners applying the trained models, a simple graphical user interface (GUI) was created to allow engineers to quickly evaluate CS and refine concrete mix designs. The combination of sophisticated ML methods with the data on RHA concrete will therefore support the overarching strategy to achieve sustainability in construction and high operational reliability of structures.

**Keywords** Reinforced concrete, Compressive strength, Rice husk ash, Machine learning, Software development

### Abbreviations

ML	Machine learning
SHAP	SHapley additive exPlanations
GPR	Gaussian process regression
SVR	Support vector regression
RFR	Random forest regression
XGBR	Extreme gradient boosting regression
ANN	Artificial neural network
KNN	K-nearest neighbors
SFRC	Steel fiber-reinforced concrete
MSE	Mean squared error
MAE	Mean absolute error
$R^2$	Coefficient of determination
RMSE	Root mean squared error
PDP	Partial dependence plot

<sup>1</sup>Department of Computer Science, College of Computer & Information Sciences, Prince Sultan University, Raffha Street, Riyadh 11586, Saudi Arabia. <sup>2</sup>Department of Civil Engineering, College of Engineering, Abu Dhabi University, Abu Dhabi, UAE. <sup>3</sup>Rock Mechanics Division, School of Engineering, Tarbiat Modares University, Tehran, Iran. <sup>4</sup>Department of Civil Engineering, Faculty of Engineering, University of Tabuk, Tabuk 47512, Saudi Arabia. <sup>5</sup>Department of Industrial Engineering, Faculty of Engineering, University of Tabuk, Tabuk 47512, Saudi Arabia. <sup>6</sup>Mining Research Center, Northern Border University, Arar 73213, Saudi Arabia. <sup>7</sup>Department of Civil Engineering, Faculty of Engineering, Al-Baha University, Alaqiq 65779, Saudi Arabia. ✉email: m.arsalan@modares.ac.ir

PFI	Permutation feature importance
RHA	Rice husk ash
CS	Compressive strength
W/C	Water-to-cement ratio
GUI	Graphical user interface
CA	Coarse aggregates
FA	Fine aggregates
C	Cement
W	Water
SP	Superplasticizer
W/B	Water-to-binder ratio
AIC	Akaike information criterion
BIC	Bayesian information criterion
UHPC	Ultra high performance concrete
Bi-LSTM	Bidirectional long short-term memory
SHAP	SHapley ADDITIVE exPlanations

Concrete is the predominant building material in structural engineering, with Portland cement being the most prevalent binder in its production, and it is employed globally for constructing various structures. Despite its widespread use, the production of Portland cement is associated with substantial carbon dioxide emissions, prompting the exploration of alternative binders to mitigate the environmental impact<sup>1</sup>. A multitude of agricultural and industrial byproducts, including sugarcane bagasse ash, rice husk ash (RHA), and silica fume, are being incorporated as supplementary cementitious materials in concrete manufacturing, offering a viable means to curtail reliance on ordinary Portland cement<sup>2</sup>.

RHA, abundant in amorphous silica, becomes a valuable addition to the concrete production process. Its inclusion not only enhances the pore structure of concrete but also contributes significantly to improving the overall strength and durability of the material<sup>3</sup>. Moreover, using RHA as a concrete component serves a dual purpose by addressing environmental concerns. RHA can lead to severe environmental pollution when left untreated and directly discharged. Integrating RHA into the concrete mix benefits not only the concrete's performance but also reduces the environmental impact. Consequently, extensive research is underway to explore the applications and advantages of RHA in concrete formulations<sup>4</sup>.

The compressive strength (CS) of concrete stands out as a pivotal parameter, bearing profound implications for the durability of structures<sup>5–8</sup>. Previous research indicates that concrete incorporating RHA tends to exhibit elevated strengths in both early and later stages<sup>9–11</sup>. Ganesan et al.<sup>12</sup> delved into the nuanced patterns of CS variation in concrete with varying RHA content. Their findings underscored that concrete strength surpassed the control group up to 30% RHA, reaching its zenith at 15%. However, a decline in strength commenced beyond the 30% threshold. Similarly, Kishore et al.<sup>13</sup> noted the peak CS of concrete occurring at a 10% RHA substitution. Bhanumathidas and Mehta<sup>14</sup> concurred, reporting that concrete strength surpassed that of the control until a 40% RHA content.

Examining the impact of water-to-cement ratio (W/C) and RHA on concrete CS, Bui et al.<sup>15</sup> observed an inverse relationship between W/C and RHA content, with the addition of RHA contributing positively to concrete CS. This aligns with the conclusions drawn by de Sensale<sup>16</sup> in the case of Uruguayan and American RHA. Additionally, established studies highlight the influence of various factors on the CS of RHA concrete, including cement content, age, W/C, water content, coarse aggregate (CA) content, fine aggregate (FA) content, and superplasticizer (SP) content<sup>9,15–17</sup>.

While traditional laboratory tests remain the conventional means of determining concrete CS, they are beset by costliness, labor intensity, and time consumption issues<sup>4</sup>. Therefore, the imperative lies in adopting a pragmatic approach for predicting the CS of RHA concrete, enabling swift assessments of concrete quality.

Several prediction equations have been devised to forecast the CS of RHA concrete. Saridemir et al.<sup>18</sup> introduced an explicit formulation rooted in gene expression programming to anticipate the CS of RHA concrete. The resulting correlation coefficient ( $R^2$ ) stood at 0.9535, attesting to its notably high prediction accuracy. Employing statistical regression analysis, Islam et al.<sup>19</sup> crafted a predictive model for the CS of RHA high-performance concrete, achieving a commendable fit with an  $R^2$  of 0.8160. Liu et al.<sup>20</sup> delved into the study of hydration products in cement slurry using X-ray analysis. They developed an optimal model for RHA replacement rate and formulated a CS prediction model for the concrete model. The mechanical properties of concrete reached their zenith with a 20% RHA content. The prediction model exhibited outstanding efficacy, with a maximum error of merely 14.4%.

Constructing a multi-factor equation for predicting the CS of RHA concrete encounters challenges arising from the nonlinear relationship between the CS of RHA concrete and various factors. This complexity can be effectively addressed through machine learning (ML). Today, ML methods have shown potential in solving problems related to structural engineering<sup>21–23</sup>. Utilizing ML approaches within concrete technology offers an intelligent perspective towards sustainability in the construction industry<sup>24–28</sup>. Recently, there has been a growing emphasis on leveraging advanced ML methods for predicting concrete CS<sup>28–32</sup>. Topcu et al.<sup>33</sup> devised an artificial neural network (ANN) and fuzzy logic model to predict CS, highlighting its significant potential for predicting the CS of fly ash concrete. In their study on high-volume fly ash self-compacting concrete, Kumar et al.<sup>34</sup> employed advanced hybrid gradient boosting models to predict the CS and developed an open-source Graphical User Interface (GUI) to support mix design optimization and enhance model transparency. Kumar et al.<sup>35</sup> developed and evaluated advanced ML models to predict the CS of ultra high performance concrete (UHPC) based on 15 input variables. Among these, the bidirectional long short-term memory (Bi-LSTM) model

achieved the highest accuracy. Sathvik et al.<sup>36</sup> replaced conventional cement and river sand with recycled fly ash and manufactured sand in concrete, testing CS over 3–90 days. They also employed ML models to accurately predict concrete CS. Erdal et al.<sup>37</sup>, in their work on predicting the CS of high-performance concrete using wavelet ensemble models, found that the discrete wavelet transform substantially enhances the prediction accuracy of ANN. Behnood et al.<sup>38</sup> employed the M5P model tree algorithm to predict CS across different concrete types based on 1912 datasets, demonstrating that the M5P model tree can serve as a viable method for CS prediction in concrete.

Recent research has started to address the prediction of CS in RHA concrete using ML methods. For instance, Li et al.<sup>39</sup> developed a hybrid neural-network model grounded on a dataset of 192 records and six input parameters, resulting in satisfactory forecasting accuracy. In a parallel effort, Iqtidar et al.<sup>40</sup> reproduced the analysis using the same dataset and an ANN approach. Hamidian et al.<sup>41</sup> have gone a step farther by integrating an ANN framework with advanced optimization methods, yielding a model where the correlation coefficient exceeds 0.95. Amin et al.<sup>42</sup> expanded the toolbox by applying bagging regressors, decision trees, and AdaBoost regressors, all demonstrating commendable precision in estimating the CS of RHA concrete. Alyami et al.<sup>43</sup> demonstrated the effectiveness of ensemble ML models (exceptionally light gradient boosting machine) in predicting the CS of RHA concrete. They used 348 values of CS collected from the experimental studies, including five characteristics of RHA concrete. The study concluded that the light gradient boosting machine is the most effective ML model for accurately predicting the CS of RHA concrete. SAPley Additive ExPlanations (SHAP) analysis further revealed that the W/C ratio is the most influential parameter in the prediction process.

Despite the progress, a knowledge gap remains in using ML to predict CS of RHA concrete when factoring in a range of multiple input variable combinations. Past investigations relied on neither exhaustive nor sufficiently large databases, resulting in a limited number of data entries for training and validation. The restricted selection of ML algorithms used in past research inhibits a thorough assessment of which models perform best under varied data scenarios. This underscores an urgent requirement to broaden the range of techniques examined, integrating ensemble methods, deep learning, support vector adaptations, and hybrid designs to more effectively capture the diverse microstructural responses of the material. Additionally, deploying sophisticated statistical protocols such as nested cross-validation, mutual information screening, permutation-based feature importance, SHAP explanatory models, and thorough uncertainty quantification remains necessary for an exacting and systematic evaluation of predictive accuracy. Together, these coordinated approaches promise to lead to predictive models of RHA concrete that are not only more precise but also more interpretable and robust across differing field applications.

This article aims to explore the efficacy of twelve ML methods in accurately estimating the CS of RHA concrete with a professional and detailed approach. The ML techniques under scrutiny encompass a diverse array of algorithms, including ANN, support vector regression (SVR), Gaussian process regression (GPR), extra tree regressor (ETR), decision tree regressor (DTR), gradient boosting regressor (GBR), histogram-based gradient boosting regressor (HGBR), extreme gradient boosting (XGBoost), null-space SVR (NuSVR), voting regressor (VR), random forest (RF), and multilayer perceptron regression (MLPR). These techniques are chosen for their capability to handle intricate relationships and patterns in data, making them widely used in predictive modeling. To guarantee reliability and the ability to generalize findings beyond the training set, the models are built and evaluated on a unique, high-fidelity dataset consisting of 500 data points created through rigorously controlled laboratory tests. This dataset covers a broad spectrum of compositions containing RHA concrete and systematically varies key material properties and mixing ratios. Unseen benchmark datasets, sourced from previous studies, supplement this core data to evaluate how well the models translate across different experimental conditions and settings. Stratified k-fold cross-validation is implemented to systematically partition the data, thereby balancing representation across the various groups and preventing both overfitting and underfitting. The resulting performance metrics are therefore more representative of the models' true capabilities. For deeper understanding of the models' inner workings, SHAP is calculated to attribute prediction variance to individual input features. Uncertainty quantification is also performed, yielding prediction intervals that inform engineers about the likelihood of varying material behavior. The Stepwise regression technique isolates the features that exert the greatest influence on CS. Complementary to this, Pearson correlation, mutual information, and distance correlation analyses together reveal both linear and nonlinear interactions among input variables and the CS target.

We created a dedicated GUI built on the trained ML models to translate our findings into practical use. The interface offers civil engineers and material scientists a straightforward tool for predicting the CS of RHA concrete by simply entering the relevant parameters. In this way, we connect sophisticated ML techniques with the routine demands of engineering practice.

## Research significance

This research study represents an original contribution to concrete technology and ML by systematically investigating the ML-based estimation of CS of RHA concrete. One of the contributions is the analysis and comparison of the twelve distinct ML algorithms, which provides an extensive including global and local perspectives into algorithmic performance. This helps in knowing the most suitable and consistent algorithms for CS estimation, paying attention to the evaluation of the algorithms based on several metrics. This study can serve as a reference in the ML algorithms selection framework for predicting concrete properties, which due to the complexity of the features, provides high prediction error.

This study advances the understanding of ML applications by showcasing the focus of this research on RHA concrete as a model demonstrates the ease with which ML techniques can handle highly complex, nonlinear, and interdependent datasets that are characteristic of concrete materials. This further demonstrates the effective

potential of ML. It can change the construction industry by making eco-friendly construction materials and techniques more widely usable and available.

The study enhances the contribution on mix design optimization by applying Stepwise regression technique to determine critical input parameters impacting concrete CS the most in a concrete structure. This enables engineers and materials scientists to tailor mix design compositions on the offered strength parameters.

From a model building perspective, this research collected a unique dataset of 500 experimental data points and put them through rigorous model training and testing. To achieve an enhanced measure of model robustness, another 30 independent data points were collected from freely available literature sources and were used for external validation. This combination of datasets not only makes concrete the model evaluation, but also strength their confidence on the results obtained from the model for various concrete mixtures.

The most critical contribution of the research is providing actionable knowledge by identifying the optimal content of RHA that yields the highest CS, thus enhancing the mix design optimization. The work strengthens the understanding of RHA and justifies its strategic use in high-performance concrete.

In order to confirm the reliability and generalizability of the created ML models, a thorough validation strategy incorporating k-fold cross-validation was employed to reduce bias and variance for the performance metrics of the models across different data splits. In addition, the uncertainty quantification, is the focus of the study because it offers predictive intervals, which is an important insight crucial for understanding the confidence of model outputs. This is particularly important for engineering fields where decisions are often made under uncertainty. In support of explainable AI and to increase interpretable results, SHAP analysis was performed which allowed for estimating feature importance and explaining individual predictions in greater detail. Integrating the quantification, validation, and explanation techniques increases the trust and confidence in the results of the ML framework while providing the transparency that is often missing in such advanced techniques.

A practitioner's ability to estimate the CS of RHA concrete and create optimal mix designs has been simplified with the development of a practical and GUI. This interface combines powerful ML tools with the practical needs of the civil engineering and research communities, enabling professionals to utilize predictive approaches with low computational proficiency.

This research is particularly remarkable for the thoroughness of its methodology, practical impact, and focus on sustainability, providing significant advancements in both materials engineering and applied ML.

## Research methodology

The ML techniques scrutinized in this investigation encompass an array of algorithms, including SVR, GPR, NuSVR, ANN, XGBoost, MLP, DTR, GBR, RF, HGBR, VR, and ETR. Here's a concise overview of these techniques and their comparative advantages within the realm of ML:

GPR is a flexible tool from Bayesian statistics, offering a full distribution over functions instead of committing to single value estimates<sup>44</sup>. This feature becomes indispensable in civil engineering, where knowing the range of possible material strengths is key to robust, risk-aware design. By treating the observed measurements as noisy samples from a latent function, GPR automatically infers uncertainty, supplying credible intervals alongside predicted values. Our dataset of 500 observations sits comfortably within the range where GPR shines, delivering fluid, continuous estimates that respect the underlying nonlinearity and tolerate the moderate noise typical of material testing.

SVR stands out for predicting CS because it adeptly captures non-linear relationships through kernel tricks like the radial basis function. By focusing on minimizing structural risk rather than merely fitting the data, SVR achieves solid generalization across even large, complex feature sets<sup>45</sup>. This quality is particularly valuable for RHA concrete, where the interplay between variables such as RHA dosage and W/C ratio often introduces intricate non-linear behaviors. SVR manages these interactions with consistency, leading to dependable strength forecasts.

NuSVR incorporates the  $\nu$  (nu) parameter so it can be dynamically adjusted that how many support vectors we want and how many margin errors we're willing to tolerate<sup>46</sup>. This extra knob for tuning not only makes the model more interpretable but also lends it extra stiffness against noise and outliers. These pesky perturbations often creep into experimental data, whether from the intrinsic variability of the materials involved or from slight inconsistencies in the testing procedures.

DTR breaks the feature space into clearly defined regions, applying simple decision rules at each split. Its structure lends it a high level of interpretability, and it can naturally capture non-linear patterns without needing normalized inputs<sup>47</sup>. Although a single tree is at risk of overfitting, it is a solid cornerstone for building more robust ensemble approaches.

Drawing motivation from biological neural networks, an ANN serves as a robust function approximator, adept at modeling intricate, non-linear interactions among input features<sup>48</sup>. Because of its flexible architecture, it can effectively memorize and generalize the intricate dependencies present in concrete mixture formulation and mechanical performance. Nonetheless, the technique mandates deliberate regularization strategies and a disciplined training regimen to mitigate the risks of overfitting, which is particularly salient when working with datasets of intermediate scale.

ETR extends the RF framework by substituting optimized split thresholds with randomly chosen cut-points. This tweak ramps up the method's overall stochasticity, yielding both lower variance in the learned predictions and lighter computational loads<sup>49</sup>. Such a setup excels when the goal is to probe potential feature interactions without committing to the heavier cost of optimization, making it particularly useful in the early, exploratory sweeps of a competitive modeling effort.

GBR constructs its trees one after another, deliberately focusing each new model on the mistakes the prior models were incapable of addressing. This layer-by-layer focus on refining the residuals allows the ensemble growing

process to home in on faint, retirees patterns lurking in the idiosyncratic space of the data<sup>49,50</sup>. GBR is ideal for high-precision tasks like CS estimation in RHA concrete, especially when nonlinear and interaction effects are present.

HGBR streamlines GBR by organizing continuous inputs into histograms, which cuts training time and memory use sharply<sup>51</sup>. It keeps the predictive strength of the original method and improves how well the model scales, so it shines in situations requiring either big data processing or fast iteration during model development.

XGBoost is an optimized framework built on the gradient boosting framework that prioritizes both speed and flexibility<sup>49</sup>. By incorporating L1 and L2 penalties, it strengthens the model against overfitting. Innovations such as histogram-based quantile sketching, distributed tree building, and dynamic tree pruning minimize waste and improve execution time. This library shines with structured or tabular datasets, regularly finishing at the top of Kaggle leaderboards in regression and ranking tasks alike.

RFs consist of many decision trees built from different bootstrapped samples, with features also divided randomly during splits. This averaging process cuts the overall model variance while lowering the risk of overfitting<sup>52</sup>. As such, random forests serve as dependable starter models for predicting CS. Their robustness to noisy measurements and competence with mixed variable types make them particularly suited for the varied constituents in concrete mix designs.

MLPR is a specialized implementation of ANN focused on regression tasks. It consists of input, hidden, and output layers and uses backpropagation for training<sup>53</sup>. Its layered architecture enables it to learn hierarchical feature representations, making it ideal for modeling intricate dependencies in concrete compositions and strength outcomes.

VR leverages various complementary regressors by either averaging their predictions or by weighting each one according to its reliability on the given hold-out sample<sup>54</sup>. Here, it operates as a high-level ensemble, boosting generalization and guarding against the idiosyncratic errors of any single base learner. This gains particular traction in situations where merit is spread unevenly among different measures, and no single model consistently takes the lead.

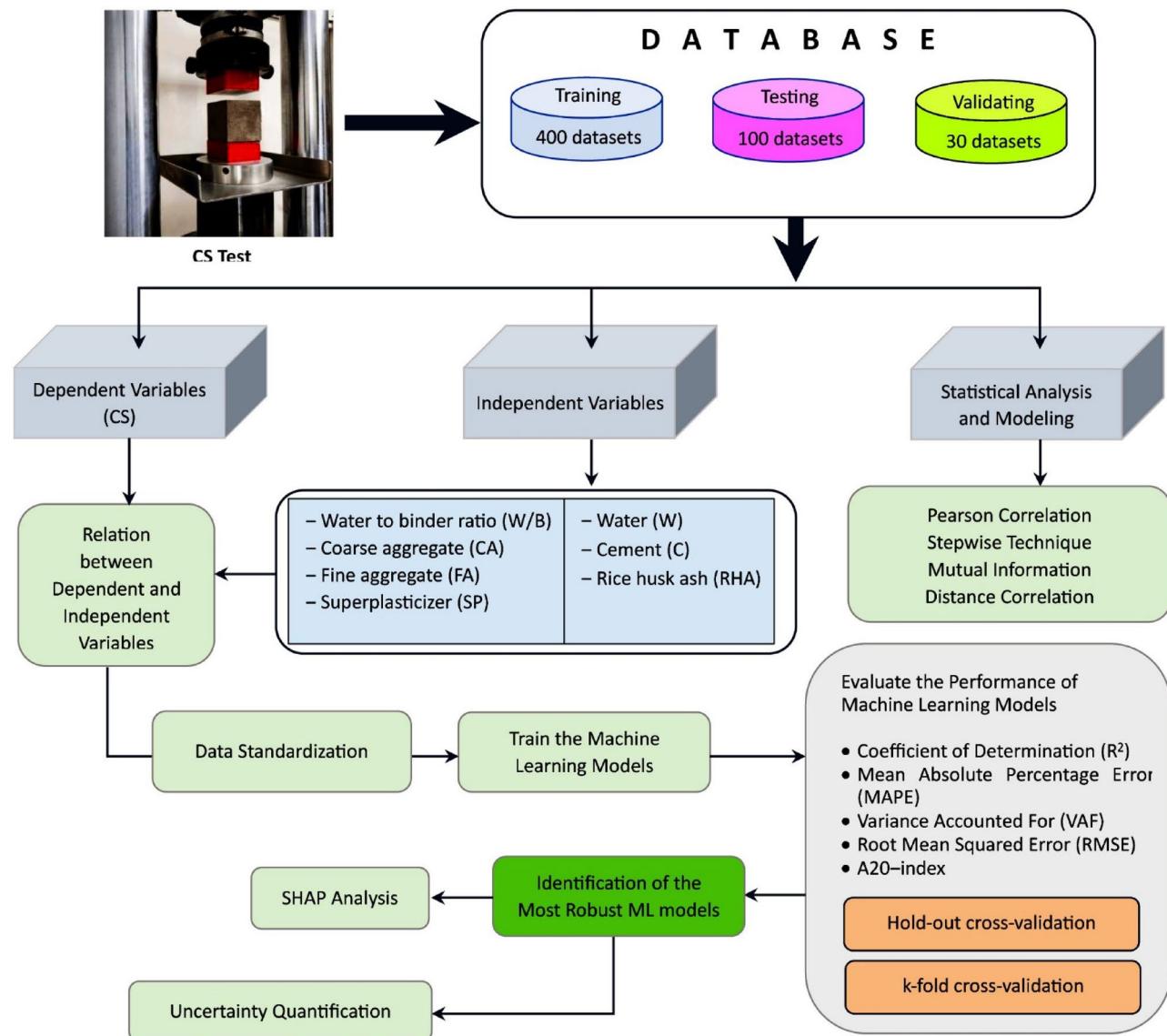
The ML model development and analysis were implemented using the Jupyter Notebook interface with Python 3.7 within the Anaconda Navigator distribution. Python 3.7 was selected due to its robust compatibility with key ML libraries (such as scikit-learn, XGBoost, SHAP, and dcor) at the time of model development. While newer Python versions (e.g., 3.10 and beyond) offer improved syntax features, they occasionally introduce dependency conflicts or deprecated functionalities with some of the specialized packages used in scientific computing and interpretability (e.g., SHAP or dcor), especially in combination. Also, the version of 3.7 still retains its popularity for broad support and stability which makes it a good environment for reproducible scientific calculations. At the same time, it was verified that there is forward compatibility with the deployment and GUI integration for version 3.9 and later. With a current configuration that consists of an Intel Core i7-10750 H CPU (2.60 GHz) and a workstation with 32GB of RAM, the performance is not an issue with multiple ML model trainings together with SHAP and cross-validation tasks. The described configuration enables reliable performance for the given tasks.

Every ML model received hyperparameter tuning utilizing a Grid Search approach on the training dataset. This approach exhausts through a given collection of hyperparameter combinations to find the optimal one for model performance, as evaluated by cross-validation on performance metrics. By definition, Grid Search provides a way to formally document optimization steps taken, ensuring no steps are skipped, and is more systematic than ad hoc optimization efforts.

The methodological framework adopted in this study, illustrated comprehensively in (Fig. 1), integrates experimental data generation, statistical analysis, ML development, and model deployment through a GUI interface. The aim is to establish a robust and interpretable pipeline for predicting the CS of RHA concrete based on seven fundamental mix design parameters. The first step in the process is the generation of rich datasets, in this case, through novel data generation, a rigorous laboratory process was used to yield an abundant dataset. The data was obtained through mechanical testing, where a total of 530 data points in the form of concrete samples straddling different RHA to aggregate ratios were split into three sets to allow robust model training (400), external validation (30), and model tuning (100) on the provided dataset. The CS of the samples, tested through standardized mechanical tests, was extracted as the dependent variable for model training. For the modeling framework, the independent variables include key ingredients of the concrete mix: water-to-binder ratio (W/B), coarse aggregate (CA), fine aggregate (FA), superplasticizer (SP), water content (W), cement content (C), and RHA content. These variables bring several dimensions of a potential input space likely to have a non-linear and dependent relationship with the CS, thus requiring intricate modeling and ML approaches.

Before model training, a thorough statistical analysis was performed to discover relationships among the input features and the target variable. Apart from the classical Pearson correlation analysis, the study also used mutual information and distance correlation to capture dependencies of both types. These approaches strengthen the exploration of the data, particularly among relationships that are non-monotonic or driven by thresholds. In addition, stepwise regression with Akaike Information Criterion (AIC), its corrected version AICc, and Bayesian Information Criterion (BIC) were applied to highlight the significant input features aimed at reducing redundancy, and risks of overfitting. The input features were standardized using the StandardScaler method. Centering a dataset to zero mean and scaling to unit variance ensures numerical stability which prevents features with large values from dominating the model during training.

Twelve distinct ML models were created and analyzed: SVR, GPR, Nu-SVR, ANN, XGBoost, GBR, HGBR, RF, MLPR, DTR, VR, and ETR. The predictive performance of these models was evaluated using well-known metrics such as coefficient of determination ( $R^2$ ), RMSE, MAPE, VAF, and a20-index. To enhance statistical rigor and mitigate biases from data pruning, two validation techniques including hold-out and k-fold were utilized.



**Fig. 1.** Flowchart of the current study.

Beyond accuracy, model interpretability was approached via SHAP, which describes and quantifies the contribution of a single feature to the prediction made. This improves the model's transparency and concrete interpretability, revealing how the mix design variables affect concrete performance. In addition, ML models were examined to estimate the bounds of the predicted values, which is necessary for engineering applications, so uncertainty quantification methods were utilized.

### Data preparation

#### RHA characteristics

The leftover RHA was procured from rice fields situated in the north of Iran. Rice husk pellets were burned in a steam boiler RHA between the temperatures of 650–750 °C. The RHA obtained was analyzed for its structural properties and was found to have predominantly amorphous silica with some crystalline silica. RHA is made up of irregularly shaped particles with a porous cellular structure. The average particle size of RHA was measured using Mastersizer 2000 and found to be 68 µm in diameter. RHA was ground for one hour in a ball mill which reduced the average particle size to 15 µm. The RHA was found to have high silica content and loss on ignition which was consistent with other studies. More details about the RHA used in this study are provided in (Table 1).

The sources of the rice husk, combustion conditions, and subsequent processing can greatly alter the peculiar physico-chemical characteristics of RHA, silica content, particle size distribution, and degree of amorphousness. This variability introduces some degree of uncertainty in the generalization of the performance of the ML models trained for different batches of RHA. Consequently, the models trained in this study would best predict the CS of concrete containing RHA of the type characterized in this study. Wider generalization for the other sources of

	Item	Unit	Cement	RHA
Physical properties	Specific gravity	–	3.11	2.10
Chemical compositions	$\text{Al}_2\text{O}_3$	(%)	5.45	0.32
	$\text{MgO}$	(%)	2.48	0.77
	$\text{SiO}_2$	(%)	21.78	87.30
	$\text{SO}_3$	(%)	1.98	1.32
	$\text{Fe}_2\text{O}_3$	(%)	3.55	0.52
	$\text{K}_2\text{O}$	(%)	0.82	3.58
	$\text{Na}_2\text{O}$	(%)	0.68	0.19
	$\text{P}_2\text{O}_5$	(%)	0.14	1.33
	Loss on ignition	(%)	0.72	9.31

**Table 1.** Physical and chemical properties of cement and RHA.

RHA would likely need retraining of the models or adaptation of the domain models. This limitation should be imposed in practical applications and future work.

### Materials used in concretes

We utilized type I Portland cement in this study. The physical and chemical attributes of the Portland cement employed are outlined in (Table 1). Our crushed coarse aggregate (CA), sourced from local quarries, boasted a maximum size of 17 mm, a density 2.62, and an absorption capacity of 1.42%. Additionally, natural sand from the same quarries featured a modulus of fineness of 3.3, a density of 2.68, and an absorption capacity of 1.3%. Local tap water was used to mix water. We incorporated a Type-G superplasticizer (SP) with a 40% solid content and a specific gravity of 1.21 to attain the desired workability for all concrete mixtures.

The properties of fresh concrete were evaluated using the slump test as per ASTM C143/C143M-15a and the unit weight, yield, and air content measurement (via the gravimetric method) according to ASTM C138/C138M-17a. All materials and processes used were in compliance with the relevant ASTM standards, which provided uniformity and reproducibility in the concrete manufacturing and testing processes.

### Testing program and database

RHA was used as a pozzolanic material in concrete. The concrete was tested for evaluating the CS. In a rotating concrete drum mixer, CA and FA, along with powder materials such as C and RHA, were meticulously proportioned. The initial dry mixing lasted for two minutes, followed by an additional three minutes after introducing W. Subsequently, the concrete blend underwent a final three-minute mixing phase upon incorporating the SP to achieve the desired consistency. Immediately slump and unit weight assessments were conducted on the freshly mixed concrete.

For the casting process, 100 mm-sized cubes were formed and compacted in dual layers atop a vibrating table, with each layer subjected to a 10-second vibration. Post-casting, molds were promptly covered with polyethylene sheets and moistened burlap, left undisturbed for 24–24-hour. Afterward, the specimens were demolded and submerged in water at 20 °C for curing until the day of testing. This involved the meticulous preparation of 500 cubic specimens, ultimately determining their 28-day CS. Ultimately, 500 data points, including seven input parameters such as W/B, C, RHA, W, SP, FA, and CA were documented, each corresponding to a specimen with distinct characteristics. Table 2 outlines the overall specifications of these data points.

Feature engineering stands as a critical stride in ML, involving the meticulous selection, transformation, and creation of features derived from raw data to optimize our model's performance<sup>55</sup>. Within this article, we harness the Stepwise method (a widely embraced technique in ML and statistical models) to automatically cherry-pick the most pertinent features from the given dataset. This method engages in an iterative dance of adding or discarding features contingent on their statistical significance or predictive prowess.

We used three statistical measures to improve the accuracy of the feature selection, including the AIC, its corrected version AICc, and BIC. These criteria assess the adequacy of a model in relation to a certain dataset while applying a penalty in terms of complexity to avoid overfitting<sup>56</sup>. This means that the model will have more accuracy in the prediction if it uses a lower number of variables. For instance, if an additional variable is added and no considerable increase in the model accuracy is observed, AIC, AICc, and most importantly BIC, will discourage its inclusion. Among them, AICc is preferred for smaller datasets, as was the case with this study, and BIC is more aggressive with increased penalties for extra variables. As seen in this study, the chosen model had the lowest AIC and AICc values and thus, showed considerable balance between predictive performance and simplicity.

To pinpoint the key factors affecting concrete CS, we used the *StepAIC()* function from the MASS package in R. This technique performs stepwise variable selection by repeatedly fitting models and comparing their AIC values. The procedure balances goodness-of-fit and model size, and was configured to perform simultaneous forward selection and backward elimination.

Table 3 shows that all seven variables retained in the final model have p-values well below the 0.05 threshold, establishing their statistical significance. In particular, the W/B ratio, W, SP, and C present extremely small p-values ( $p < 2e-16$ ) and indicate powerful links to compressive strength. The most significant p-value, corresponding to the RHA content, is 0.0083 and is still firmly below the significance cutoff, justifying its retention. The analysis

		W/B	C (kg/m <sup>3</sup> )	RHA (kg/m <sup>3</sup> )	W (kg/m <sup>3</sup> )	SP (kg/m <sup>3</sup> )	FA (kg/m <sup>3</sup> )	CA (kg/m <sup>3</sup> )	CS (MPa)
Training data	Count	400	400	400	400	400	400	400	400
	Mean	0.32	496.96	81.80	179.55	35.49	650.09	1125.08	57.41
	Std	0.06	165.14	53.08	38.53	25.56	171.81	156.19	15.97
	Min	0.21	188.00	0.00	115.00	0.00	316.00	881.00	4.97
	25%	0.27	391.00	41.00	149.60	11.33	542.75	995.75	47.48
	50%	0.33	475.00	74.00	171.00	30.85	644.95	1094.00	58.11
	75%	0.37	623.00	127.00	212.00	58.00	785.00	1272.25	67.23
	Max	0.50	826.00	187.00	255.00	85.00	980.00	1420.00	98.10
Testing data	Count	100	100	100	100	100	100	100	100
	mean	0.31	492.67	82.14	181.58	38.57	626.21	1126.87	55.91
	Std	0.06	187.83	51.55	38.74	23.94	186.13	143.12	15.27
	Min	0.21	197.00	3.00	117.00	1.00	325.00	890.00	18.29
	25%	0.25	304.75	34.00	150.00	16.75	450.50	992.75	47.60
	50%	0.30	489.00	75.00	179.00	37.00	617.00	1132.50	56.42
	75%	0.36	647.25	130.75	206.25	57.25	777.25	1216.75	66.02
	Max	0.42	819.00	185.00	255.00	85.00	947.00	1406.00	89.08

**Table 2.** General specification of the datasets.

	Estimate	Std. error	t-value	p-value	Code
Intercept	56.8168	0.8179	69.467	<2e-16	***
W/B	-24.2189	1.7443	-13.884	<2e-16	***
C	-15.1332	1.4961	-10.115	<2e-16	***
RHA	-2.1259	1.4928	-1.424	0.0083	**
W	-17.9758	1.3744	-13.079	<2e-16	***
SP	-27.7828	1.4518	-19.137	<2e-16	***
FA	14.8466	1.4573	10.188	<2e-16	***
CA	17.1440	1.1418	15.015	<2e-16	***

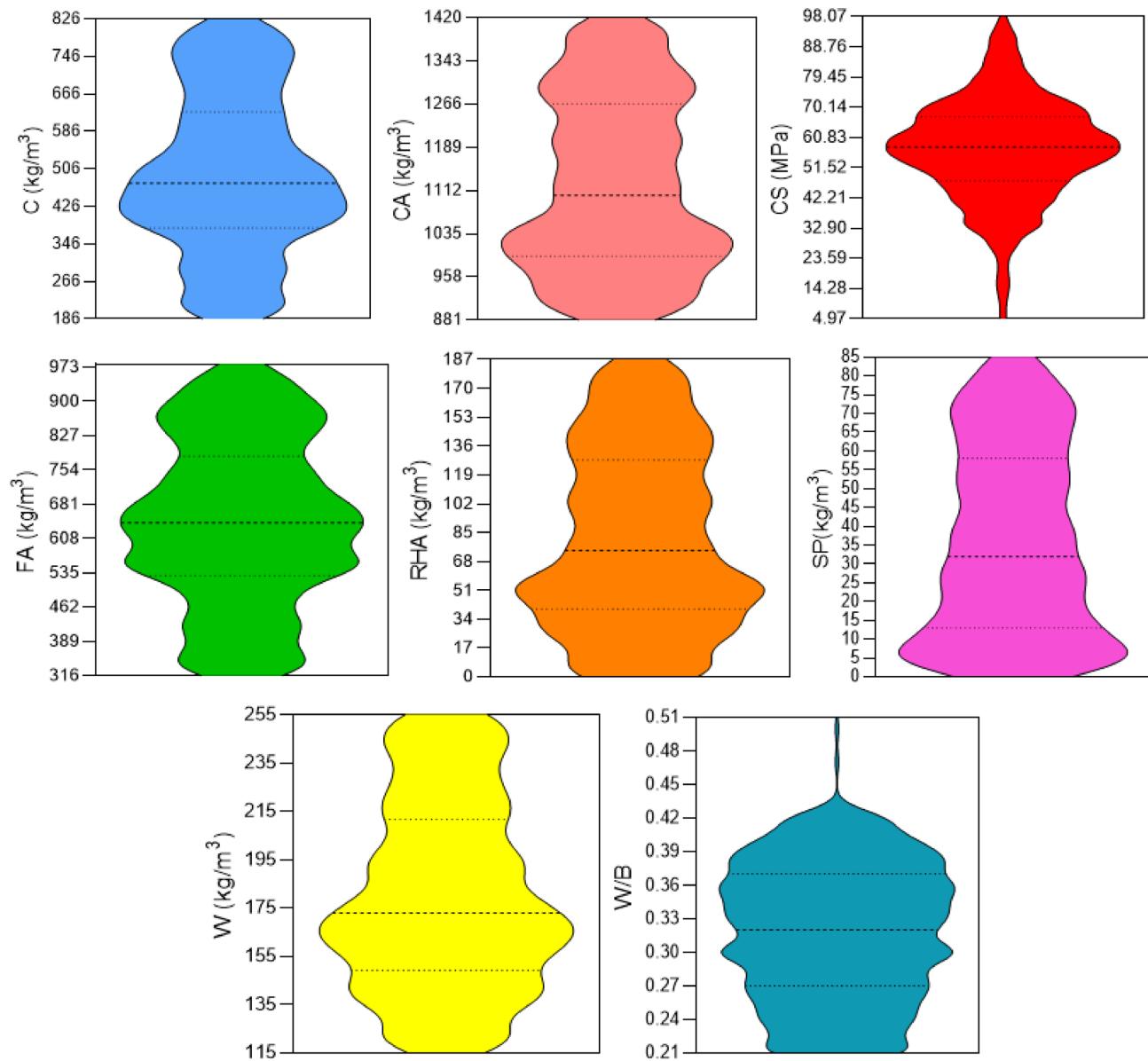
**Table 3.** Leveraging feature engineering through the Stepwise method. Significant codes: 0–0.001 (\*\*); 0.001–0.01 (\*\*); 0.01–0.05 (\*); 0.05–0.1 (...); 0.1–1 (.); >1 (.)

verifies that each included predictor is meaningful and supports the validity of the selection technique. The retained factors also resonate with established concrete science, where the interplay between binder formulation, water volume, and aggregate properties is recognized as decisive for mechanical strength.

Figure 2 shows the comprehensive portrayal of input and output parameter values through violin plots. The utility of violin plots lies in their versatility, providing an insightful means to capture the nuances of data distribution. This makes them indispensable tools in exploratory data analysis and facilitating nuanced statistical comparisons. Moreover, violin plots effectively convey intricate data patterns to a diverse audience.

Considering the details of the violin plots, it is clear that most of the input parameters are typically distributed, confirming the appropriateness of the dataset for ML. The exception is the W/B ratio, which has an unusual distribution showing a marked density dip between 0.43 and 0.51. This pronounced non-uniformity is undoubtedly more than a random statistical occurrence and has important theoretical implications. The W/B ratio is arguably the most critical factor that governs the degree of cement hydration, the associated porosity, and the concrete CS in a concrete element. A dip suggests a lack of experimental data in the neighborhood of critical transition zones where, beyond certain limits, the W/B ratio can increase or decrease the rate of strength gain. This is especially true for RHA concrete given its peculiar high surface area and pozzolanic activity which alters its water demand. This distributional irregularity may enhance model sensitivity in that area while calling for more targeted sampling thresholds. So, the unique viola plot shape of W/B reflects its clarity of statistical and mechanistic significance in CS prediction, confirming the need for meticulous treatment in model design and explanation.

Within Fig. 3, a matrix meticulously lays out the intricate connections linking the input and output parameters. This visual representation distinctly reveals a tenuous correlation between the input parameters and their interplay with the output parameter CS. The discernible implication is the absence of a discernible linear relationship among these parameters. In simpler terms, conventional linear methods are ill-suited to unravel the underlying patterns governing these interactions. This conspicuous non-linearity in the relationship demands the prowess of advanced non-linear ML algorithms. It serves as a clarion call to embrace sophisticated methodologies that can adeptly navigate and comprehend the intricate complexities inherent in this dynamic interplay. The conventional constraints of linear approaches are transcended by the exigencies of a non-linear



**Fig. 2.** Violin plots for each feature and the target. C: Cement content; CA: Coarse aggregate; CS: Compressive strength; FA: Fine aggregate; RHA: Rice husk ash; SP: Superplasticizer; W: Water content; W/B: Water-to-binder ratio.

landscape, necessitating a paradigm shift towards more nuanced and intricate analytical techniques to unravel the subtleties embedded in these connections.

Figure 4 displays the distribution of all the continuous input features employed in predicting the CS of the cementitious mixtures. Before proceeding to the visual representation, we undertook a systematic procedure to detect and remove outliers, thereby enhancing the statistical integrity of the modeling process. The Interquartile Range (IQR) technique, a standard tool in robust statistical analysis, served to isolate and discount data points that deviated too far from the interquartile range. This method defines outliers as data points lying outside the range  $[Q1 - 1.5 \times IQR, Q3 + 1.5 \times IQR]$ .  $Q1$  and  $Q3$  denote the 25 and 75th percentiles, respectively, and  $IQR = Q3 - Q1$ . Application of this protocol effectively eliminated all extreme deviations, yielding a cleaned dataset of 500 reliable samples that were then available for training and analytical procedures. The box plots shown in Fig. 4 are based on this post-IQR-cleaned cohort.

The majority of the features, including W, SP, C, and RHA, display distributions that are approximately symmetric and exhibit little skewness. Such uniformity promotes stable convergence properties in model training. Each box's vertical length illustrates the interquartile range, offering a snapshot of how spread out the central half of the data is. CA and FA, with noticeably taller boxes, indicate a greater dispersion of aggregate proportions in their respective mix designs. After applying the interquartile-range rule, every variable reveals no extreme data points. This silence suggests that the preprocessing steps succeeded in filtering out outlying

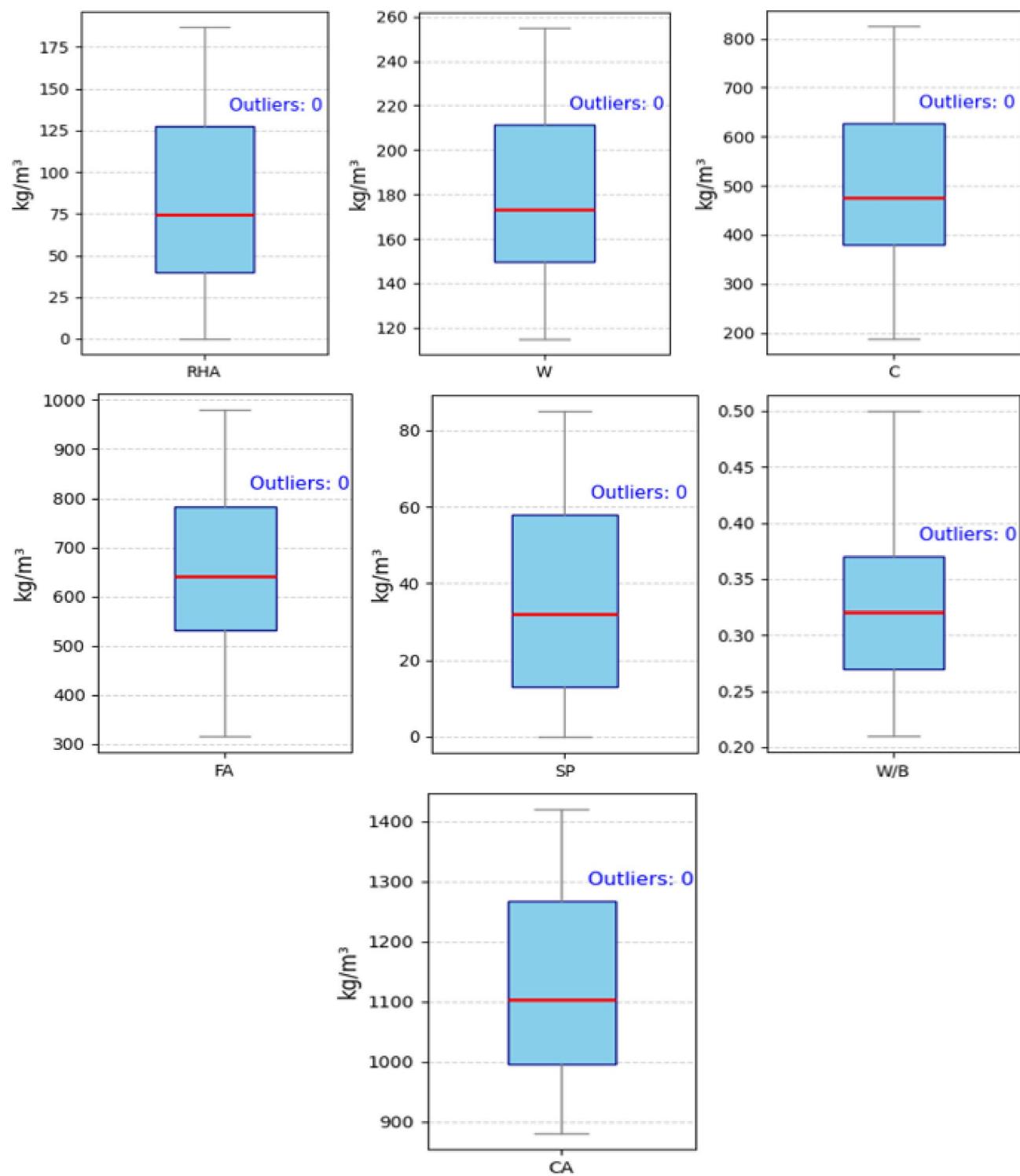


**Fig. 3.** Correlation matrix between the predictors and the target. C: Cement content; CA: Coarse aggregate; CS: Compressive strength; FA: Fine aggregate; RHA: Rice husk ash; SP: Superplasticizer; W: Water content; W/B: Water-to-binder ratio.

noise. The cleaned data is now more stable for the subsequent ML tasks, mitigating the risks of fitting to aberrant values.

We adopted mutual information and distance correlation methods to deepen our insight into feature-to-target links beyond the linear scope captured by Pearson correlation. These nonlinear and model-agnostic metrics expose the intricate, non-parametric ties between the predictive variables and the target<sup>57,58</sup>. Figure 5 presents mutual information scores for each feature in relation to CS. Because mutual information encompasses both linear and nonlinear ties, it highlights variables that shape the target through intricate, possibly threshold-like, or saturation behaviors. The SP emerged as the most informative predictor, with mutual information score = 0.188, underscoring its nonlinear capacity to enhance workability without adding moisture. W and C ranked next, with mutual information scores of 0.129 and 0.127, quantifying their marked yet partially overlapping influences. The mutual information scores for RHA and W/B, though still non-trivial, were lower, indicating that their contributions to CS might hinge on specific contexts or dosage ranges.

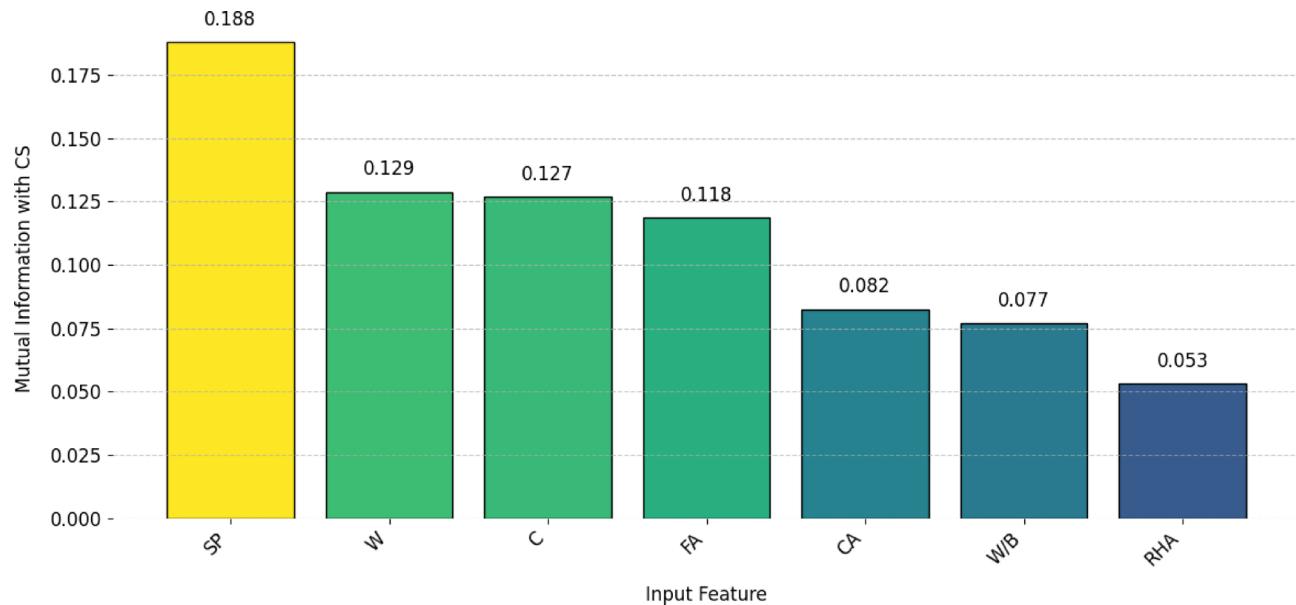
Figure 6 presents the distance correlation findings. Unlike Pearson or mutual information, distance correlation captures any form of statistical dependency, whether linear or nonlinear, producing values that span from 0 to 1. SP once more leads the rankings (distance correlation = 0.514), affirming its decisive influence on CS. W (distance correlation = 0.374) and the W/B (distance correlation = 0.328) reveal noteworthy distance-based connections to strength, perhaps reflecting subtler hydration phenomena that Pearson fails to condense.



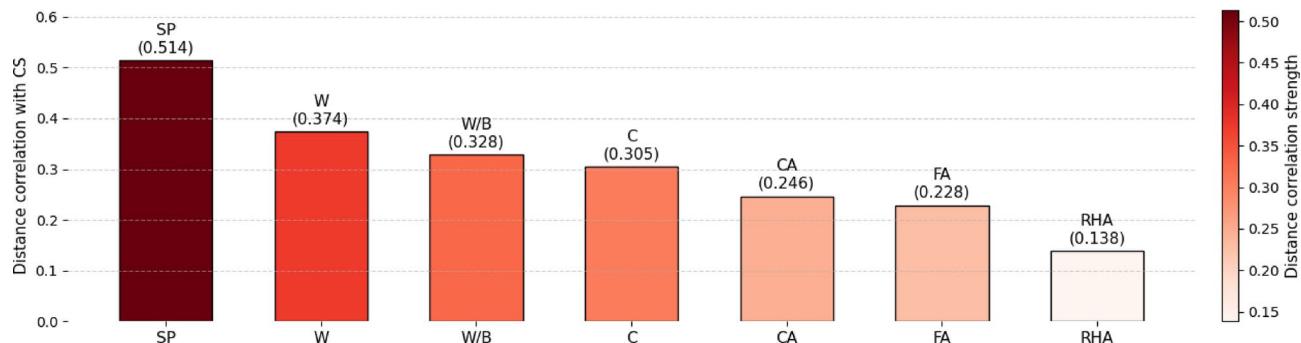
**Fig. 4.** Box plots of input features. C: Cement content; CA: Coarse aggregate; FA: Fine aggregate; RHA: Rice husk ash; SP: Superplasticizer; W: Water content; W/B: Water-to-binder ratio.

RHA, celebrated for its environmental merit and pozzolanic activity, registers the weakest distance correlation (0.138), suggesting its advantages can only be unlocked through more tailored incorporation.

These advanced correlation tools augment Pearson by disclosing nonlinear structures and cross-validating the primacy of SP, W, and C regardless of technique. The union of mutual information and distance correlation enriches variable selection and bolsters model robustness and interpretability (essential ingredients for refining ML pipelines aimed at predicting concrete properties).



**Fig. 5.** Feature-wise mutual information scores with the target variable. C: Cement content; CA: Coarse aggregate; FA: Fine aggregate; RHA: Rice hush ash; SP: Superplasticizer; W: Water content; W/B: Water-to-binder ration.



**Fig. 6.** Distance correlation scores with target variable. C: Cement content; CA: Coarse aggregate; FA: Fine aggregate; RHA: Rice hush ash; SP: Superplasticizer; W: Water content; W/B: Water-to-binder ration.

### Data standardization

In the realm of ML algorithms, standardization of the data emerges as one of the most important preprocessing steps, neutralizing all features to a common scale without bias<sup>59,60</sup>. This step is essential for algorithms that use distance metrics or rely on gradients to optimize, like SVR, ANN, and GPR. Learning is usually misled with biases due to differences in the features used. In the absence of standardization, features that are numerically larger in scale can overshadow smaller features. This leads to model performance and interpretation misaligned with the intended goals.

This study used the Standard Scaler implementation from the *scikit-learn* library. This implementation standardizes features by first *rg* the mean and scaling to unit variance, leading to feature values with a mean of zero and a standard deviation of 1. This is favorable for the dataset employed in this study, which contains W, RHA, and SP, all in different units and magnitudes. StandardScaler's assumption of Gaussian-like feature distribution is reasonable with this study's data after we remove outliers and normalize.

Alternative scaling options such as MinMaxScaler would scale features to a fixed range of [0,1] and were tried in earlier experiments. Those models proved to be unhelpful for certain models, particularly SRV and GPR. MinMaxScaler issues where features with a narrow range of variability tend to be compressed and extreme values overemphasized, with severe consequences to how well the models can generalize. In contrast, StandardScaler kept the equilibrium of the distribution and model-defined features, bringing stability, interpretability, sensitivity analysis, and SHAP values feature attribution.

## Performance evaluation of the ML models

The following evaluation criteria have been meticulously employed to assess the performance of the ML models in estimating the CS of concrete specimens. These criteria bear paramount significance, offering unique insights into the accuracy and efficacy encapsulated within the models.

This metric scrutinizes the extent to which the ML models account for the variance in target parameter values. A higher  $R^2$  value signifies a more robust alignment between the models and the observed data Eq. (1).

MAPE acts as a yardstick, quantifying the average percentage difference between predicted and measured CS values. A lower MAPE value indicates a higher accuracy level in the models' estimations Eq. (2). RMSE, the square root of the average squared differences between predicted and measured values, gauges the overall error in the models' predictions. A lower RMSE underscores superior predictive performance Eq. (3). VAF measures the proportion of variance in the predicted values attributed to the ML models. A higher VAF implies a more substantial contribution from the models in elucidating the variability in target parameter values Eq. (4). The a20-index emerges as a specific performance metric, meticulously evaluating the accuracy of the models within a predefined tolerance range. It quantifies the percentage of predicted values falling within  $\pm 20\%$  of the target parameter values. A higher a20-index underscores the models' prowess in providing precise estimates within the specified tolerance range Eq. (5).

These evaluation criteria collectively serve as a comprehensive toolkit, dissecting the nuanced facets of the ML models' performance and fortifying the reliability of their estimations for the CS of concrete specimens.

$$R^2 = \left( \frac{\sum_{i=1}^n (f(x_i) - \bar{f}(x))(f^*(x_i) - \bar{f}^*(x))}{\sqrt{\sum_{i=1}^n (f(x_i) - \bar{f}(x))^2 \sum_{i=1}^n (f^*(x_i) - \bar{f}^*(x))^2}} \right)^2 \quad (1)$$

$$MAPE = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{f(x_i) - f^*(x_i)}{f(x_i)} \right| \quad (2)$$

$$RMSE = \sqrt{\left( \frac{1}{n} \right) \sum_{i=1}^n (f(x_i) - f^*(x_i))^2} \quad (3)$$

$$VAF = 1 - \left[ \frac{var(f(x_i) - f^*(x_i))}{var(f(x_i))} \right] \times 100\% \quad (4)$$

$$a20 - inxex = \frac{n - EquationNumber of points between x = 1.10y and x = 0.90y}{n} \quad (5)$$

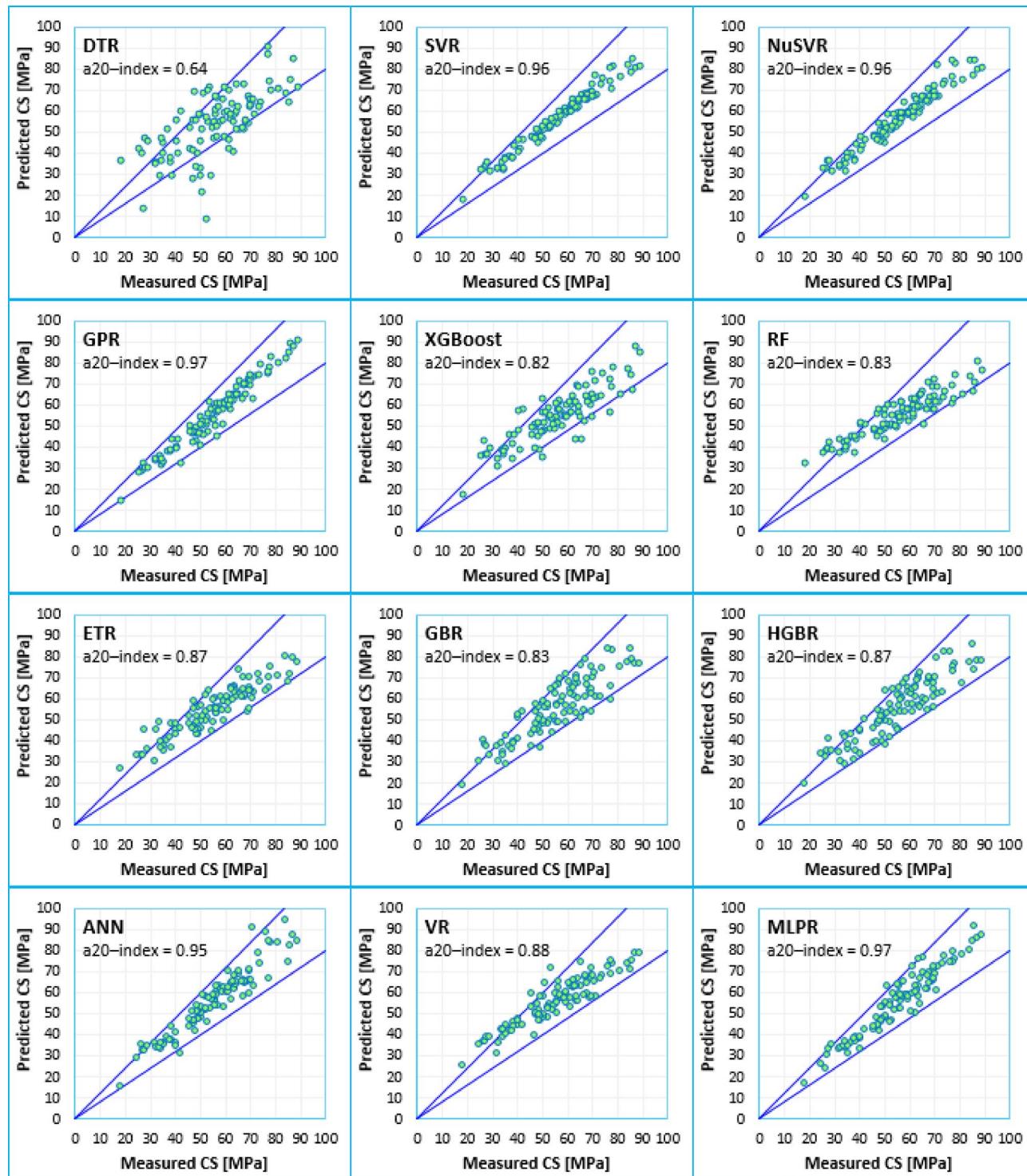
Where  $f(x_i)$  and  $f^*(x_i)$  are the measured and estimated values of parameter  $x$  for the  $i^{\text{th}}$  dataset, respectively.  $n$  is the total number of test datasets.

To gauge the efficacy of the ML algorithms, each algorithm undergoes a rigorous evaluation process, garnering scores based on predefined criteria. The culmination of these scores across all evaluation criteria is then meticulously calculated for each algorithm. The algorithm with the highest total points is consequently endorsed as the most fitting and accurate choice for estimating the CS of concrete. This methodology ensures a thorough appraisal of the ML models' performance, enabling pinpointing the most precise and suitable algorithm for the task at hand—estimating the CS of concrete.

## Results analysis and comparison

In Fig. 7, we meticulously scrutinize the estimated values produced by each algorithm against the individually measured 28-day CS values. This scrutiny unfolds through graphs employing the a20-index metric, revealing a noteworthy alignment of the majority of points within the  $x = 1.20y$  and  $x = 0.80y$  lines. This alignment signifies the commendable accuracy of predictions across all ML algorithms. The a20-index was selected because it is widely recognized in civil engineering and materials science as a clear and trustworthy gauge of predictive competence, especially when forecasting concrete behaviors. The a20-index determines the fraction of modelled values that lie within  $\pm 20\%$  of the corresponding measured values, giving a concrete benchmark for error that engineering practitioners regard as tolerable. In contrast to summary statistics like  $R^2$  and RMSE, which summarize the entire data set, the a20-index focuses exclusively on the portion of predictions that satisfy a tolerance cut-off that is meaningful in practice, especially when the stakes include safety margins and the inherent variability of construction materials. Additionally, when placed alongside the other a20 indices (notably a10 and a30), the a20 threshold settles into a widely accepted compromise. a10 is often considered too harsh, penalizing models for disparities that would not compromise workability. a30, on the other hand, is frequently dismissed as too forgiving, allowing models to appear trustworthy even when significant inaccuracies go unnoticed. The a20-index thus occupies a sound, practical middle ground, revealing the model's dependability in contexts where engineering judgement is paramount.

According to Fig. 7, the a20-index values span a range of 0.64 to 0.97, with the DTR algorithm exhibiting the lowest accuracy and the GPR and MLPR algorithms showcasing the highest accuracy. The other algorithms present acceptable accuracy levels, excluding the DTR model. Consequently, based on the a20-index results, all models, except for DTR, exhibit satisfactory performance in estimating concrete CS.

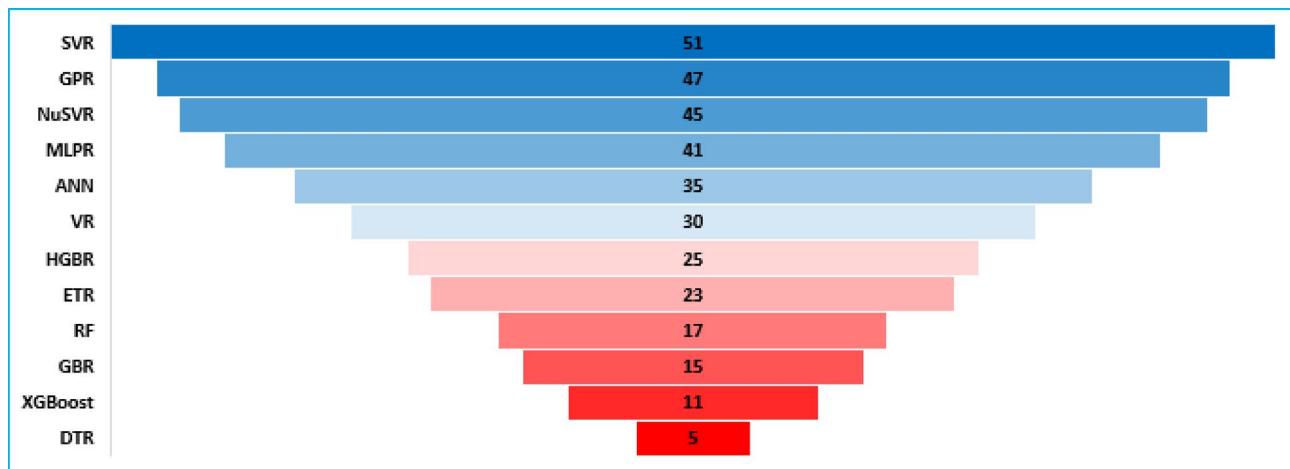


**Fig. 7.** Evaluating the ML algorithms' accuracy in predicting the concrete CS using the a20-index. ANN: Artificial neural network; SVR: Support vector regression; GPR: Gaussian process regression; ETR: Extra tree regressor; DTR: Decision tree regressor; GBR: Gradient boosting regressor; HGBR: Histogram-based gradient boosting regressor; XGBoost: Extreme gradient boosting; NuSVR: Null-space SVR; VR: Voting regressor; RF: Random forest; MLPR: Multilayer perceptron regression.

The results from evaluating the performance of a ML model may differ from one to another based on the metric which is selected as the focal point of the evaluation. Each of the metrics capture different elements of performance which include the level of error, the explanation of variance, the error in prediction, and the degree of robustness. In this case, a multi-criteria scoring model is preferred to provide a more balanced outcome and therefore, in this case the metrics selected are  $R^2$ , MAPE, RMSE, VAF, and a20-index. Each ML model

	R <sup>2</sup>	Score	MAPE	Score	RMSE	Score	VAF [%]	Score	a20-index	Score	Ranking score
DTR	0.3401	1	0.20	1	12.33	1	60.1	1	0.64	1	5
SVR	0.9647	12	0.04	8	2.85	12	98.2	12	0.96	7	51
NuSVR	0.9418	11	0.05	7	3.66	10	97.1	10	0.96	7	45
GPR	0.9400	10	0.05	7	3.50	11	97.3	11	0.97	8	47
XGBoost	0.7391	2	0.11	3	7.75	2	86.4	2	0.82	2	11
RF	0.7659	4	0.12	2	7.34	4	87.6	4	0.83	3	17
ETR	0.8009	5	0.10	4	6.77	5	89.8	5	0.87	4	23
GBR	0.7595	3	0.11	3	7.44	3	87.2	3	0.83	3	15
HGBR	0.8060	6	0.11	3	6.69	6	89.9	6	0.87	4	25
ANN	0.8947	8	0.07	5	4.92	8	94.6	8	0.95	6	35
VR	0.8158	7	0.10	4	6.51	7	90.3	7	0.88	5	30
MLPR	0.9126	9	0.06	6	4.49	9	95.6	9	0.97	8	41

**Table 4.** Ranking of ML models based on statistical performance metrics using the hold-out validation method.



**Fig. 8.** Ranking of ML models based on using the hold-out validation method. ANN: Artificial neural network, SVR: Support vector regression; GPR: Gaussian process regression; ETR: Extra tree regressor; DTR: Decision tree regressor GBR: Gradient boosting regressor; HGBR: Histogram-based gradient boosting regressor; XGBoost: Extreme gradient boosting; NuSVR: Null-space SVR; VR: Voting regressor; RF: Random forest; MLPR: Multilayer perceptron regression.

was ranked per metric based on its raw performance. For example, the model with the highest R<sup>2</sup> received a score of 12 (indicating 1st place out of 12 models), the next best received 11, and so forth down to the model with the lowest R<sup>2</sup>, which received a score of 1. Each model was evaluated with the other models in a given metric competition. No weighing bias was introduced, and therefore, the final score was the outcome of every metric score. The ranking score in the final column of Table 4 is simply the sum of individual metric scores, reflecting the aggregate performance of each model. This clear model evaluation avoids bias towards models due to championing one performance metric. For instance, while SVR excelled in R<sup>2</sup> (0.9647), MAPE (0.04), and RMSE (2.85), it also ranked highly across VAF (98.2%) and a20-index (0.96), giving it a cumulative score of 51, the highest among all contenders. Similarly, GPR and NuSVR also demonstrated consistently high scores across metrics, securing strong overall rankings.

In Fig. 8, a schematic depiction delineates the total scores for each algorithm based on the comprehensive evaluation criteria. These overarching results distinctly position the SVR algorithm as the current frontrunner among its counterparts. However, it's imperative to note that these estimates rest on test datasets, and the algorithms' performance awaits confirmation through rigorous testing on new unseen datasets to ensure their sustained accuracy.

K-fold cross-validation is a widely accepted method for validating the performance of ML models. The entire dataset is partitioned into K equally sized subsets, known as folds. For every iteration, a single fold serves as the holdout test set, while the concatenation of the remaining K-1 folds is utilized to train the model. This rotation is carried out K distinct times, guaranteeing that every fold is designated as the test set once. The performance metrics from every cycle are subsequently averaged, yielding a composite score that mitigates the influence of any one particular split. By employing this procedure, the analysis confirms the model's capacity to generalize,

as every observation is subjected to testing while simultaneously being part of the training pool across the entire K passes. K-fold cross-validation serves a crucial role in ML by discouraging overfitting. Instead of having the model latch onto the idiosyncrasies of just one training set, K-fold forces it to encounter varied subsets, compelling it to learn useful patterns that hold across the entire dataset. Averaging performance scores across these multiple folds yields a reliability that a lone train-test split cannot match. This is especially beneficial in situations where the dataset is small, as every observation gets its day in court both for training and for validation. K-fold also streamlines the model-selection process, offering a fair playground to compare multiple algorithms or fine-tuned hyperparameters. The end result is a clearer, more detailed picture of how well a model might perform on unseen data.

We applied 5-fold cross-validation (K=5) to rigorously evaluate the ML models. The complete dataset was divided into five equal parts; one part served as the test set for every fold while the remaining four were combined to form the training set. By rotating the test set across all five parts, we guaranteed that every observation contributed to the training and the validation process. This practice produces a robust and dependable estimate of how well the models can predict the CS of concrete incorporating RHA. The choice of 5 folds strikes a good balance, granting us reliable performance metrics without excessively prolonging training times, thereby enhancing our understanding of each model's capacity to generalize to unseen samples.

Table 5 summarizes the comparative performance of the ML approaches assessed using three primary evaluation metrics ( $R^2$ , RMSE, and VAF). For every fold, the metrics are computed and thereafter used to order the models. Reviewing the table, the SVR model consistently records the highest  $R^2$  scores across every fold, suggesting its strong capability to predict concentrated solids. For instance, in the first fold, SVR attains an  $R^2$  benchmark of 0.9518, outperforming every competitor. In contrast, the DTR variant consistently appears at the foot of the  $R^2$  hierarchy, evidencing weaker predictive quality. In Fold 1, DTR earns an  $R^2$  of merely 0.3312, a value that falls well below that of any alternative model considered. For RMSE, the SVR model records the smallest values, pointing to the least prediction error on record. In Fold 1, it settles at 2.98, staking a strong claim to the model's accuracy. On the other hand, the DTR model shows the highest RMSE values, especially in Fold 1,

Fold no.		SVR	GPR	NuSVR	MLPR	ANN	VR	HGBR	ETR	RF	GBR	XGBoost	DTR
Fold 1	$R^2$	0.9518	0.9312	0.9344	0.9087	0.8862	0.8123	0.7983	0.7931	0.7562	0.7503	0.7288	0.3312
	Score	12	11	10	9	8	7	6	5	4	3	2	1
	RMSE	2.98	3.82	3.99	5.13	5.87	7.21	7.66	8.03	8.57	8.94	9.25	13.81
	Score	12	11	10	9	8	7	6	5	4	3	2	1
	VAF (%)	97.2	96.14	95.83	94.12	94.01	88.86	88.04	87.43	87.12	86.39	85.74	58.31
	Score	12	11	10	9	8	7	6	5	4	3	2	1
Fold 2	$R^2$	0.9716	0.9594	0.9643	0.95	0.9349	0.8744	0.8649	0.8634	0.8319	0.8297	0.8113	0.4593
	Score	12	10	11	9	8	7	6	5	4	3	2	1
	RMSE	2.88	3.72	3.89	5.04	5.78	7.12	7.57	7.94	8.48	8.85	9.16	13.75
	Score	12	11	10	9	8	7	6	5	4	3	2	1
	VAF (%)	98.69	98.05	97.83	96.69	96.95	92.47	91.87	91.45	91.41	90.86	90.37	65.22
	Score	12	11	10	8	9	7	6	5	4	3	2	1
Fold 3	$R^2$	0.9905	0.9825	0.9883	0.9797	0.9683	0.9146	0.9073	0.9077	0.8789	0.8785	0.8617	0.5328
	Score	12	10	11	9	8	7	5	6	4	3	2	1
	RMSE	2.43	3.97	3.48	5.72	6.83	8.95	9.64	10.21	11.03	11.60	12.07	19.49
	Score	12	11	10	9	8	7	6	5	4	3	2	1
	VAF (%)	98.52	97.84	97.61	96.41	96.63	92.09	91.47	91.03	90.96	90.40	89.89	64.51
	Score	12	11	10	8	9	7	6	5	4	3	2	1
Fold 4	$R^2$	0.9713	0.9671	0.9623	0.9461	0.9298	0.8671	0.8568	0.8547	0.8223	0.8195	0.8006	0.4411
	Score	12	11	10	9	8	7	6	5	4	3	2	1
	RMSE	2.76	3.21	3.79	5.29	6.16	7.76	8.29	8.73	9.36	9.80	10.16	15.68
	Score	12	11	10	9	8	7	6	5	4	3	2	1
	VAF (%)	98.14	97.34	97.09	95.74	95.86	91.14	90.46	89.97	89.83	89.22	88.66	62.68
	Score	12	11	10	8	9	7	6	5	4	3	2	1
Fold 5	$R^2$	0.9590	0.9370	0.9439	0.9156	0.8931	0.8187	0.8046	0.7993	0.7624	0.7564	0.7349	0.3345
	Score	12	10	11	9	8	7	6	5	4	3	2	1
	RMSE	2.75	3.70	3.79	5.43	6.34	8.04	8.61	9.07	9.74	10.20	10.59	16.47
	Score	12	11	10	9	8	7	6	5	4	3	2	1
	VAF (%)	98.34	97.67	97.38	96.21	96.43	91.88	91.26	90.82	90.75	90.18	89.67	64.30
	Score	12	11	10	8	9	7	6	5	4	3	2	1
All folds	Sum of scores	180	162	153	131	124	105	89	76	60	45	30	15

**Table 5.** Ranking of ML models based on statistical performance metrics using the 5-fold cross-validation method.

where it reaches 13.81, indicating that the predictions are farther from the true values compared to other models. Turning to VAF, SVR again leads with a score that shows it explains the greatest proportion of variance. In Fold 1, it notches up 97.2%, the summit among all contenders. DTR, however, posts the lowest VAF, 58.31% in the same Fold, a reading that reveals it captures only a fraction of the data's underlying variation.

The model rankings provided in the accompanying table are derived from the overall score calculations made over the complete set of five cross-validation folds; in this system, a greater cumulative score indicates superior model performance. The SVR model exceeds every other candidate by this measure in each individual fold and attains the highest total score of 180. This result underlines SVR's consistent merit and stability from fold to fold in the validation process. Analyzing performance on a fold-by-fold basis confirms SVR's dominant position. In Fold 1, the model records the best  $R^2$ , the lowest RMSE, and the leading VAF, which together assure its first rank. Fold 2 sees SVR again on top, producing similarly strong  $R^2$  and VAF values, though the RMSE rises by a small margin. The same pattern persists in Fold 3, where  $R^2$  and VAF remain elevated and RMSE stays comparatively low. Fold 4 registers identical results: top  $R^2$  and VAF, a slight RMSE increase. Finally, Fold 5 again delivers peak  $R^2$  and VAF, paired with the best RMSE, reaffirming SVR's overall superiority.

The SVR, NuSVR, and GPR models outperformed the other methods in this work for three mutually reinforcing reasons that matched the problem's conditions. First, their architectures suit moderate datasets (like the 500 samples here) where deep learners, including ANNs, risk overfitting without heavy and sometimes unbalanced regularizations. Second, they employ kernel functions (specifically the radial basis function) that enable the mapping of input features into high-dimensional spaces where nonlinear trends can be effectively captured. Lastly, the three methods embed regularization: SVR and NuSVR impose it via the penalty parameters, while GPR incorporates it through the Bayesian priors. Together, these design choices supported reliable generalization and strong predictive accuracy through every evaluation phase.

In summary, the SVR model leads in every validation fold, showing the highest predictive accuracy. Its strength across all five partitions supports the model's reliability for estimating the CS of RHA concrete. Conversely, the DTR model places at the bottom in each measure, underscoring its relative unsuitability for this application. The 5-fold cross-validation adopted here enhances the credibility of the results by preventing reliance on a single data split; instead, it confirms model behavior through a thorough evaluation over multiple data segments. This multi-partition method delivers a robust and consistent basis for judging the model's potential to generalize and produce precise forecasts.

To ensure a robust evaluation of the trained algorithms in estimating concrete CS, we employ previously unused datasets from prior publications as validation datasets. Initially, we scrutinize the models' performance on the 24 data points presented in Bui et al.<sup>15</sup>, outlined in (Table 6). These data points share identical parameters with our study, encompassing the same considerations in concrete sample preparation and testing methodology to ascertain the 28-day CS. These external data points were used exclusively for generalization assessment; no external samples were used for training or hyperparameter adjustment. This insertion aims to gauge how

W/B	C (kg/m <sup>3</sup> )	RHA (kg/m <sup>3</sup> )	W (kg/m <sup>3</sup> )	SP (kg/m <sup>3</sup> )	FA (kg/m <sup>3</sup> )	CA (kg/m <sup>3</sup> )	CS (MPa)
0.30	550.00	0.00	165.00	5.60	550.00	1283.00	63.50
0.30	495.00	55.00	165.00	5.80	546.00	1273.00	72.80
0.30	468.00	82.00	165.00	6.10	543.00	1267.00	75.10
0.30	440.00	110.00	165.00	6.30	540.00	1261.00	78.20
0.32	500.00	0.00	160.00	5.50	567.00	1324.00	59.60
0.32	450.00	50.00	160.00	5.72	562.00	1313.00	68.80
0.32	425.00	75.00	160.00	6.00	560.00	1307.00	72.20
0.32	400.00	100.00	160.00	6.22	558.00	1301.00	72.70
0.34	500.00	0.00	170.00	5.00	560.00	1305.00	57.90
0.34	450.00	50.00	170.00	5.00	555.00	1294.00	66.60
0.34	425.00	75.00	170.00	5.00	553.00	1290.00	67.20
0.34	400.00	100.00	170.00	5.00	551.00	1285.00	69.30
0.30	550.00	0.00	165.00	6.60	550.00	1283.00	88.50
0.30	495.00	55.00	165.00	6.80	546.00	1273.00	95.20
0.30	468.00	82.00	165.00	7.19	543.00	1267.00	96.00
0.30	440.00	110.00	165.00	7.43	540.00	1261.00	98.10
0.32	500.00	0.00	160.00	6.50	567.00	1324.00	85.70
0.32	450.00	50.00	160.00	6.76	562.00	1313.00	91.60
0.32	425.00	75.00	160.00	7.08	560.00	1307.00	93.40
0.32	400.00	100.00	160.00	7.36	558.00	1301.00	94.30
0.34	500.00	0.00	170.00	6.40	560.00	1305.00	82.80
0.34	450.00	50.00	170.00	6.40	555.00	1294.00	89.70
0.34	425.00	75.00	170.00	6.40	553.00	1290.00	90.30
0.34	400.00	100.00	170.00	6.40	551.00	1285.00	91.10

**Table 6.** Evaluation data points available in Bui et al.<sup>15</sup>.

the trained models perform on independent data produced under differing experimental arrangements. This approach strengthens the credibility of the models. It addresses model robustness and transferability, which is especially important in ML applications to concrete materials where variability in raw materials and test conditions is common.

In Fig. 9, the outcomes predicted by each algorithm are showcased on these data points, juxtaposed with the experimental results. While a majority of the models exhibit behavior akin to the experimental outcomes, it is noteworthy that not all algorithms yield acceptable and accurate results, reflected in  $R^2$  values spanning from 0.46 to 0.94. Notably, the SVR, GPR, and NuSVR models demonstrate superior accuracy on the test dataset in our study, showcasing the best performance on these data points among other algorithms. This attests to the sound training of these algorithms. The MLP and ANN algorithms secure the fourth and fifth positions in terms of accuracy, achieving  $R^2$  values of 0.84 and 0.82, respectively. Conversely, other algorithms exhibit subpar performance, registering  $R^2$  values within the range of 0.46 to 0.71. Notably, the DTR algorithm delivers the least accuracy, with an  $R^2$  value of 0.46.

In this analytical phase, a meticulous investigation was conducted into the performance of each intricately trained ML model using an additional set of six data points, which underwent CS testing as detailed by Chao-Lung et al.<sup>61</sup>. These specific data points, elucidated in (Table 7), deviate solely in the geometric configuration of samples, transitioning from cubic to cylindrical. The primary objective of this comparative analysis was to assess the adaptability of the ML models developed in our study to the diverse structural forms of concrete samples. We acknowledge that the differing shapes (cube vs. cylinder) can influence the CS results because each geometry redistributes stress and triggers failure in distinct patterns. Nonetheless, the goal of this comparison was to assess how well the ML models can generalize and remain robust when applied to datasets that contain only mildly different specimen silhouettes, even in the absence of direct geometric normalization.

Figure 10 serves as a visual representation, illustrating the correlation between the CS values estimated by each ML algorithm and the corresponding values obtained from laboratory tests conducted by Chao-Lung et al.<sup>61</sup>. The  $R^2$  values derived from these ML algorithms present a spectrum ranging from 0.50 to 0.98. A standout performer is the SVR model, showcasing exceptional accuracy with an impressive  $R^2$  of 0.98. The NuSVR and GPR models also exhibit noteworthy precision, achieving  $R^2$  values of 0.95 and 0.93, respectively. Conversely, models such as DTR, XGBoost, RF, GBR, HGBR, and VR, with  $R^2$  values below 0.80, demonstrate comparatively lower accuracy. Meanwhile, MLP, ANN, and ETR models showcase acceptable accuracy, with  $R^2$  more than 0.80. It is essential to highlight that SVR and DTR models record the highest and lowest accuracies, with  $R^2$  values of 0.98 and 0.50, respectively, echoing trends observed in previous evaluations.

A comprehensive examination of these results reveals the proficiency of the SVR model in accurately estimating the concrete CS, particularly within the context of the dataset utilized in this study. This finding not only underscores the robustness of the SVR model but also prompts further exploration into the factors contributing to its superior predictive performance in this specific application. Additionally, these insights into the comparative accuracies of various ML models provide valuable guidance for selecting appropriate models in similar contexts, contributing to the ongoing refinement of predictive methodologies in the domain of concrete CS estimation.

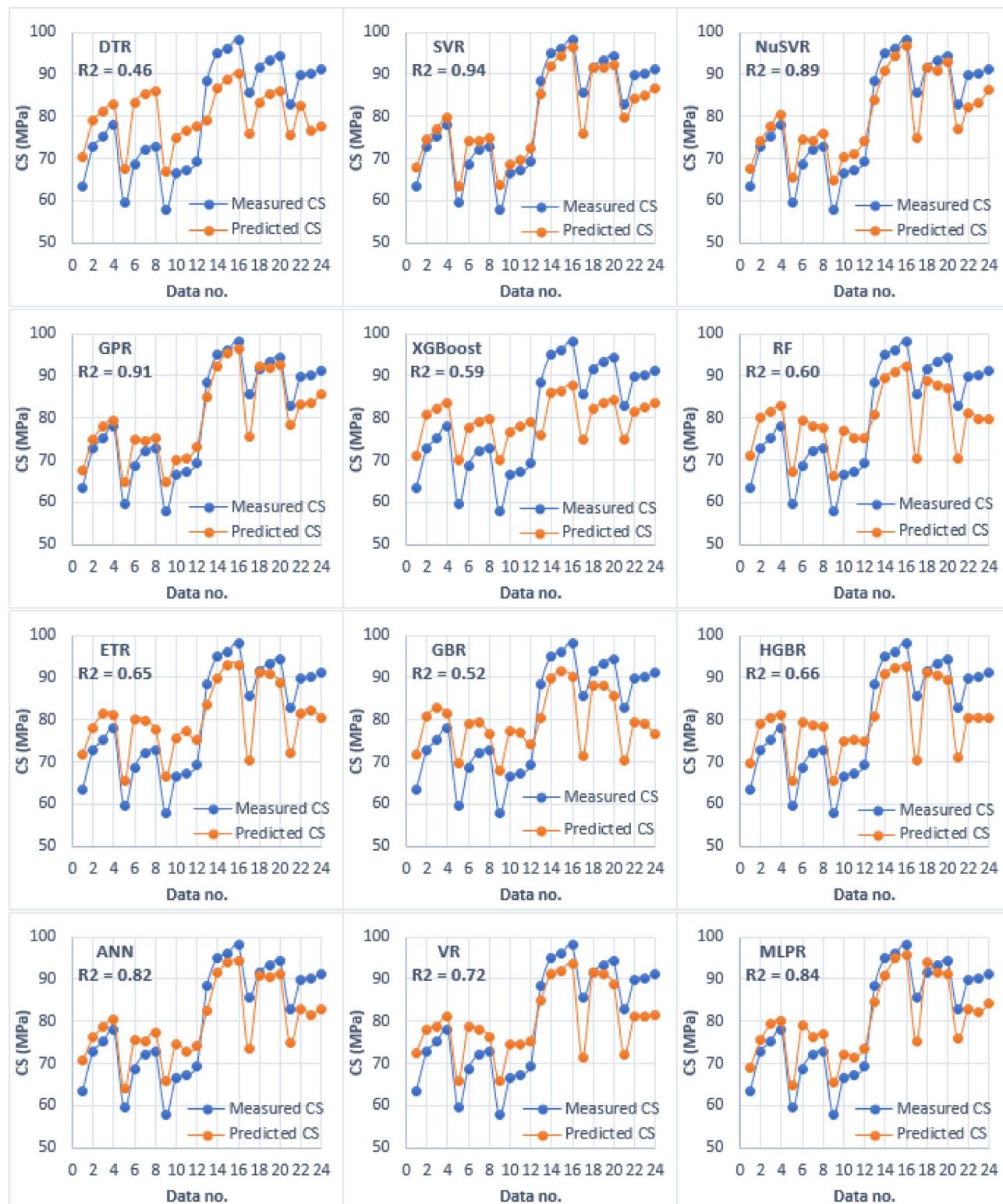
The profound expertise demonstrated by the SVR model in estimating concrete CS, as evidenced through the comprehensive evaluation of results in this study, underscores its efficacy as a robust predictive tool. The successful application of the SVR model to the dataset employed herein attests to its nuanced understanding of various parameters and their intricate relationships with the model output (CS). This mastery positions the SVR model as a valuable asset for predictive modeling in concrete engineering.

Motivated by the proficiency of the SVR model, a meticulous exploration is initiated to unravel the influence of the RHA parameter in the concrete mixing plan on CS. This investigation is methodically conducted using three distinct datasets comprising 20 data points as novel test datasets. The systematic variation of the RHA parameter's value within its range (0 to 190 kg/m<sup>3</sup>) in 10 kg/m<sup>3</sup> increments while holding other parameters constant, according to (Table 8), forms the basis of this inquiry. The predictions, shown in Fig. 11, clearly demonstrate a parabolic pattern where CS increases with increasing RHA content to an optimal level (around 80–100 kg/m<sup>3</sup>). Beyond this, further increases in RHA content result in a gradual reduction in strength. This phenomenon illustrates a saturation effect, typically due to the pozzolanic reactivity of RHA, which improves strength to a certain level of replacement and then negatively affects it as the replacement level increases because of excessive RHA leading to dilution of cementitious materials and increased workability problems. It should be noted that the best observed RHA value is bound to a specific dataset and the chemical and physical properties of RHA relevant to this study such as particle size and the degree to which it has been charred, along with the complete mixture design which includes the W/B, SP, and aggregate size distribution. To provide example, finer RHA particles with greater amorphous silica content is more reactive and hence optimal dosage is shifted to higher value. In contrast, coarser and less reactive RHA shifted the optimum to lower value. Thus, the optimal range as described above should only be considered as relevant in the context of the experimental materials and proportions. Extrapolation to other contexts would necessitate recalibration or retraining of the relevant models with localized material properties and mix designs to maintain accurate strength and optimal performance value targets.

The examination firmly establishes that the addition of RHA to the concrete mixture holds the potential to enhance its CS, contingent upon the intricate interdependence of various parameters. This nuanced insight contributes to the ongoing discourse on optimizing concrete mix designs for superior performance.

### GUI for practical deployment

To enable seamless integration of the ML models into everyday engineering workflows, a dedicated standalone GUI was crafted using the PyQt5 toolkit in Python. This user-friendly desktop application, illustrated in (Fig.



**Fig. 9.** Evaluating the ML algorithm predictions against the test results conducted by Bui et al.<sup>15</sup>.

<sup>12</sup>), prompts input of seven critical mix design parameters, including W/B, C, RHA, total W, SP, FA, and CA. Users can swiftly obtain the predicted 28-day CS of concrete mixtures incorporating RHA by entering these values. Additionally, the interface permits selection from twelve pre-trained ML models with the underlying models serialized via the joblib library to guarantee rapid initialization and optimal computational performance.

The GUI was crafted to function seamlessly across Windows, Linux, and macOS, making it easy to access and use during critical on-site decision points. Built-in input checks confirm that the mix design parameters

W/B	C (kg/m <sup>3</sup> )	RHA (kg/m <sup>3</sup> )	W (kg/m <sup>3</sup> )	SP (kg/m <sup>3</sup> )	FA (kg/m <sup>3</sup> )	CA (kg/m <sup>3</sup> )	CS (MPa)
0.23	783.00	22.76	212.00	3.60	344.00	933.00	66.00
0.35	571.00	0.00	219.00	1.00	633.00	933.00	56.00
0.35	514.00	22.74	218.00	1.40	608.00	933.00	61.00
0.35	457.00	43.81	216.00	2.60	582.00	933.00	60.00
0.35	400.00	63.26	215.00	3.70	557.00	933.00	54.00
0.47	383.00	22.74	221.00	0.30	737.00	933.00	47.00

**Table 7.** Evaluation data points available in Chao-Lung et al.<sup>61</sup>.

stay within proven, empirically grounded ranges from the authors' dataset. For instance, if a user tries an unrealistically high W/B or RHA value, the tool instantly highlights it for revision. The application serves two main audiences: practitioners can quickly test different mix designs without incurring the expense and delay of full lab testing, while researchers can tweak parameters systematically to produce synthetic datasets for simulations or optimization studies.

Figure 12 presents the interface returning a predictive CS value derived from the chosen ML model. In the current scenario, the SVR model estimates a CS of 75.40 MPa. Though this level may appear surprising for concretes incorporating RHA, the forecast is supported by a carefully optimized set of parameters: W/B = 0.3, C = 468 kg/m<sup>3</sup>, RHA = 82 kg/m<sup>3</sup>, SP = 6.1 kg/m<sup>3</sup>, FA = 543 kg/m<sup>3</sup>, and CA = 1267 kg/m<sup>3</sup>. Together, these variables encourage a compact microstructure and improved pozzolanic reactivity. The result underlines the interface's ability to quantify nonlinear and synergistic interactions that govern strength gain.

It should be emphasized that how suitable the predictions are and how well they can be applied in practice is dependent on the materials and data used in the model building process. For instance, the RHA used in this study had particular characteristics such as high amorphous silica content and low loss on ignition (LOI) and fine particle size around 15 micrometers, which was obtained through grinding and burning at 650–750 °C. Hydrothermally processed RHA with coarse particles, increased crystalline content, or high LOI can drastically modify the pozzolanic activity and hydration speeds and CS in a manner that the existing models do not account for. In order to solve this, users from different sources or grades of RHA are advised to re-train the models with datasets most relevant to their materials. Other approaches, such as transfer learning or implementation of corrective factors based on material property testing can make the model more flexible. The reliability of the model for various applications can be enhanced by adding metadata on the RHA properties for later versions of the GUI.

### Model interpretability using SHAP analysis

ML models frequently present themselves as black boxes, obscuring the reasoning behind their predictions. SHAP values counter this opacity by providing a rigorously grounded way to dissect model decisions, quantifying how much each feature sways a given prediction. In contrast to standard feature importance metrics, SHAP discloses not just whether a feature is influential but also the precise amount and direction of its effect. It elegantly accommodates interactions among variables, permits interpretation on both local and global scales, and thus serves diagnostic efforts and specialized domains alike, including engineering design tasks. Consequently, SHAP empowers practitioners to grasp the parameters the model weighs most heavily and the underlying logic, revealing paths for targeted refinement or domain-oriented tuning.

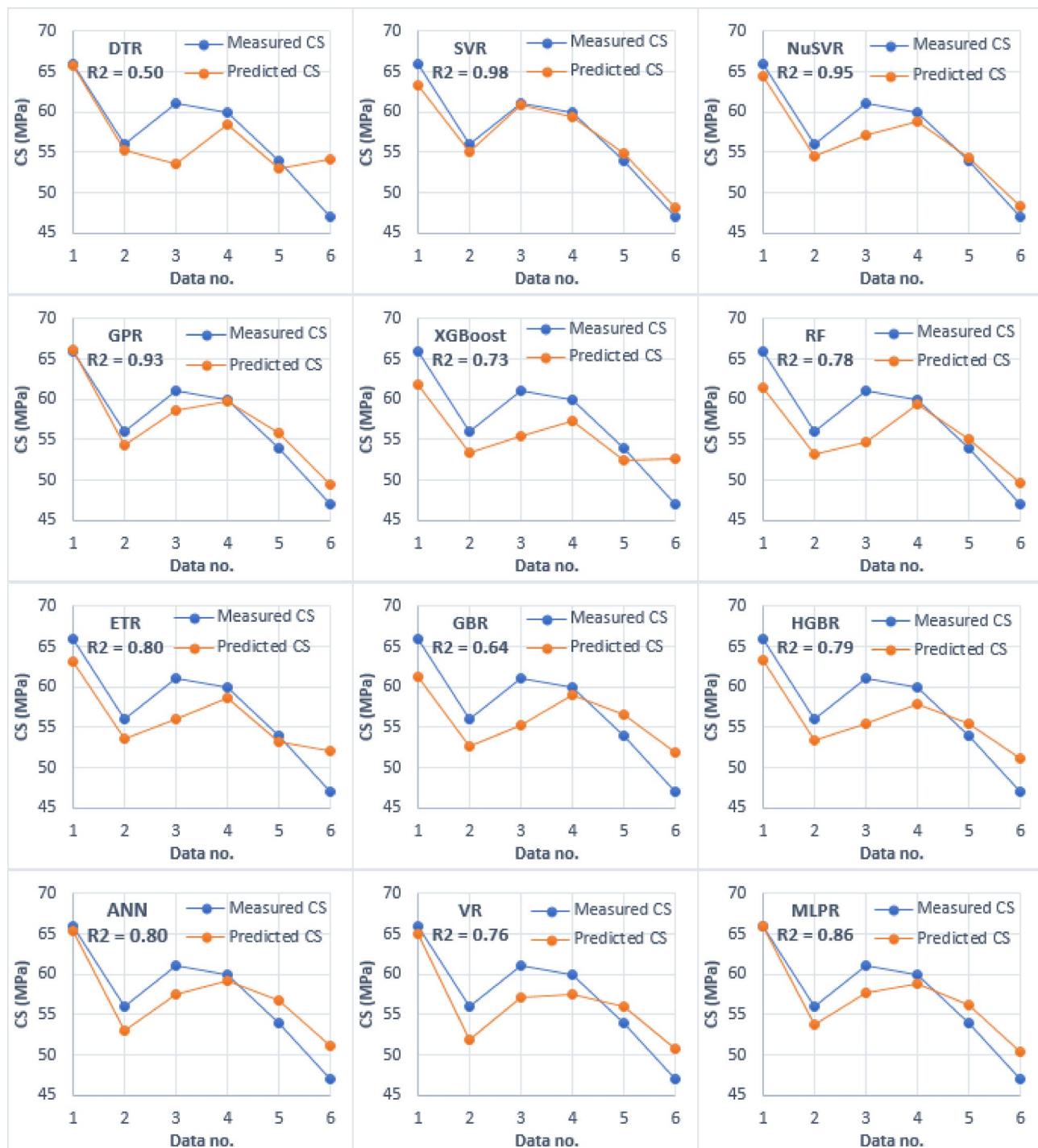
This section delivers a thorough SHAP investigation for the three top-performing regression models (SVR, NuSVR, and GPR). SHAP results for these models are shown in (Fig. 13). In every summary plot, the vertical axis lists features, while the horizontal axis quantifies each feature's SHAP contribution to predicting CS. Every dot corresponds to a single instance, color-coded by the feature value (red for high, blue for low).

The SHAP summary plot for the SVR model pinpoints SP, CA, and W/B as the leading influential features. SP exhibits a wide scatter of positive SHAP values at elevated levels, pointing to a vigorous positive impact on the predicted output. In contrast, W/B and CA yield more concentrated SHAP value distributions, indicating their contributions vary more sensitively around a mid-range effect. This confirms that SVR adeptly maps both linear and non-linear interactions, especially among mix design variables.

The NuSVR model reaffirms the dominance of SP, W/B, and CA as top features, yet its SHAP values cluster more widely around zero for W/B and CA, highlighting a subtler but pervasive influence across the sample space. The gradient coloring (shifting from blue to red) corroborates that below-average SP values correlate with negative SHAP, whereas elevated SP levels consistently raise predictions. This more consistent response pattern likely underpins NuSVR's superior generalization, particularly when extrapolating to samples with moderate or borderline values in the key features.

The GPR model tells a different story altogether. Its SHAP summary plot reveals SHAP values that are tightly grouped for every feature, all skewed negatively; C, FA, and W are the strongest offenders. The SHAP values are tightly packed and consistently below zero, indicating that GPR uniformly forecasts lower outcomes, suggesting a built-in bias toward conservative, smooth estimates. Unlike the SVR and NuSVR models, which exhibit sharper reactions to small perturbations, GPR's probabilistic framework mitigates the impact of individual features, yielding gentler and less erratic surface predictions.

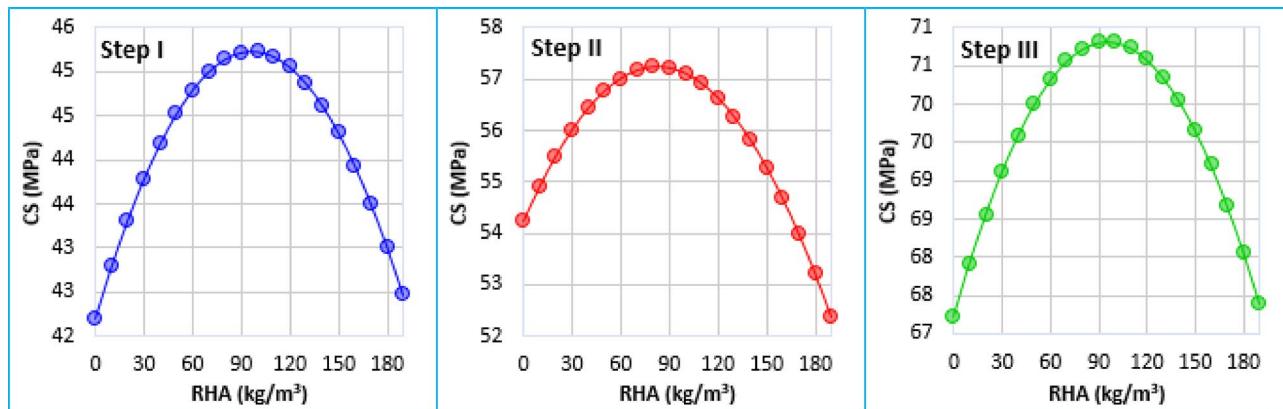
When viewed side by side, these SHAP findings highlight the distinct feature hierarchies that each algorithm adopts. SVR and NuSVR grant higher leverage to the blend-centric parameters (specifically SP, W/B and CA),



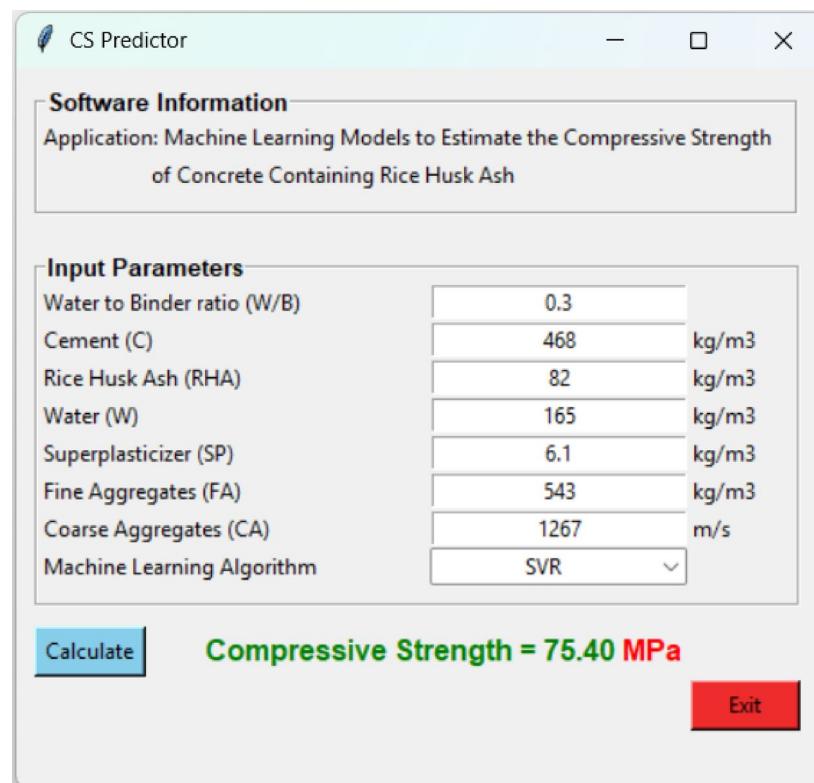
**Fig. 10.** Evaluating the ML algorithm predictions against the test results conducted by Chao-Lung et al.<sup>61</sup>.

Step	W/B	Cement (kg/m <sup>3</sup> )	W (kg/m <sup>3</sup> )	SP (kg/m <sup>3</sup> )	FA (kg/m <sup>3</sup> )	CA (kg/m <sup>3</sup> )
Step I	0.35	407	178	36	620	944
Step II	0.32	490	180	35	650	1120
Step III	0.30	440	165	7.43	540	1261

**Table 8.** The values of constant parameters considered in each testing step.



**Fig. 11.** Progressive evaluation of concrete CS using the SVR model across three steps, modulating RHA parameter from 0 to 190 kg/m<sup>3</sup> in 10 kg/m<sup>3</sup> increments.

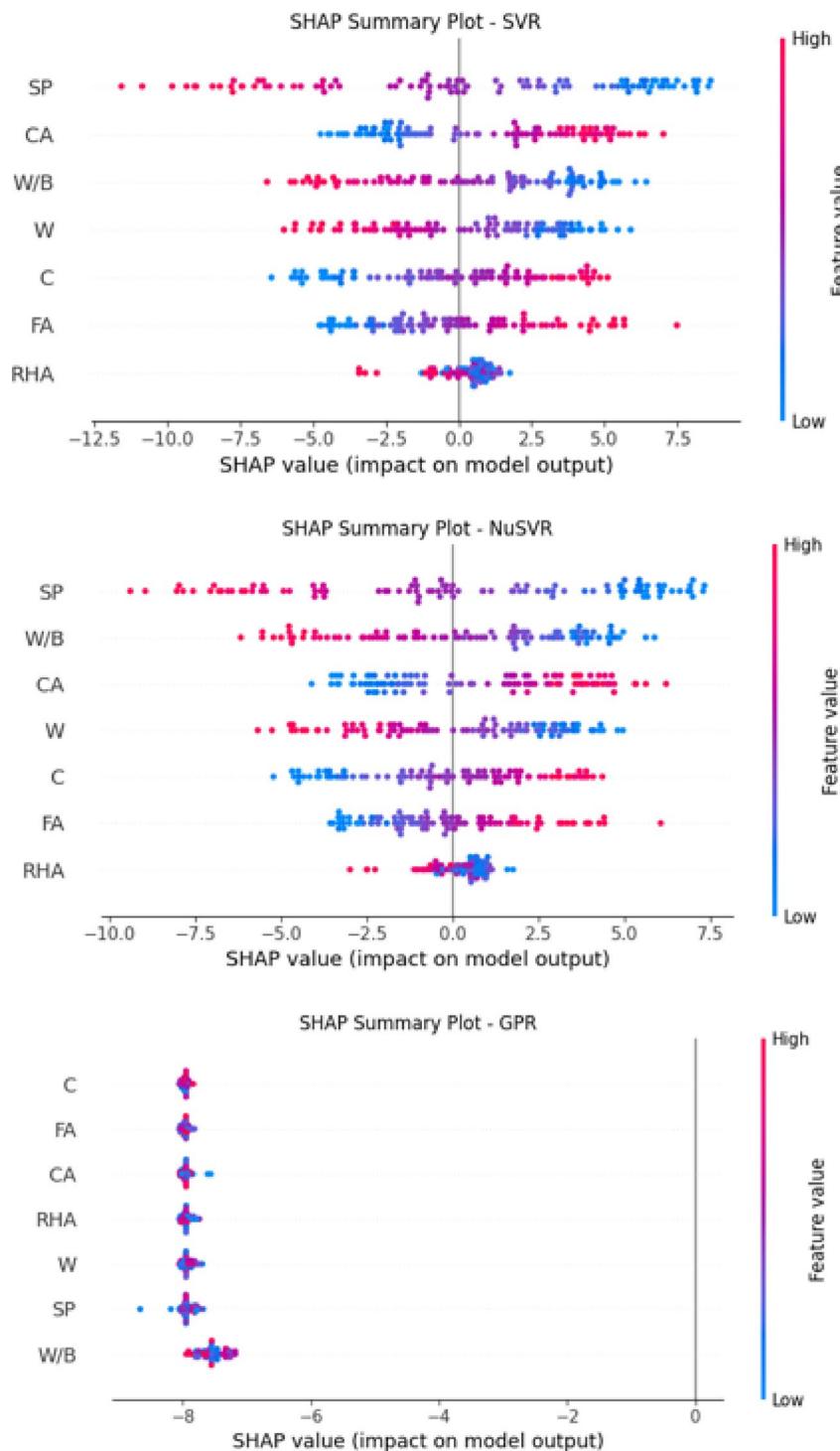


**Fig. 12.** A ML-based GUI to estimate concrete CS.

while GPR focuses on the binder and fluid constituents, yet with far smaller swings in output. Such divergence in explanatory power clarifies each model's reasoning and enables practitioners to choose the most suitable approach for optimizing concrete mix designs and anticipating durability within a single, unified interpretive framework.

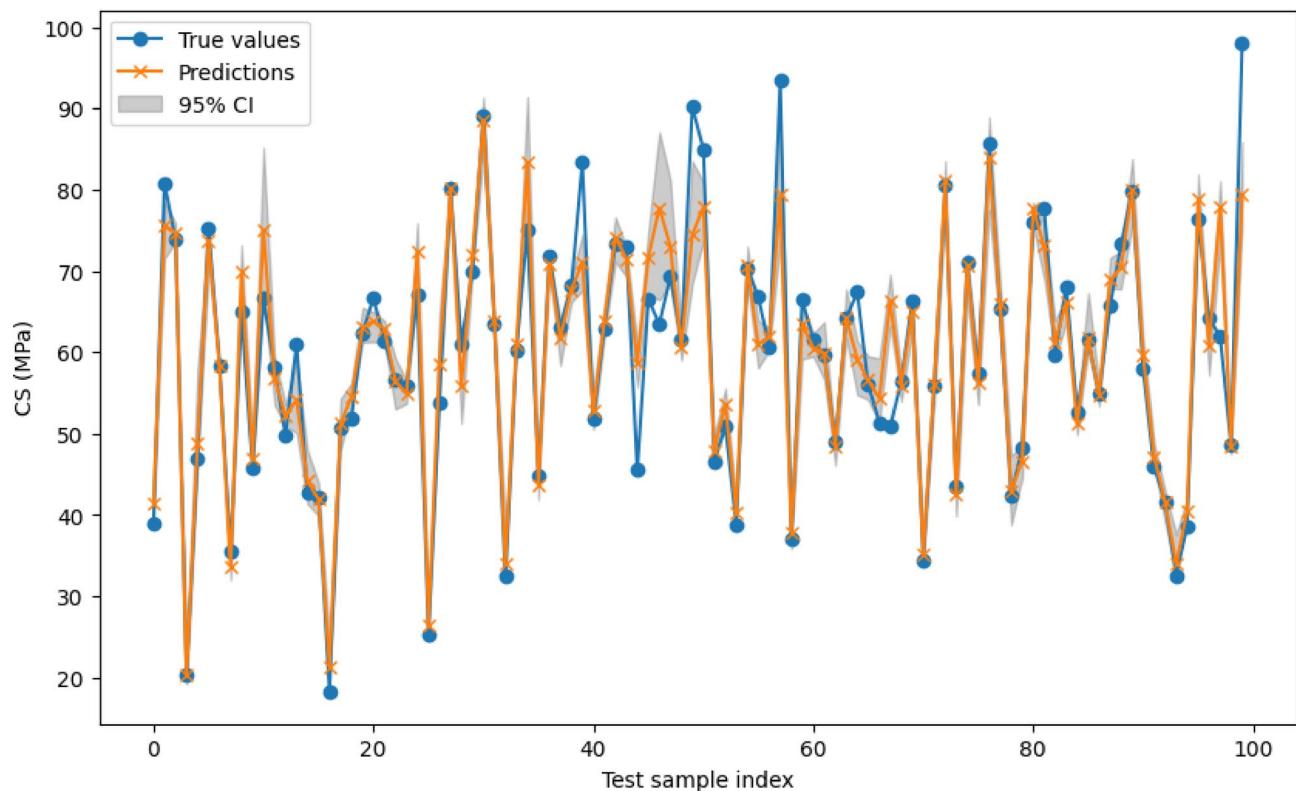
### Uncertainty quantification

To evaluate the reliability of the predictions made by the most robust model (SVR), an uncertainty quantification approach based on bootstrapping was implemented. Specifically, the SVR model was used to forecast the CS on each of the 1000 bootstrap resamples created from the test set. This approach was adopted to estimate empirical 95% confidence intervals for each prediction, based on the perturbation of the model's inputs and performance variability. The SVR model also provides predictive distributions; hence, point estimates and uncertainty metrics can be derived. Using the predictive distributions, the 95% confidence bounds for each CS were calculated, representing the ranges expected to contain the actual CS values with 95% certainty.



**Fig. 13.** SHAP summary plots for SVR, GPR, and NuSVR models.

In Fig. 14, the predicted CS values are surrounded by shaded areas that depict the corresponding confidence intervals. The figure displays the model's prediction uncertainty at different test points by the width of the shaded areas, with narrower regions indicating higher confidence and broader regions representing higher uncertainty. This disparity in confidence is most noticeable in regions where the model's predictions diverge considerably from the actual CS values. The uncertainty shifts in these different regions are a result of the posterior predictive distribution of the uncertainty of the model's confidence in the data and the noise present in the data.



**Fig. 14.** SVR predictions with bootstrapped 95% confidence intervals.

### Key limitations and suggestions

This study outlines specific limitations that require consideration, accompanied by valuable suggestions for future research to overcome these challenges:

- The investigation looked into a limited set of input variables predicting the CS of RHA concrete. However, the model's predictive capability and range of applicability could still be improved by incorporating additional relevant considerations such as curing temperature and duration, and ambient humidity, as well as the type and dosage of cementitious materials used, the aggregate shape and gradation, the chemical characteristics of RHA (like silica or alkali content), and certain chemical admixtures including retarders or accelerators. Considering these variables could help better explain intricate material and environmental interactions and improve the model's applicability across different conditions and materials, thus enhancing the model's overall robustness.
- Perhaps the most significant drawback of this study is the singular emphasis on 28-day CS as the target for prediction. While 28-day strength is regarded as a benchmark during the evaluation of concrete's quality, it inadequately assesses the performance of RHA-based concrete over time, especially considering the prolonged pozzolanic activity and hydration kinetics that can improve strength well beyond 28 days. The lack of 56 and 90-day strength data limits the model's use for structural durability life cycle analysis. Researchers in the future should strive to incorporate the extended curing age strengths into multi-output or time-evolving predictive models to better capture the mechanics of sustainable concrete.
- The assessment primarily centered around twelve ML models. To comprehensively evaluate various algorithms for predicting CS, future research could broaden the comparison by including a more extensive range of ML models.
- Depending solely on laboratory data for training and testing ML models, this study lacks validation in real-world scenarios. Future research should apply the models to engineering projects, comparing predicted CS values with observed values for robust validation.
- The study did not highlight the interpretability of ML models. Future research might explore techniques to enhance model interpretability, offering engineers insights into the underlying factors influencing CS predictions.
- The GUI developed for estimating the CS of RHA concrete is user-friendly and effective, but its predictions are limited to the ranges of input parameters seen during training. If predictions are made for mixtures that fall outside the training domain (such as very high or very low binder-to-aggregate ratios, non-traditional aggregate sources, or unusually reactive pozzolans), the accuracy of the results may decline. To avoid this, users are encouraged to restrict their input to conditions that mirror the validated dataset or to supplement

the GUI with new experimental data followed by additional training to strengthen its predictive capability in unexplored material domains.

By addressing these limitations and incorporating the proposed directions for future work, this study has the potential to significantly improve the accuracy, reliability, and practicality of ML models in predicting concrete compressive strength.

## Conclusions

This research compared the predictive accuracy of twelve ML algorithms for estimating the CS of concrete containing RHA. A dataset comprising 500 laboratory specimens was supplemented with 30 additional validation records drawn from published studies. Model performances were quantified through a range of indicators. The Stepwise selection procedure distilled seven principal input parameters that most significantly govern the material's strength development.

The main findings and their practical significance are:

- SVR, GPR, and NuSVR stood out as the top techniques, achieving deep accuracy for both the hold-out and k-fold cross-validation tests with  $R^2$  values exceeding 0.93. On the other hand, the DTR showed the weakest performance with  $R^2$  values below 0.53. This was due to DTR's limitations, primarily its sensitivity to overfitting small. Unlike ensemble or regularized models, single decision trees tend to overfit to the training data, drastically reducing their performance on new and unseen data. This is a critical drawback for capturing the complex and nonlinear interactions that define the CS of RHA concrete. The results strongly demonstrate the impact of the model choice on small and heterogeneous datasets, where SVR and GPR provide more robustness due to regularization and probabilistic frameworks.
- Since each metric captures a particular facet of predictive performance, basing model selection solely on one could mislead. Incorporating several indicators in the evaluation process affords a clearer, more complete picture of model behavior across the operational range.
- The widening performance disparity between the training dataset and the independent validation set highlights the critical need for external verification. Such validation helps gauge how well a model will perform in practical applications, thereby curbing the dangers of overfitting to the original sample.
- Findings showed that adding RHA improves the CS of concrete, especially when carefully balanced with variables such as the W/B, C, and the gradation of aggregate particles.
- To empower civil engineers, an intuitive GUI was created that puts ML prediction tools at their fingertips. Through this software, users can quickly estimate CS and iteratively design the ideal mix, equipping practitioners with a timely, data-driven resource that promotes greener building practices. The GUI and the best-performing predictive models can slot seamlessly into existing design platforms or site-level workflows, translating sophisticated data science into streamlined workflows and reducing the need for exhaustive experimental programs. This functional interface connects cutting-edge analytics with everyday engineering tasks, shortening the path from discovery to onsite performance.

## Data availability

The employed dataset and the developed GUI not available due to restrictions imposed by research sponsors, ongoing analysis for future studies, and the necessity to maintain data confidentiality until further validation and publication. If someone wants to request the dataset or the GUI from this study, he/she should contact Dr. Arsalan Mahmoodzadeh. In this way, they are shared with them on responsible request.

## Code availability

The codes used in this paper are available in the following link. <https://mega.nz/file/KGgDFb4J#UUId5-SIaJO76V3fRoyTFr6lMDlxot3cMJDgQwm4Bk4>.

Received: 24 April 2025; Accepted: 9 October 2025

Published online: 07 November 2025

## References

1. Mikulčić, H., Klemić, J. J., Vujanović, M., Urbaniec, K. & Duić, N. Reducing greenhouse gasses emissions by fostering the deployment of alternative Raw materials and energy sources in the cleaner cement manufacturing process. *J. Clean. Prod.* **136**, 119–132. <https://doi.org/10.1016/j.jclepro.2016.04.145> (2016).
2. Apranti, E. A huge number of artificial waste material can be supplementary cementitious material (SCM) for concrete production – a review part II. *J. Clean. Prod.* **142**, 4178–4194. <https://doi.org/10.1016/j.jclepro.2015.12.115> (2017).
3. Poon, C. S., Kou, S. C. & Lam, L. Compressive strength, chloride diffusivity and pore structure of high performance Metakaolin and silica fume concrete. *Constr. Build. Mater.* **20** (10), 858–865. <https://doi.org/10.1016/j.conbuildmat.2005.07.001> (2006).
4. Huang, Y., Lei, Y., Luo, X. & Fu, C. Prediction of compressive strength of rice husk Ash concrete: A comparison of different metaheuristic algorithms for optimizing support vector regression, *Case Studies in Construction Materials*, **18**, e02201 <https://doi.org/10.1016/j.cscm.2023.e02201> (2023).
5. Parhi, S. K. & Panigrahi, S. K. Application of metaheuristic spotted hyena optimization in strength prediction of concrete. *Metaheuristics-Based Materials Optimization* 229–248. <https://doi.org/10.1016/B978-0-443-29162-3.00008-3>. (Elsevier, 2025).
6. Parhi, S. K. & Patro, S. K. Compressive strength prediction of PET fiber-reinforced concrete using Dolphin echolocation optimized decision tree-based machine learning algorithms. *Asian J. Civil Eng.* **25** (1), 977–996. <https://doi.org/10.1007/s42107-023-00826-8> (2024).
7. Parhi, S. K. & Patro, S. K. Prediction of compressive strength of geopolymers concrete using a hybrid ensemble of grey Wolf optimized machine learning estimators. *J. Building Eng.* **71**, 106521. <https://doi.org/10.1016/j.jobe.2023.106521> (2023).

8. Parhi, S. K., Dwibedy, S. & Patro, S. K. Managing waste for production of low-carbon concrete mix using uncertainty-aware machine learning model. *Environ. Res.* **279**, 121918. <https://doi.org/10.1016/j.envres.2025.121918> (2025).
9. Giaccio, G., de Sensale, G. R. & Zerbino, R. Failure mechanism of normal and high-strength concrete with rice-husk Ash. *Cem. Concr. Compos.* **29** (7), 566–574. <https://doi.org/10.1016/j.cemconcomp.2007.04.005> (2007).
10. Saraswathy, V. & Song, H. W. Corrosion performance of rice husk ash blended concrete. *Construct. Build. Mater.* **21** (8), 1779–1784 <https://doi.org/10.1016/j.conbuildmat.2006.05.037> (2007).
11. Paris, J. M., Roessler, J. G., Ferraro, C. C., DeFord, H. D. & Townsend, T. G. A review of waste products utilized as supplements to Portland cement in concrete. *J. Clean. Prod.* **121**, 1–18. <https://doi.org/10.1016/j.jclepro.2016.02.013> (2016).
12. Ganesan, K., Rajagopal, K. & Thangavel, K. Rice husk Ash blended cement: assessment of optimal level of replacement for strength and permeability properties of concrete. *Constr. Build. Mater.* **22** (8), 1675–1683. <https://doi.org/10.1016/j.conbuildmat.2007.06.011> (2008).
13. Kishore, R., Bhikshma, V. & Prakash, P. J. Study on strength characteristics of high strength rice husk Ash concrete. *Procedia Eng.* **14**, 2666–2672. <https://doi.org/10.1016/j.proeng.2011.07.335> (2011).
14. Bhanumathidas, N. & Mehta, P. Concrete mixtures made with ternary blended cements containing fly ash and rice-husk ash. In *Seventh CANMET/ACI International Conference on Fly Ash, Silica Fume, Slag and Natural Pozzolans in Concrete*, 379–391. <http://worldcat.org/isbn/0870310267> (2001).
15. Bui, D. D., Hu, J. & Stroeven, P. Particle size effect on the strength of rice husk Ash blended gap-graded Portland cement concrete. *Cem. Concr. Compos.* **27** (3), 357–366. <https://doi.org/10.1016/j.cemconcomp.2004.05.002> (2005).
16. de Sensale, G. R. Strength development of concrete with rice-husk Ash. *Cem. Concr. Compos.* **28** (2), 158–160. <https://doi.org/10.1016/j.cemconcomp.2005.09.005> (2006).
17. Sam, J. Compressive strength of concrete using fly Ash and rice husk ash: A review. *Civil Eng. J.* **6** (7), 1400–1410. <https://doi.org/10.28991/cej-2020-03091556> (2020).
18. Saridemir, M. Genetic programming approach for prediction of compressive strength of concretes containing rice husk Ash. *Constr. Build. Mater.* **24** (10), 1911–1919. <https://doi.org/10.1016/j.conbuildmat.2010.04.011> (2010).
19. Islam, M. N., Zain, M. F. M. & Jamil, M. Prediction of strength and slump of rice husk Ash incorporated high-performance concrete. *J. Civil Eng. Manage.* **18** (3), 310–317. <https://doi.org/10.3846/13923730.2012.698890> (2012).
20. Liu, C., Zhang, W., Liu, H., Lin, X. & Zhang, R. A compressive strength prediction model based on the hydration reaction of cement paste by rice husk Ash. *Constr. Build. Mater.* **340**, 127841. <https://doi.org/10.1016/j.conbuildmat.2022.127841> (2022).
21. Mohamed, H. S. et al. Compressive behavior of elliptical concrete-filled steel tubular short columns using numerical investigation and machine learning techniques. *Sci. Rep.* **14** (1), 27007. <https://doi.org/10.1038/s41598-024-77396-5> (2024).
22. Xu, C. et al. Numerical and machine learning models for concentrically and eccentrically loaded <sc>CFST columns confined with <sc>FRP wraps. *Struct. Concrete.* <https://doi.org/10.1002/suco.202400541> (2024).
23. Tipu, R. K., Batra, V., Suman, K. S., Pandya & Panchal, V. R. Enhancing load capacity prediction of column using eReLU-activated BPNN model. *Structures* **58**, 105600. <https://doi.org/10.1016/j.istruc.2023.105600> (2023).
24. George, C. et al. Predicting the fire-induced structural performance of steel tube columns filled with SFRC-enhanced concrete: using artificial neural networks approach. *Front. Built Environ.* **10** <https://doi.org/10.3389/fbui.2024.1403460> (2024).
25. Satyanarayana, A., Dushyanth, V. B. R., Riyam, K. A., Geetha, L. & Kumar, R. Assessing the seismic sensitivity of Bridge structures by developing fragility curves with ANN and LSTM integration. *Asian J. Civil Eng.* **25** (8), 5865–5888. <https://doi.org/10.1007/s42107-024-01151-4> (2024).
26. Satyanarayana, A. et al. A multifaceted comparative analysis of incremental dynamic and static pushover methods in Bridge structural assessment, integrated with artificial neural network and genetic algorithm approach. *Discover Mater.* **5** (1), 84. <https://doi.org/10.1007/s43939-025-00262-2> (2025).
27. George, C., Kumar, R. & Ramaraju, H. K. Comparison of experimental and analytical studies in light gauge steel sections on CFST using SFRC in beams subjected to high temperatures. *Asian J. Civil Eng.* **26** (2), 667–681. <https://doi.org/10.1007/s42107-024-01213-7> (2025).
28. Wani, S. R. & Suthar, M. A comparative analysis of the predictive performance of Tree-Based and artificial neural network approaches for compressive strength of concrete utilising waste. *Int. J. Pavement Res. Technol.* <https://doi.org/10.1007/s42947-024-00454-8> (2024).
29. Wani, S. R. & Suthar, M. Using machine learning approaches for predicting the compressive strength of ultra-high-performance concrete with SHAP analysis. *Asian J. Civil Eng.* **26** (1), 373–388. <https://doi.org/10.1007/s42107-024-01195-6> (2025).
30. Wani, S. R. & Suthar, M. Using soft computing to forecast the strength of concrete utilized with sustainable natural fiber reinforced polymer composites. *Asian J. Civil Eng.* **25** (8), 5847–5863. <https://doi.org/10.1007/s42107-024-01150-5> (2024).
31. Velay-Lizancos, M., Perez-Ordoñez, J. L., Martinez-Lage, I. & Vazquez-Burgo, P. Analytical and genetic programming model of compressive strength of eco concretes by NDT according to curing temperature. *Constr. Build. Mater.* **144**, 195–206. <https://doi.org/10.1016/j.conbuildmat.2017.03.123> (2017).
32. Cheng, M. Y., Firdausi, P. M. & Prayogo, D. High-performance concrete compressive strength prediction using genetic weighted pyramid operation tree (GWPOT). *Eng. Appl. Artif. Intell.* **29**, 104–113. <https://doi.org/10.1016/j.engappai.2013.11.014> (2014).
33. Topcu, İ. B. & Saridemir, M. Prediction of compressive strength of concrete containing fly Ash using artificial neural networks and fuzzy logic. *Comput. Mater. Sci.* **41** (3), 305–311. <https://doi.org/10.1016/j.commatsci.2007.04.009> (2008).
34. Kumar, R., Kumar, S., Rai, B. & Samui, P. Development of hybrid gradient boosting models for predicting the compressive strength of high-volume fly Ash self-compacting concrete with silica fume. *Structures* **66**, 106850. <https://doi.org/10.1016/j.istruc.2024.106850> (2024).
35. Kumar, R. et al. Estimation of the compressive strength of ultrahigh performance concrete using machine learning models. *Intell. Syst. Appl.* **25**, 200471. <https://doi.org/10.1016/j.iswa.2024.200471> (2025).
36. Sathvik, S. et al. Analyzing the influence of manufactured sand and fly Ash on concrete strength through experimental and machine learning methods. *Sci. Rep.* **15** (1), 4978. <https://doi.org/10.1038/s41598-025-88923-3> (2025).
37. Erdal, H. I., Karakurt, O. & Namli, E. High performance concrete compressive strength forecasting using ensemble models based on discrete wavelet transform. *Eng. Appl. Artif. Intell.* **26** (4), 1246–1254. <https://doi.org/10.1016/j.engappai.2012.10.014> (2013).
38. Behnood, A., Behnood, V., Gharehveran, M. M. & Alyamac, K. E. Prediction of the compressive strength of normal and high-performance concretes using M5P model tree algorithm. *Constr. Build. Mater.* **142**, 199–207. <https://doi.org/10.1016/j.conbuildmat.2017.03.061> (2017).
39. Li, C., Mei, X., Dias, D., Cui, Z. & Zhou, J. Compressive strength prediction of rice husk Ash concrete using a hybrid artificial neural network model. *Materials* **16** (8), 3135. <https://doi.org/10.3390/ma16083135> (2023).
40. Iqtidar, A. et al. Prediction of compressive strength of rice husk Ash concrete through different machine learning processes. *Crystals* **11** (4), 352. <https://doi.org/10.3390/cryst11040352> (2021).
41. Hamidian, P., Alidoust, P., Golafshani, E. M., Niavol, K. P. & Behnood, A. Introduction of a novel evolutionary neural network for evaluating the compressive strength of concretes: A case of rice husk Ash concrete. *J. Building Eng.* **61**, 105293. <https://doi.org/10.1016/j.jobe.2022.105293> (2022).
42. Nasir Amin, M. et al. Prediction model for rice husk Ash concrete using AI approach: boosting and bagging algorithms. *Structures* **50**, 745–757. <https://doi.org/10.1016/j.istruc.2023.02.080> (2023).
43. Alyami, M. et al. Estimating compressive strength of concrete containing rice husk Ash using interpretable machine learning-based models. *Case Stud. Constr. Mater.* **20**, e02901. <https://doi.org/10.1016/j.cscm.2024.e02901> (2024).

44. Rasmussen, C. E. Gaussian processes in machine learning, pp. 63–71. [https://doi.org/10.1007/978-3-540-28650-9\\_4](https://doi.org/10.1007/978-3-540-28650-9_4) (2004).
45. Cortes, C. & Vapnik, V. Support-vector networks. *Mach. Learn.* **20** (3), 273–297. <https://doi.org/10.1007/BF00994018> (1995).
46. Prasad, D. V. V. & Jaganathan, S. Null-space based facial classifier using linear regression and discriminant analysis method. *Cluster Comput.* **22**, 9397–9406. <https://doi.org/10.1007/s10586-018-2178-z> (2019).
47. Quinlan, J. R. Induction of decision trees. *Mach. Learn.* **1** (1), 81–106. <https://doi.org/10.1007/BF00116251> (1986).
48. Gad, A. F. Artificial neural networks, in Practical Computer Vision Applications Using Deep Learning with CNNs, Berkeley, CA: A, 45–106. doi: [https://doi.org/10.1007/978-1-4842-4167-7\\_2](https://doi.org/10.1007/978-1-4842-4167-7_2). (2018).
49. Chen, T. & Guestrin, C. XGBoost. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 785–794. <https://doi.org/10.1145/2939672.2939785> (2016).
50. Jin, R. & Agrawal, G. Communication and memory efficient parallel decision tree construction. In *Proceedings of the 2003 SIAM International Conference on Data Mining* 119–129 <https://doi.org/10.1137/1.9781611972733.11> (2003).
51. Gayathri, R., Rani, S. U., Čepová, L., Rajesh, M. & Kalita, K. A comparative analysis of machine learning models in prediction of mortar compressive strength. *Processes* **10** (7), 1387. <https://doi.org/10.3390/pr10071387> (2022).
52. Ho, T. K. The random subspace method for constructing decision forests. *IEEE Trans. Pattern Anal. Mach. Intell.* **20**, 832–844. <https://doi.org/10.1109/34.709601> (1998).
53. Murtagh, F. Multilayer perceptrons for classification and regression. *Neurocomputing* **2**, 5–6. [https://doi.org/10.1016/0925-2312\(91\)90023-5](https://doi.org/10.1016/0925-2312(91)90023-5) (1991).
54. Phylo, P. P., Byun, Y. C. & Park, N. Short-term energy forecasting using machine-learning-mased ensemble voting regression. *Symmetry* **14**, 1, 160 <https://doi.org/10.3390/sym14010160> (2022).
55. Bomrah, S. et al. A scoping review of machine learning for sepsis prediction- feature engineering strategies and model performance: a step towards explainability. *Crit. Care.* **28** (1), 180. <https://doi.org/10.1186/s13054-024-04948-6> (2024).
56. Vrieze, S. I. Model selection and psychological theory: A discussion of the differences between the Akaike information criterion (AIC) and the bayesian information criterion (BIC). *Psychol. Methods.* **17** (2), 228–243. <https://doi.org/10.1037/a0027127> (2012).
57. Yu, Z. G. & Jiang, P. Distance, correlation and mutual information among portraits of organisms based on complete genomes. *Phys. Lett. A.* **286** (1), 34–46. [https://doi.org/10.1016/S0375-9601\(01\)00336-X](https://doi.org/10.1016/S0375-9601(01)00336-X) (2001).
58. Hall, M. Correlation distance and bounds for mutual information. *Entropy* **15** (9), 3698–3713. <https://doi.org/10.3390/e15093698> (2013).
59. García, S., Ramírez-Gallego, S., Luengo, J., Benítez, J. M. & Herrera, F. Big data preprocessing: methods and prospects. *Big Data Analytics.* **1** (1), 9. <https://doi.org/10.1186/s41044-016-0014-0> (2016).
60. Luengo, J., García-Gil, D., Ramírez-Gallego, S., García, S. & Herrera, F. *Big Data Preprocessing* <https://doi.org/10.1007/978-3-030-39105-8> (Springer International Publishing, 2020).
61. Chao-Lung, H., Le Anh-Tuan, B. & Chun-Tsun, C. Effect of rice husk Ash on the strength and durability characteristics of concrete. *Constr. Build. Mater.* **25** (9), 3768–3772. <https://doi.org/10.1016/j.conbuildmat.2011.04.009> (2011).

## Acknowledgements

The authors would like to thank Prince Sultan University for their support. The authors would also like to acknowledge The Office of Research and Sponsored Programs, Abu Dhabi University, Abu Dhabi (U.A.E) for offering Research, Innovation, and Impact Grant (Cost Center # 19300933). The authors extend their appreciation to the Deanship of Scientific Research at Northern Border University, Arar, KSA for funding this research work through the project number “NBU-FFR-2025-2105-12”.

## Author contributions

Ala'a R. Al-Shamasneh, Arsalan Mahmoodzadeh, and Manish Kewalramani contributed to the conceptualization, data curation, methodology, original draft preparation, and investigation. Abdulaziz Alghamdi and Jasim Alnahas were responsible for investigation, visualization, and review and editing of the manuscript. Mohammed Sulaiman, Nejib Ghazouani, and Ibrahim Albaijan contributed to visualization as well as review and editing.

## Declarations

### Competing interests

The authors declare no competing interests.

### Additional information

**Correspondence** and requests for materials should be addressed to A.M.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025