Materiales de Construcción 75 (358)
April-June 2025, e375
ISSN-L: 0465-2746, eISSN: 1988-3226
https://doi.org/10.3989/mc.2025.382624

Predictive modelling on the compressive strength and fresh properties of self-compacting recycled aggregate concrete using machine learning approaches

T. Singh, K. Kapoor ✉, S.P. Singh

a. Department of Civil Engineering, Dr B R Ambedkar National Institute of Technology, (Jalandhar, India)
✉: kapoork@nitj.ac.in

Received 16 April 2024
Accepted 11 October 2024
Available on line 16 October 2025

ABSTRACT: Self-Compacting Recycled Aggregate Concrete (SCRAC) has emerged as a practical engineering solution to recycle the construction aggregates. This study focusses on developing and comparing six popular Machine Learning (ML) models based on Multi Layer Perceptron regression (MLP), Gradient Boost Regression (GBR), Bagging regression (BG), Extreme Gradient Boost regression (XGB), K Nearest Neighbours regression (KNN), and Support Vector Regression (SVR), along with six metamodels prepared by stacking of these ML models, for prediction of each, Compressive Strength (CS), Slump Flow (SF) and V Funnel time (VF) of the SCRAC. Eight mix ratios were used as the input features. The best performance ranking was exhibited by Gradient Boost model GBR_CS for CS prediction whereas stacked models performed best for prediction of SF and VF. Additionally, multi-objective optimization and partial Pareto fronts were constructed to analyse trade-offs between CS, SF, cost, and CO_2 emission.

KEY WORDS: Self compacting recycled aggregate concrete; Multi layer perceptron; Gradient boost regression; Bagging regression; Extreme gradient boost regression; K nearest neighbours regression; Support vector regression; K fold cross validation; Multi objective pareto optimisation.

Citation/Citar como: Singh T, Kapoor K, Singh SP. 2025. Predictive modelling on the compressive strength and fresh properties of self-compacting recycled aggregate concrete using machine learning approaches. Mater. Construcc. 75(358):e375. https://doi.org/10.3989/mc.2025.382624

RESUMEN: *Modelado predictivo de la resistencia a la compresión y las propiedades en fresco del hormigón autocompactante con áridos reciclados utilizando enfoques de aprendizaje automático*. El Hormigón Autocompactante con Áridos Reciclados (SCRAC) es una solución de ingeniería para reciclar áridos de la construcción. Este estudio desarrolla y compara seis modelos de aprendizaje automático (ML), incluidos perceptrón multicapa (MLP), regresión de aumento de gradiente (GBR), regresión de ensacado (BG), regresión de aumento de gradiente extremo (XGB), K vecinos más cercanos (KNN) y regresión de vectores de soporte (SVR), junto con seis metamodelos basados en el apilamiento de estos modelos, para predecir la resistencia a la compresión (CS), flujo de asentamiento (SF) y tiempo de embudo V (VF) del SCRAC. Se utilizaron ocho proporciones de mezcla como características de entrada. El mejor rendimiento lo obtuvo GBR_CS para la predicción de CS, mientras que los modelos apilados fueron superiores en SF y VF. Además, se realizaron frentes de Pareto parciales y una optimización multiobjetivo para analizar las compensaciones entre CS, SF, costo y emisiones de CO_2.

PALABRAS CLAVE: Hormigón autocompactante con áridos reciclados; Perceptrón multicapa; Regresión por gradiente ascendente; Regresión por ensacado; Regresión por gradiente ascendente extremo; Regresión por los K vecinos más cercanos; Regresión de vectores de soporte; Validación cruzada K-fold; Optimización multiobjetivo de Pareto.

Graphical abstract:

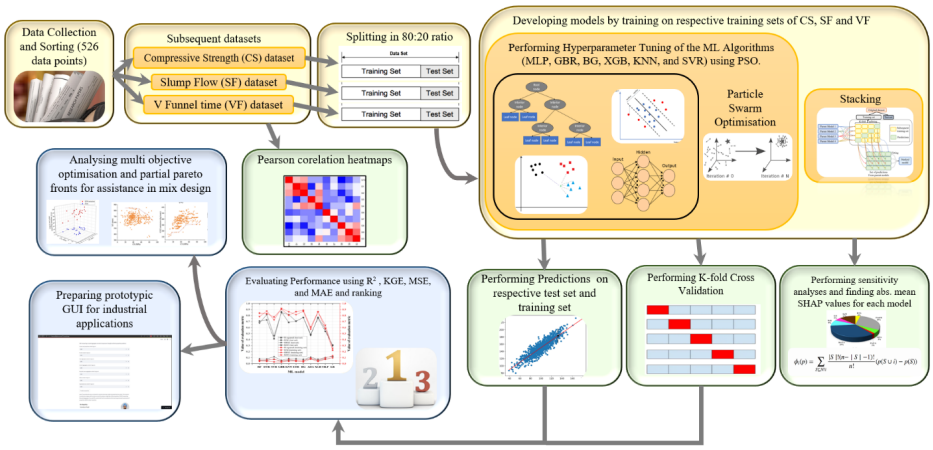


Table 1. Acronyms.

Acronym	Expanded form	Acronym	Expanded form
CS	Compressive Strength	B/TA	Binder/Total Aggregates
SCC	Self Compacting Concrete	FAsh/B	Fly Ash /Binder
SCRAC	Self Compacting Recycled Aggregate Concrete	RA/CA	Recycled Aggregates/natural Coarse Aggregates
C&D	Construction and Demolition	W/B	Water/Binder
RA	Recycled Aggregates	L/S	Liquid/Solids
NA	Natural Aggregates	Sp/B	Superplasticizer/Binder
ITZ	Interstitial Transition Zone	PSO	Particle Swarm Optimisation
SF	Slump Flow	GUI	Graphical User Interface
VF	V Funnel time	GBR	Gradient Boost Regression
Sp	Superplasticizer	XGB	Extreme Gradient Boost
STS	Split Tensile Strength	SVR	Support Vector Regression
EMV	Expanded Mortar Volume	BG	Bagging
ML	Machine Learning	DNN	Deep Neural Network
KNN	K Nearest Neighbours	GEP	Gene Expression Programme
DT	Decision Tree	ADA	Adaboost regression
RF	Random Forest	LR	Linear Regression

1. INTRODUCTION

With the rising urban population, the burden of accommodating more people on limited land has given rise to the vertical growth of cities worldwide, may it be in the form of multi-storeyed residential and commercial buildings or multi-level roads and railways (1, 2). To ensure the safety, stability and durability of such structures, the design of their structural elements mostly end up accommodating heavy and closely spaced reinforcement. Since the conventional concrete would require heavy vibratory compaction to completely settle and cover the reinforcement in such cases, Self-Compacting Concretes (SCC) have emerged as suitable alternative due to its ability to flow and compact under the effect of gravity (3). However, another problem associated with concrete sector is the exploitation of natural stones for use as aggregates in concrete. Parallelly, the Construction and Demolition (C&D) waste generated per annum in USA, Europe, China and India has surpassed 569 Million tonnes (Mt), 2500 Mt, 923.9 Mt and 15Mt respectively (4-7). The environmental as well as land resource issues related with the disposal of C&D wastes in landfills are well known (5). Unequivocally, the use of C&D wastes in the form of crushed Recycled coarse Aggregates (RA) as a substitute to Natural coarse Aggregates (NA) in developing SCC has moved the construction sector towards a more sustainable engineering solution (8). Abundant research has been reported on the effect of partial to full substitution of RA with NA on the fresh properties, mechanical properties, and durability properties of SCC. Although the RAs consist of a layer of adhered mortar over the natural aggregates, which provide an extra Interfacial Transition Zone (ITZ) between the aggregates and the matrix having a tendency to reduce the Compressive Strength (CS) of the Self Compacting Recycled Aggregate Concrete (SCRAC), Abed. et. al. (9) observed that the RA can be replaced with NA up to 50% without significant reduction in CS. Additionally it was reported that the SCRAC retained its Slump Flow (SF) class and V Funnel (VF) class with same amount of Superplasticizer (Sp) admixture. However, the study by Gracia et.al (10) reveals that the increase in substitution of the RA with NA from 20% to 50% and 100% in 1:2.26:1.75, cement: sand: coarse

aggregate SCRAC having w/c of 0.47, tends to reduce self-compacting properties of SCRAC, hence requiring extra Sp for the concrete to retain the SF and VF class. Since the Fly Ash (FAsh) is a promoter of workability owing to its rounded particles and consequent ball bearing effect in the matrix, it is considered as another important supplementary cementitious constituent in SCC. The effect of reducing the content of FAsh from 24.9% to 8.47% and 0% with respect to cement in 1:1.95:2.077, cement: sand: aggregate SCRAC with w/c=0.4 was observed by Sun et. al. (11). The effect was reported as detrimental on the fresh state properties like SF, T500, VF and L box time. In another recent investigation on the effect of varying substitution levels of NA with RA and the content of FA in SCRAC (11), it was observed that up to 30% replacement of cement with FA, significantly improves the workability of 1:1.56:1.878, binder: sand: aggregate SCRAC with w/c of 0.35. Additionally, it was observed that the increase in RA content improves the CS of SCRAC up to a 50% substitution, owing to the unreacted cement content present in the RA from its previous cycle of casting.

Few attempts to link the effects of different constituents and mix ratios to the compressive strength and fresh properties of SCRAC for developing suitable mix design procedures have been reported in past research. A recent study by Wang and Wu (12) fitted the 28 day CS of SCRAC (f_{cu}) to the 28 day CS of conventional concrete (f_{ce}), RA/NA replacement ratio (η_{RA}), Fash/cement replacement ratio (β), and w/c ratio (λ) as per Equation [1].

$$f_{cu} = 0.55 f_{ce}(1-0.18\eta_{RA})(1-0.82\beta)\lambda^{-0.86} \quad [1]$$

It can be observed that the f_{cu} is directly related to the η_{RA} and β but inversely related to λ, which is in conformity with basic experimental investigations reported in substantial previous research. However, the previous studies on SCRAC properties, simply state the assumption of 600mm to 800mm of Slump flow in the mix design and makes no comment about foreseeing other fresh properties (9,11). Another attempt by Chen et.al. (13) aimed at statistically modelling the rheological properties of the SCRAC at fixed FAsh content yielded only a moderate accuracy of 59.7% until another parameter of mortar volume was considered in the calculations. The reliability of these statistical mix design approaches stands bleak, owing to their constrained nature of applicability subject to the set of assumptions made in respective studies.

Nonetheless, apart from the conventional approaches, predictive modelling of concrete properties using Machine Learning (ML) approaches has been a successful attempt in vast domains of concrete research (14). Mahmood et.al. (15) developed few such ML models to predict the CS of SCC by training and testing on small datasets of 106 and 27 respectively. They reported best R^2 values close to 0.99 for K Nearest Neighbours (KNN), Decision Tree (DT), Random Forest (RF), Gradient Boost (GBR) and Extreme Gradient Boost (XGB) based models. However, the datasets were small and employing K fold cross validation on these models could have given a more reliable idea about their predictive efficacy. Similarly, another study by (14) moved a step further by predicting CS as well as the SF of SCC by training Support Vector Regression (SVR) and RF. Datasets of 255 and 349 were taken for prediction of CS and SF respectively, with a test train split of 90:10. Despite the larger dataset of SF, lower R^2 of 0.90 to 0.95 were observed in case of SF predictions as compared to CS predictions on the test set. This is attributed to the more volatile and random nature of the fresh properties of SCC rendering them slightly harder to predict than strength properties such as CS and STS (14). On similar lines as SCC, the CS prediction of RA concrete has also been successfully studied over last decade. As per a recent study (16) prediction of the CS of RA concrete after training the Deep Neural Networks (DNN) and Genetic Programming (GP) models, yielded R^2 of 0.94 and 0.83 respectively. Since it is evident the ML can be effectively applied for modelling the properties of SCC and RA concrete, it has moved significant research focus in recent years, towards SCRAC as well. The ML modelling for the CS prediction of SCRAC, documented by (17-22) employed SVR, ANN, ANN, RF, XGB, Catboost and XGB algorithms, respectively. These studies used datasets of 111, 210, 515, 515, 337, 515 and 368, respectively yielding prediction R^2 values of 0.89, 0.90, 0.93, 0.71, 0.87, 0.96 and 0.97, respectively.

1.1. Research significance

The fresh properties of any SCC (SCRAC in context of this study) hold prime importance determining its use case. However, most of the reported research on ML based predictive modelling primarily focusses on the CS of SCRAC. Sparse attempts are available on the ML based prediction of the fresh properties of SCRAC. This study is aimed at addressing this gap by performing predictions of the CS as well as the SF and VF time of SCRAC. Further, since the range of values of the mix constituents vary tremendously, mix ratios such as Binder/Total Aggregates (B/TA), FAsh/Binder (FAsh/B), Binder/Fine Aggregates (B/FA), Fine Aggregates/Coarse Aggregates (FA/CA), Recycled Aggregates/total Coarse Aggregates (RA/CA), Water/Binder (W/B), Liquid/Solids (L/S) and Superplasticizer/Binder (Sp/B) were considered as the eight input features instead of quantities of mix constituents. The RA/CA ratio was primarily introduced as the input feature to gauge the effect of RA on SCRAC properties. The modelling was based on six popular ML model algorithms: Multi-Layer Perceptron (MLP), GBR, Bagging (BG), XGB, KNN, SVR hypertuned using the Particle Swarm Optimiser (PSO) and six metamodels developed by stacking of the former models. The robustness of the ML models was assessed by K-fold cross validation. Additionally, absolute mean SHAP analysis and sensitivity analysis were performed to understand the effect of

every input feature on the target feature. In addition to this, pareto optimisation plots were prepared to get an insight into the behavioural patterns of SCRAC in terms of CS, SF, cost, and carbon footprint. Lastly, a Graphical User Interface (GUI) for prediction of SCRAC properties, was prepared as a preliminary prototype of the industrial application of the research.

2. METHODOLOGY

2.1. Preparation of datasets

The literature on SCRAC was explored to collect the data on the mix constituents and target features such as CS, SF and VF summing to a total dataset of 526 mixes from 52 research publications. Subsequently, the mix ratios of B/TA, FAsh/B, B/FA, FA/CA, RA/CA, W/B and L/S were computed to act as input features in the dataset for ML modelling. The B/TA and B/FA shall represent the role of the total binder relative to the total aggregate and FA, whereas FAsh/B represents the role of rounded particles of FAsh on target features. Similarly, the FA/CA represents the effect of fine aggregate paste whereas W/B, L/S and Sp/B represent the role of liquids on the target features. However, not all the referred research papers reported all the target features, therefore three separate subsets of the dataset were prepared for prediction of CS, SF and VF as per the availability of results of these target features. The datasets thus obtained for CS, SF and VF consisted of 431, 507 and 317 samples, respectively. Table 2 shows the descriptive statistics of the datasets for the prediction of CS, SF and VF. The skewness of all the variables except FA/CA, were small and comparable thus confirming the reliability of the dataset. The uneven distribution of FA/CA ratio is manifestable to the variation in mix designs and research approaches of different authors referred for gathering the dataset.

Table 2. Descriptive statistics of the datasets for prediction of CS, SF and VF time respectively.

Dataset for CS prediction	***B/TA***	***FAsh/B***	***B/FA***	***FA/CA***	***RA/CA***	***W/B***	***L/S***	***Sp/B***	***CS***
Minimum	0.178	0.000	0.259	0.447	0.000	0.120	0.038	0.000	7.170
Maximum	0.487	0.750	1.020	4.007	1.000	0.750	0.138	0.029	88.000
Mean	0.323	0.161	0.629	1.162	0.481	0.389	0.095	0.008	43.544
Std. Dev. (SD)	0.065	0.171	0.135	0.477	0.392	0.069	0.018	0.005	15.757
Skewness	0.514	0.754	0.148	2.820	0.194	0.006	0.405	0.695	0.749
Dataset for SF prediction	***B/TA***	***FAsh/B***	***B/FA***	***FA/CA***	***RA/CA***	***W/B***	***L/S***	***Sp/B***	***SF***
Minimum	0.178	0.000	0.259	0.470	0.000	0.039	0.011	0.000	280.00
Maximum	0.487	0.750	1.020	2.802	1.000	0.750	0.138	0.029	850.00
Mean	0.328	0.157	0.634	1.160	0.502	0.375	0.093	0.008	666.01
SD	0.062	0.169	0.136	0.384	0.405	0.070	0.018	0.005	94.711
Skewness	0.428	0.886	0.039	1.760	0.092	0.081	0.216	0.666	-1.889
Dataset for VF prediction	***B/TA***	***FAsh/B***	***B/FA***	***FA/CA***	***RA/CA***	***W/B***	***L/S***	***Sp/B***	***VF***
Minimum	0.241	0.000	0.417	0.470	0.000	0.039	0.011	0.000	3.000
Maximum	0.477	0.600	0.958	2.558	1.000	0.505	0.138	0.024	30.50
Mean	0.343	0.156	0.640	1.218	0.433	0.378	0.097	0.007	10.118
SD	0.058	0.150	0.108	0.331	0.378	0.056	0.018	0.005	7.865
Skewness	0.543	0.413	0.261	1.289	0.385	-0.69	0.643	0.605	6.162

2.2. Data visualisation

The combined scatter plots representing the datasets of CS, SF and VF have been depicted in Figure 1(a), Figure 1(b) and Figure 1(c), respectively. The scatter plots show the distribution of every feature with respect to every other feature in the whole dataset. It can be observed that the distribution of the data is not excessively patchy except for the FA/CA. Further, the corelation heatmaps of the three datasets have been presented in Figure 2.

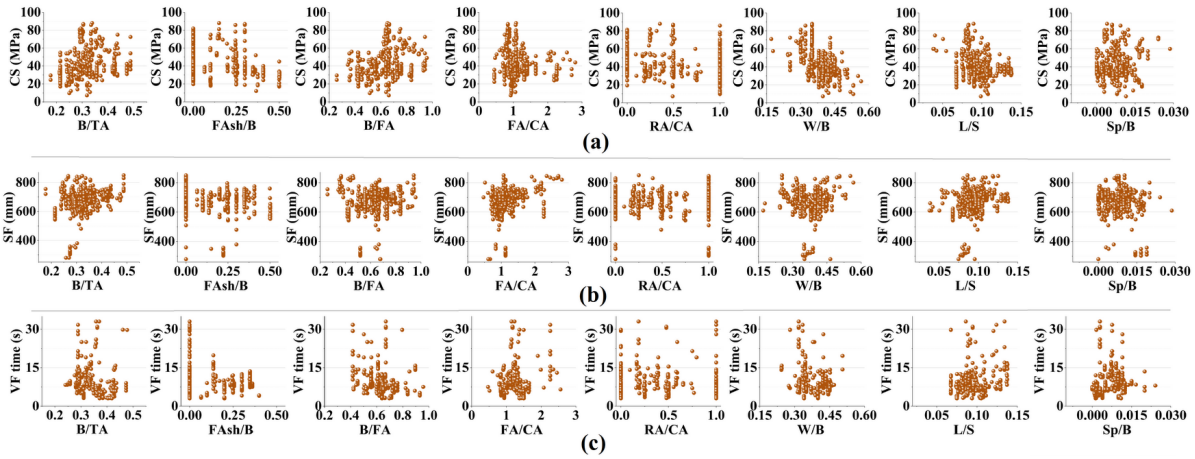


Figure 1. Scatter plots of (a) CS dataset, (b) SF dataset and (c) VF dataset.

Correlation matrix heatmap of CS dataset

	B/TA	FAsh/B	B/FA	FA/CA	RA/CA	W/B	L/S	Sp/B	CS
B/TA	1	0.15	0.76	0.10	-0.02	-0.36	0.48	-0.07	0.16
FAsh/B	0.15	1	0.16	-0.19	0.11	0.20	0.36	-0.14	-0.32
B/FA	0.76	0.16	1	-0.47	-0.09	-0.43	0.20	-0.09	0.26
FA/CA	0.10	-0.19	-0.47	1	0.13	0.17	0.23	0.02	-0.14
RA/CA	-0.02	0.11	-0.09	0.13	1	0.27	0.24	0.17	-0.23
W/B	-0.36	0.20	-0.43	0.17	0.27	1	0.63	-0.19	-0.52
L/S	0.48	0.36	0.20	0.23	0.24	0.63	1	-0.20	-0.34
Sp/B	-0.07	-0.14	-0.09	0.02	0.17	-0.19	-0.20	1	0.10
CS	0.16	-0.32	0.26	-0.14	-0.23	-0.52	-0.52	0.10	1

Correlation matrix heatmap of SF dataset

	B/TA	FAsh/B	B/FA	FA/CA	RA/CA	W/B	L/S	Sp/B	SF
B/TA	1	0.04	0.80	0.03	-0.03	-0.28	0.48	-0.24	0.27
FAsh/B	0.04	1	0.12	-0.25	0.12	0.23	0.28	-0.12	-0.10
B/FA	0.80	0.12	1	-0.51	-0.09	-0.34	0.29	-0.19	0.11
FA/CA	0.03	-0.25	-0.05	1	0.12	0.18	0.15	-0.01	0.28
RA/CA	-0.03	0.12	-0.09	0.12	1	0.33	0.27	0.15	-0.08
W/B	-0.28	0.23	-0.34	0.18	0.33	1	0.68	-0.11	0.08
L/S	0.48	0.28	0.29	0.15	0.27	0.68	1	-0.23	0.23
Sp/B	-0.24	-0.12	-0.19	-0.01	0.15	-0.11	-0.23	1	-0.15
SF	0.27	-0.10	0.11	0.28	-0.08	0.08	0.23	-0.15	1

Correlation matrix heatmap of VF time dataset

	B/TA	FAsh/B	B/FA	FA/CA	RA/CA	W/B	L/S	Sp/B	VF
B/TA	1	0.10	0.77	0.21	0.11	-0.13	0.57	-0.35	-0.11
FAsh/B	0.10	1	0.09	-0.13	0.11	0.36	0.36	-0.06	-0.22
B/FA	0.77	0.09	1	-0.40	0.02	-0.23	0.34	-0.28	-0.17
FA/CA	0.21	-0.13	-0.40	1	0.15	0.10	0.23	-0.08	0.15
RA/CA	0.11	0.11	0.02	0.15	1	0.26	0.29	0.07	0.03
W/B	-0.13	0.36	-0.23	0.10	0.26	1	0.73	0.02	-0.17
L/S	0.57	0.36	0.34	0.23	0.29	0.73	1	-0.17	-0.21
Sp/B	-0.35	-0.06	-0.28	-0.08	0.07	0.02	-0.17	1	0.03
VF	-0.11	-0.22	-0.17	0.15	0.03	-0.17	-0.21	0.03	1

Figure 2. Correlation matrix heatmaps of the CS, SF and VF datasets respectively.

Each of the datasets were split into training and test sets in 80:20 ratio. The splitting was done using the train_test_ split command while keeping the random_state=0 in the python script.

2.3. Machine learning methods

MLP, GBR, BG, XGB, KNN and SVR algorithms were used to develop respective models for prediction of CS, SF and VF time. A brief summary of these algorithms is presented in Table 3.

Table 3. Description of the ML algorithms.

ML algorithm	Inspiration/ fundamentals	Governing equation(s)	References
MLP	Inspired from the working of neurons in the human brain	$O = \frac{1}{1+e^{-OS}}$ [2] $OS = bias + \sum_{i=1}^{n} I_i.w_i$ [3] Where, O is the output of the perceptron, I_i= ith input value w_i= weight for the I_i	(23, 24)
GBR	Summing up of small values of residuals from a number of decision trees, to yield a final output.	$y = y' + \sum_{d} \eta.p$ [4] Where, y is the predicted target variable y' = expected value of target variable i.e., the mean, η=learning rate, d=the number of successive decision trees prepared to predict the pseudo residuals, p=the predicted pseudo residual of the respective decision tree.	(25)
BG	Bagging stands for *bootstrapping* + *aggregation*. Random sampling from dataset to form decision trees is called *bootstrapping* and the clubbing the results from decision trees is called *aggregation*.	Same as GBR $y = y' + \sum_{d} \eta.p$ [4] But replacement of samples is allowed in bootstrapping i.e., a single sample can be a part of the training of multiple decision trees.	(26)
XGB	Improvement over GBR where decision trees are fitted based on maximisation of information gain.	$S = \frac{(residuals)^2}{N_s - \lambda}$ [5] Where, S is the similarity index of the leaf or root of the tree, N_s= number of samples falling the leaf or root, λ=Regularisation parameter (an hyperparameter of the XGB algorithm)	(27)
KNN	Averaging of the target values of "n" number of nearest neighbours.	$O = \frac{\sum_{i=0}^{n} O_i}{n}$ [6] Where, O is the calculated value of output variable, O_i is the ith value of the output variable among the "n" nearest neighbours.	(28)
SVR	Fitting of a hyperplane with a widest offset ε on either side, such that minimum number of samples fall within that offset. The margins of the offset are called the support vectors.	$\{y - [w.\phi(x) + \beta] - \xi_i\} \leq \varepsilon$ [7] $\{-y + [w.\phi(x) + \beta] - \xi_i\} \leq \varepsilon$ [8] Where, y is the actual value of the target variable, w, $\phi(x)$ and β are the slope, defining function and intercept of the hyperplane, ξ_i is the allowable points within the margin and ε is the offset between the hyperplane and support vectors.	(29)

2.4. Particle swarm optimisation

Inspired from natural processes such as a swarm of bees or fishes searching for food, the conceptual idea of PSO is that every particle shall move towards three directions representing inertial, cognitive and social aspect (by equal distance) in each iteration (30). Inertial motion follows the last direction of motion of the particle in previous iteration. Cognitive motion points towards personal best position of the particle throughout all the previous iterations, called as the

Pbest position. And the social motion aligns towards the best position achieved by any particle in any iteration is known as *Gbest* position (30-32). The inertial, cognitive and social components of the swarm are represented in the form of *w,* c_1, and c_2 as shown in Equation [9] and Equation [10]. A pictorial representation of PSO have been presented in Figure 3.

$$\overrightarrow{X_i^{t+1}} = \overrightarrow{X_i^t} + \overrightarrow{V_i^{i+1}} \qquad [9]$$

$$V_i^{t+1} = w.V_i^t + c_1 r_1 (P_i^t - X_i^t) + c_2 r_2 (G^t - X_i^t) \qquad [10]$$

Where, X_i^t is position of ith particle in previous tth iteration, V_i^t is velocity of ith particle in tth iteration, P_i^t and G^t are *Pbest* and *Gbest* of the swarm in t iterations, r_1 and r_2 are random numbers.

2.5. K fold cross validation

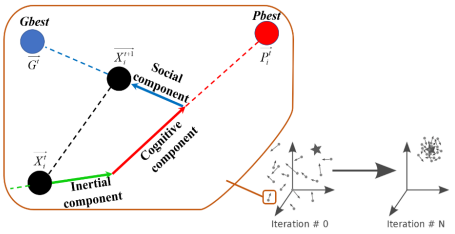


Figure 3. Typical representation of the particle swarm optimisation process (30).

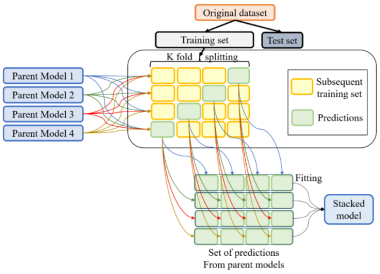


Figure 4. Schematic representation of the stacking process of models (35).

This technique involves splitting the whole dataset in "K" number of equally sized subsets called as folds. Subsequently K number of training set-test set pairs are created by assigning one of the subsets as test set and remaining all subsets collectively as the test set in each pair (33, 34). Subsequently, the model is trained and tested on each of the training set-test set pairs and the mean value of the evaluation metric from all pair is recorded as the K fold cross validated score. A model which shows a good K fold cross validated scores signifies a robust performance. The K fold cross validation with 5 folds was done for each model in this research.

2.6. Stacking

A pictorial representation of the stacking algorithm is presented in Figure 4. Stacking consists of splitting the training set into subsequent K fold training and test sets. Each of the parent models is then trained on the subsequent training set and employed to make predictions for the respective subsequent test set. Same process is repeated for all folds (35). Now we get a set of predictions made by all models for all the samples of the training set. This set of predictions is considered as the input dataset for fitting another linear regression or other model to predict the target variable. This whole process is collectively referred to as stacking (35). Six type of stacking models were developed for predictions of each property i.e., CS, SF, and VF time. The models were named as Stack1_CS to Stack6_CS, Stack1_SF to Stack6_SF and Stack1_VF to Stack6_VF, respectively.

2.7. Absolute mean SHAP values

The SHapley Additive exPlainations (SHAP) values were introduced in mid of the 20th century to represent the role of a player in a coalition (36). In context of the ML, the main idea of the SHAP analysis is to assess prediction performance using a coalition of input features with or without a particular input feature considered. Some input features can contribute in better prediction only in coalition with another input feature (36). For instance, in context of this research, for SF prediction, the FAsh/B has a meaning only if the binder content represented by the B/TA is known as the input. Contrarily, the effect of individual removal of an input feature from dataset, on the prediction is known as marginal contribution. SHAP analysis considers both, the coalition effect and the marginal contribution effect to assess the role of an input feature on the target feature prediction (36). The expression for finding the SHAP value of an input feature ϕ for a sample i is presented in Equation [11].

$$\phi_i(f,x) = \sum_{z \subseteq x} \frac{|z|!.(M-|z|-1)!}{M!} [f_x(z) - f_x(z/i)] \qquad [11]$$

Mean of the ϕ_i of all the samples of the dataset were determined and reported as significant absolute mean SHAP values in this research.

2.8. Evaluation metrics

The prediction performance of models can be evaluated by assessing the closeness of the prediction values to the actual values of the target variable or minimisation of the residuals or errors. The performance of all the developed models were assessed by R^2, Kling Gupta Efficiency (KGE), Mean Absolute Error (MAE) and Mean Squared Error (MSE) as the evaluation metrics. The details of these metrics have been presented in Table 4.

Table 4. Description of the evaluation metrics.

Evaluation metric	Equation	Significance
R^2	$R^2 = \left[\frac{n\sum_{i=1}^{n} y_a y_p - \sum_{i=1}^{n} y_a \sum_{i=1}^{n} y_p}{\sqrt{\left[n\sum_{i=1}^{n} y_a^2 - \left(\sum_{i=1}^{n} y_a^2\right)\right]\left[n\sum_{i=1}^{n} y_p^2 - \left(\sum_{i=1}^{n} y_p^2\right)\right]}}\right]^2$ Where, y_a=actual value of target variable, and y_p=predicted value of target variable	Denotes degree of corelation between predicted values and actual values of variable (37)
KGE	$KGE = 1 - \sqrt{(R-1)^2 + (\alpha-1)^2 + (\beta-1)^2}$ Where, R = pearson corelation coefficient α= (mean of y_p)/(mean of y_a) β= (variance of y_p)/(variance of y_a)	Denotes efficiency of the prediction in terms of R^2 as well as relative mean of predicted values (α) and relative variance of predicted values (β) (38)
MAE	$MAE = \frac{\sum_{i=1}^{n} \lvert y_p - y_a \rvert}{n}$ Where, n= number of data points	Denotes magnitude of error or bias of predicted values with respect to actual values of target variable (37)
MSE	$MSE = \frac{\sum_{i=1}^{n} (y_p - y_{p\_mean})^2}{n}$ Where y_{p_mean} is the mean of predicted values	Denotes scatter of the predicted values around the mean of the predicted values itself (37)

2.9. Ranking of the models

Different models perform differently depending on the dataset and evaluation metric used. For instance, some models may excel in predicting on the test set, while others might perform better when assessed through K-fold cross-validation. To determine the overall performance ranking of the models, a scoring system was implemented that considers the influence of R^2, KGE, MAE, and MSE across training set predictions, test set predictions, and K-fold cross-validation predictions.

To calculate the total score for each model, individual scores were assigned for each evaluation metric on a scale from 1 to 12. Higher R^2 and KGE values, as well as lower MAE and MSE values, were considered superior. For example, a model with the highest R^2 on the test set would receive a score of 12, while the model with the lowest R^2 would receive a score of 1. This scoring method was applied to all evaluation metrics across the test set, training set, and K-fold cross-validation. The scores for each model were then summed to obtain a total score, which was used to rank the models. The model with the highest total score was ranked 1st, while the model with the lowest total score was ranked 12th.

2.10. Sensitivity analysis

The sensitivity analysis performed in few previous research based on ML modelling was referred (20, 39). Their approach has certain limitations. The range of all input features may vary differently over a dataset, for example, the ranges of Sp/B and L/B are as small as 0.029 and 0.100 respectively whereas the ranges of RA/CA and FA/CA are as high as 1.00 and 3.56 respectively in the CS dataset as shown in table 2. Therefore, the approach was modified by considering normalised N_i with respect to the mean value of the input feature as shown in Equation [12] and Equation [13]. Further, the mean values of $f_{max}(x_i)$ and $f_{min}(x_i)$ were calculated for "p" number of maximum and minimum target feature valued mixes. These were termed as $f_{p\_max}(x_i)$ and $f_{p\_min}(x_i)$. Thus, the modified normalised N_i shall fall in comparable ranges for all the input features.

$$N_{nor\_i} = \frac{f_{p\_\max}(x_i) - f_{p\_\min}(x_i)}{x_{mean\_i}} \quad [12]$$

$$S_{nor\_i} = \frac{N_{nor\_i}}{\sum_{j=1}^{n} N_{nor\_j}} \quad [13]$$

Where x_{mean_i} is the mean value of the i^{th} input feature, and the normalised impact of every feature S_{nor_i} was calculated for predictions of CS, SF and VF and were expressed as percentage.

2.11. Multi objective pareto optimisation

Multi objective pareto optimisation is a technique used to determine the sequence of optimal solutions which perform good as compared to any other solution, at least in one objective (40). Such solutions are collectively known as a pareto front. In context of this research, the pareto multiple objectives considered for optimising the mix design of SCRAC were maximisation of CS, maximisation of SF, minimisation of Carbon footprint (CO_2 emitted in kg/kg) and minimisation of the cost. The samples in the original dataset of 526 mixes were filtered to the samples reporting both, CS and SF. The CO_2 emission and cost (USD) were computed as per the respective details of constituent materials mentioned in Table 5 (41).

Table 5. Carbon footprint and cost of SCRAC mix constituents (41).

Mix constituent	CO_2 emission in kg/kg	Cost in USD/kg
Cement	0.804	0.130
Fly ash	0.020	0.072
Natural sand	0.017	0.033
Natural aggregates	0.011	0.039
Recycled aggregates	0.004	0.013
Water	0.0003	0.003
Superplasticizer	2.388	2.470

2.12. Graphical user interface

The value added by any research is cogently dependent on the industrial applicability and viability. Although, ample research related to the implementation of various ML and statistical modelling on forecasting of concrete properties, is reported every year, few of this research concludes on drafting a GUI to give an insight into the industrial/field pertinence of the research.

In context of this research, a GUI in the form of a web app, to predict CS, EFNARC SF class and EFNARC VF class was deployed with the help of *streamlit* platform. The script for the app code was written on python 3.10.12 and the ML as well as stacked model files were saved and deployed in .joblib format.

3. RESULTS AND DISCUSSION

The prediction performance of ML models, the sensitivity analysis, SHAP analysis for CS, SF and VF have been discussed in this section. Additionally, the pareto optimisation and the GUI have also been discussed. Six ML models based on MLP, GBR, BG, XGB, KNN, SVR algorithms and six stacked models were trained on the datasets of each property i.e., CS, SF and VF. The nomenclature of these models was done according to the algorithm followed by the predicted property; for example, a model named XGB_SF is based on XGB algorithm and is developed to predict the SF of SCRAC.

3.1. Prediction of compressive strength

The CS dataset consisted of 431 samples. The ML models were hypertuned with PSO algorithm. The hyperparametric details of these models are presented in Table 6. Further the scatter plots of the model predictions on the test set and all the evaluation metrics of these models are presented in Figure 5 and Figure 6 respectively. Additionally, the ranking of all models calculated on the basis of the evaluation metrics has been presented in Table 7.

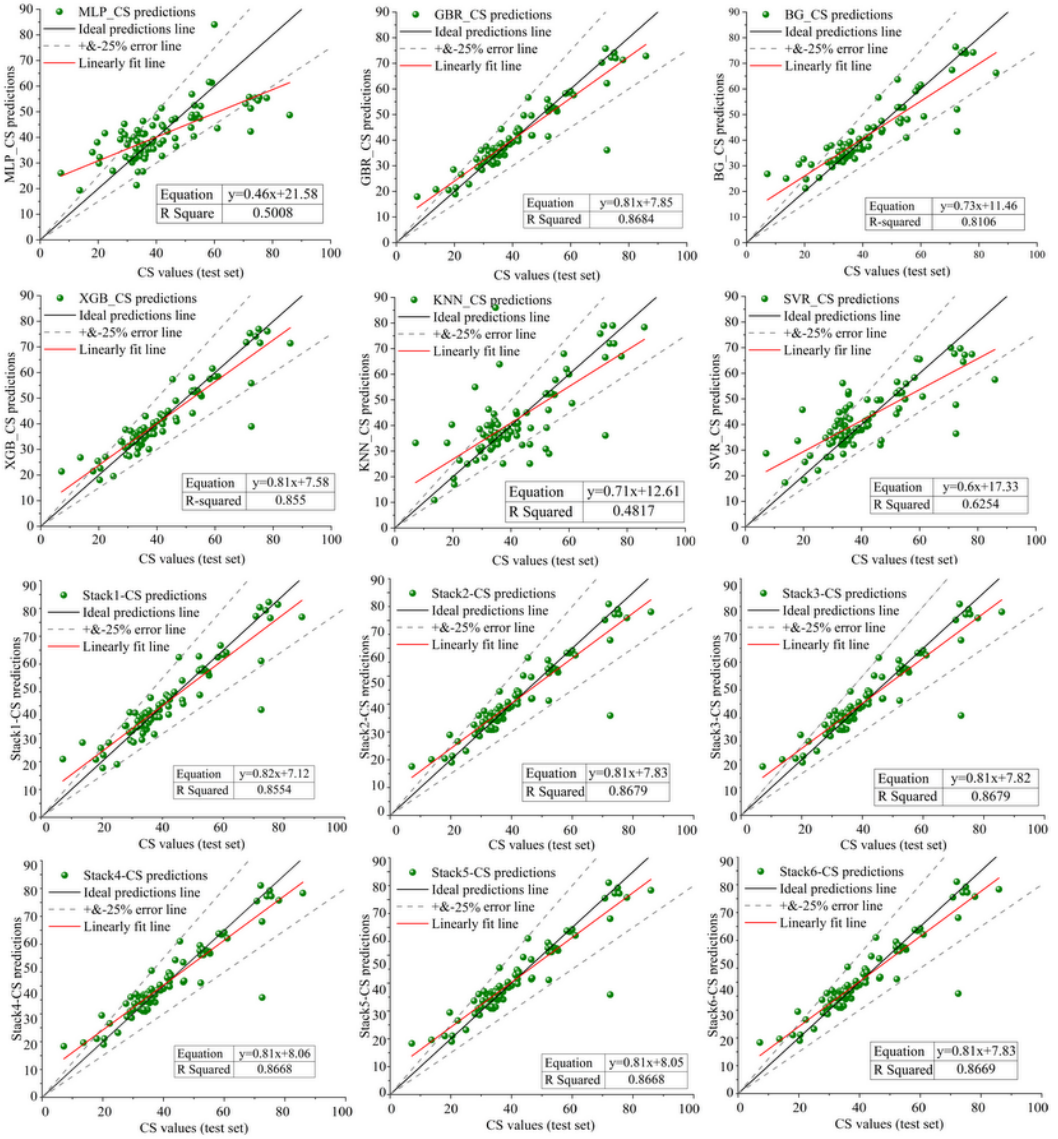


Figure 5. Scatter plots of the CS predictions.

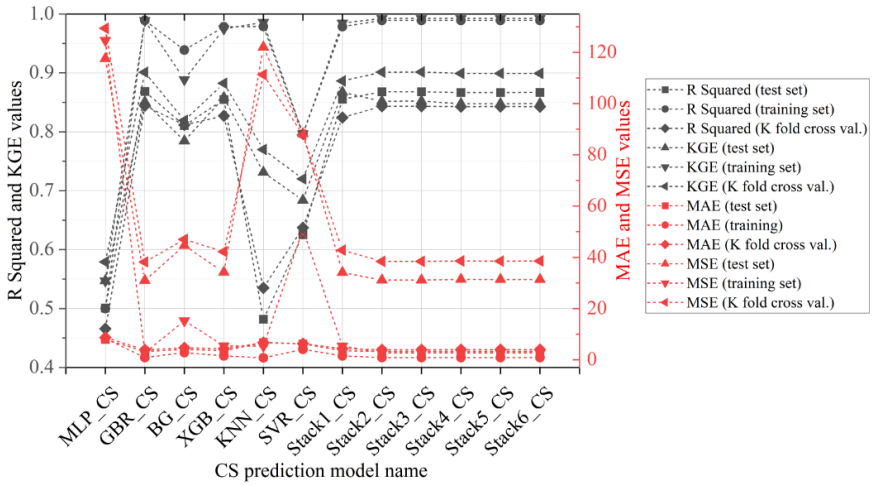


Figure 6. Evaluation metrics of the CS predictions.

Table 5. Carbon footprint and cost of SCRAC mix constituents (41).

Trained model	Hyperparameters
MLP_CS	Hidden layer size = 21, iterations = 409, alpha = 0.005, beta1= 0.0914, beta2-0.902, learning rate = 0.039, power t = 0.788, and tolerance = 2.385.
GBR_CS	No. of base estimators = 281, subsample = 0.65, max depth = 25, max features = 7, min samples split = 16, min samples in leaf = 7, min weight fraction leaf = 0.02, min leaf nodes = 374, min impurity decrease for splitting = 0.016, and learning rate = 0.2016.
BG_CS	No. of base estimators = 139, max samples =199, max features =6, and random state = 81
XGB_CS	No. of base estimators = 660, learning rate = 0.2476, max depth = 8, max leaves = 19, min child weight = 8.19, and subsample =0.5012.
KNN_CS	No. of neighbours = 1, leaf size = 188, and p = 1
SVR_CS	C=38.23, and epsilon = 0.3349
Stack1_CS	Linear stacking of BG_CS, and XGB_CS
Stack2_CS	Linear stacking of GBR_CS, BG_CS, and XGB_CS
Stack3_CS	Linear stacking of GBR_CS, BG_CS, XGB_CS, and MLP_CS
Stack4_CS	Linear stacking of GBR_CS, BG_CS, XGB_CS, MLP_CS, and KNN_CS
Stack5_CS	Linear stacking of Stack2_CS, Stack3_CS, and Stack4_CS
Stack6_CS	Linear stacking of Stack1_CS, Stack2_CS, Stack3_CS, and Stack4_CS

It can be observed that among the ML models, the best CS prediction performance in terms of R^2 and KGE, on test set, training set as well as with K fold cross validation, was exhibited by the GBR_CS, followed by XGB_CS, BG_CS, SVR_CS, MLP_CS and KNN_CS in respective order. Therefore, the best ranking is achieved by GBR_CS model, followed by the stacked models. Indeed, the tree based models have performed substantially better than other ML models such as SVR_CS, KNN_CS and MLP_CS. This can be attributed to the additive learning approach which establishes an adequate bias variance trade off in the predictions. The hyperplane soft margin (ε) maximisation function of the SVR_CS resulted in better fitting as compared to the neighbour averaging approach of KNN_CS and gradient descent in MLP_CS. The hyperplane fitting principally exhibits low variance and high bias whereas the neighbour averaging shall exhibit low bias and high variance. Therefore, the SVR_CS could fit better to the restrained but erratic nature of the CS of SCRACs.

Table 6. Carbon footprint and cost of SCRAC mix constituents (41).

Model	Test set predictions								Training set predictions								K fold cross validation								Total score	Ranking
	R^2		KGE		MAE		MSE		R^2		KGE		MAE		MSE		R^2		KGE		MAE		MSE		score	mg
	V	*S*	V	*S*	V	*S*	V	*S*	V	*S*	V	*S*	V	*S*	V	*S*	V	*S*	V	*S*	V	*S*	V	*S*		
MLP_CS	0.50	*2*	0.55	*1*	7.9	*1*	117	*2*	0.50	*1*	0.55	*1*	8.3	*1*	124	*1*	0.47	*1*	0.58	*1*	8.7	*1*	129	*1*	*14*	*12*
GBR_CS	0.87	*12*	0.85	*8*	3.2	*12*	31.0	*12*	0.99	*7*	0.99	*7*	0.8	*11*	2.7	*7*	0.84	*12*	0.90	*10*	3.9	*12*	38.2	*12*	*122*	***1***
BG_CS	0.81	*4*	0.78	*4*	4.1	*4*	44.6	*4*	0.94	*3*	0.89	*3*	2.8	*3*	15.2	*3*	0.81	*4*	0.82	*4*	4.7	*4*	47.0	*4*	*44*	*10*
XGB_CS	0.86	*5*	0.86	*11*	3.5	*6*	34.1	*5*	0.98	*4*	0.97	*4*	1.5	*4*	5.4	*4*	0.83	*6*	0.88	*5*	4.3	*6*	42.2	*6*	*66*	*8*
KNN_CS	0.48	*1*	0.73	*3*	6.7	*2*	122	*1*	0.98	*6*	0.99	*6*	0.8	*12*	5.3	*6*	0.54	*2*	0.77	*3*	6.7	*2*	111	*2*	*46*	*9*
SVR_CS	0.63	*3*	0.68	*2*	6.2	*3*	88.2	*3*	0.80	*2*	0.80	*2*	4.1	*2*	50.9	*2*	0.64	*3*	0.72	*2*	6.3	*3*	87.8	*3*	*30*	*11*
Stack1_CS	0.86	*6*	0.87	*12*	3.5	*5*	34.1	*6*	0.98	*5*	0.98	*5*	1.5	*5*	5.4	*5*	0.82	*5*	0.89	*6*	4.3	*5*	42.8	*5*	*70*	*7*
Stack2_CS	0.87	*10*	0.85	*9*	3.2	*10*	31.1	*10*	0.99	*8*	0.99	*8*	0.9	*7*	2.7	*8*	0.84	*10*	0.90	*11*	3.9	*11*	38.4	*10*	*112*	*3*
Stack3_CS	0.87	*11*	0.85	*10*	3.2	*11*	31.1	*11*	0.99	*9*	0.99	*9*	0.9	*6*	2.7	*9*	0.84	*11*	0.90	*12*	3.9	*10*	38.4	*11*	*120*	*2*
Stack4_CS	0.87	*7*	0.85	*5*	3.3	*7*	31.4	*7*	0.99	*11*	0.99	*11*	0.8	*10*	2.7	*11*	0.84	*7*	0.90	*8*	3.9	*7*	38.5	*7*	*98*	*6*
Stack5_CS	0.87	*8*	0.85	*6*	3.3	*8*	31.4	*8*	0.99	*10*	0.99	*10*	0.8	*9*	2.7	*10*	0.84	*9*	0.90	*7*	3.9	*9*	38.5	*9*	*103*	*5*
Stack6_CS	0.87	*9*	0.85	*7*	3.3	*9*	31.3	*9*	0.99	*12*	0.99	*12*	0.8	*8*	2.7	*12*	0.84	*8*	0.90	*9*	3.9	*8*	38.5	*8*	*111*	*4*

Where* **V *represents the value of evaluation metric and* ***S*** *represents the score of models from 1 to 12, based on that evaluation metric.*

Nevertheless, a decline in prediction performance of all models was observed on the test set as compared to the training set. This aligns with the most relevant literature (42-46). The maximum decline in R^2 values was observed for KNN_CS (50.78% drop), followed by SVR_CS (21.44%) and the drop in performance of all other models ranged between 12% and 14%. Further the decline in the prediction was even more as per the K fold cross validation assessment. The highest decline in R^2 values in per K fold results when compared to training set predictions, was observed for KNN_CS (45.3%) followed by SVR_CS (19.9%), and the drop in performance of all other models ranged between 14% and 16%. This follows from the fact that the ML models always perform better predictions on the data on which it has been trained as compared to the data on which is new to it (47, 48). Moreover, averaging the performance after training and testing on different K fold sets further lowers the ability of the originally chosen hyperparameters of the initial model to fit well on all other sets (33, 34). The results were confirmed even by the error metrics scores i.e., MAE and MSE scores.

Apart from this, the stacked models performed even better than most of the ML models with test set R^2 ranging from 0.86 to 0.87. This was expected since the stacked models fit the results of already fitted models to the target

variable. However, whilst increasing the number of models as input to stacking process led to slight improvement in performance of Stack2_CS as compared to Stack1_CS, the performance stagnates after further incorporation of more models into stacking for Stack3_CS, and Stack4_CS. This stagnation prevails even when second level stacking was done in models Stack5_CS and Stack6_CS indicating that there is an optimum level of improvement that could be achieved by stacking. Nonetheless, the best performing model was the GBR_CS and this model was further used in the GUI discussed ahead in the paper.

3.1.1. Sensitivity and SHAP Analysis

The impact of every input feature on the actual CS, calculated as per the Equation [13], considering p=10 and the absolute mean SHAP trends have been depicted in Figure 7. The FAsh/B ratio exhibits high impact of 29% on the CS, which is in line with the high absolute mean SHAP values relative to the other input features. Moreover, this is in line with the existing literature. The increase in FAsh content in the binder beyond a limit weakens the C-S-H gel and leads to a significant reduction in the CS (49, 50).

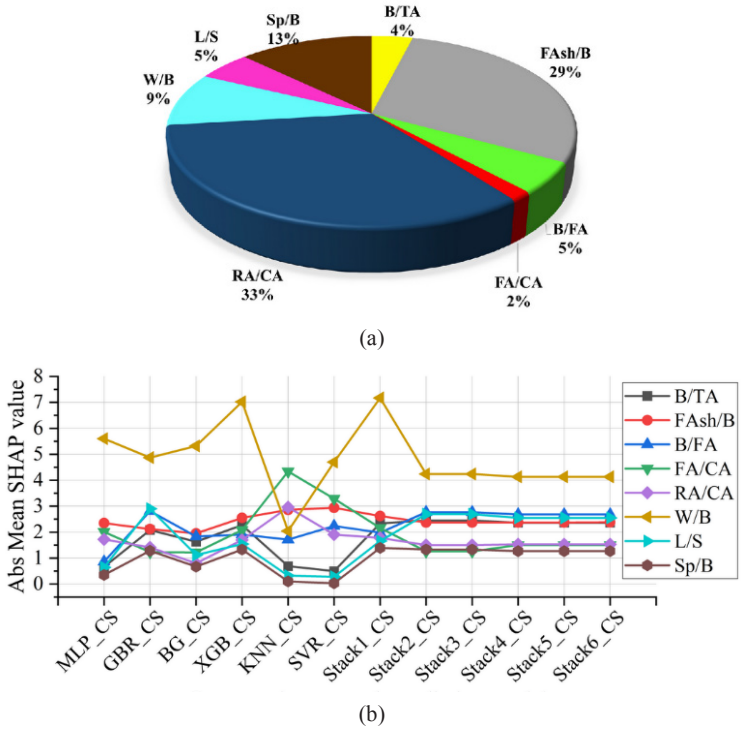


Figure 7. (a) Percentage impact as per sensitivity analysis and (b) Absolute mean SHAP values of mix ratios for all CS prediction models.

Likewise, the B/TA and FA/CA exhibit a low impact of 3.9% and 1.5% on the CS which is consistent with the low range of SHAP values relative to the other input features. However, the sensitivity analysis impact and absolute mean SHAP results of B/FA, L/S, RA/CA, Sp/B and W/B show some contradiction with each other. The B/FA, L/S and W/B exhibit a low impact of 5.2%, 5.1% and 9% on the CS respectively but high absolute mean SHAP values relative to most other input features. Whereas the RA/CA and Sp/B exhibit high impact values of 33.3% and 12.7% respectively on the CS but low range of absolute mean SHAP values relative to the other input features. This conundrum is subject to the variations in compositions of binder of B/FA, source and properties of RA in RA/CA, contents of liquids such as water and superplasticizers in L/S and W/B, and type of superplasticizer in Sp/B. Since these parameters vary significantly in the dataset gathered from wide sources, the standardisation of these parameters could give more reliable fit and is considerable as a future scope of research.

3.2 Prediction of fresh properties

The hyperparametric details of the ML and stacked models developed to predict the SF and VF have been presented in Table 8. Further, the scatter plots of the model predictions on the test set have been presented in Figure 8 and Figure 9. Additionally, the evaluation metrics of all these models on the test set, training set and with K fold cross validation, and the ranking of all models have been presented in Table 9.

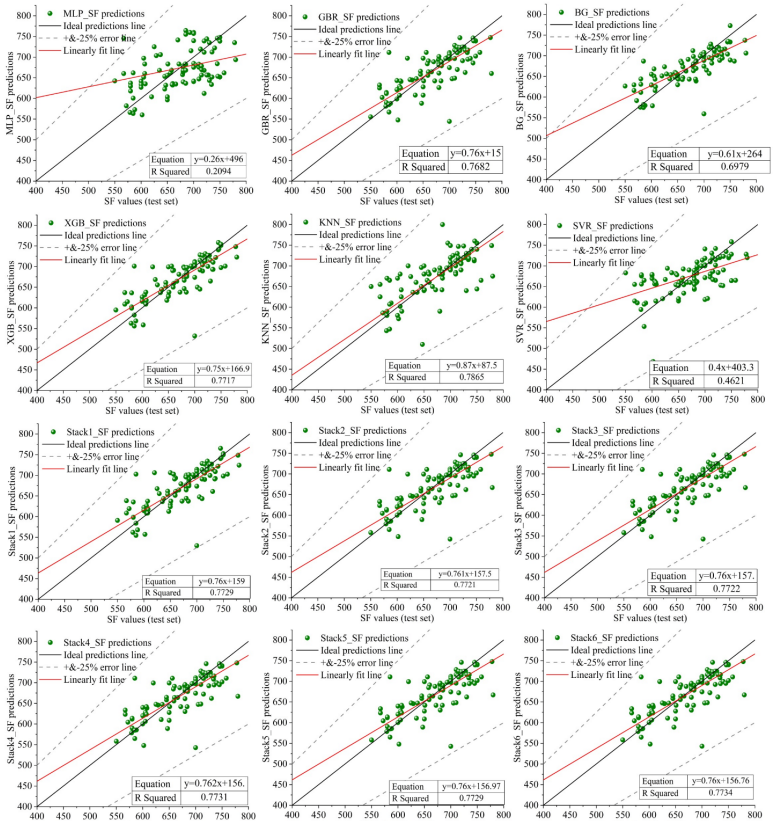


FIGURE 8. Scatter plots of the SF predictions.

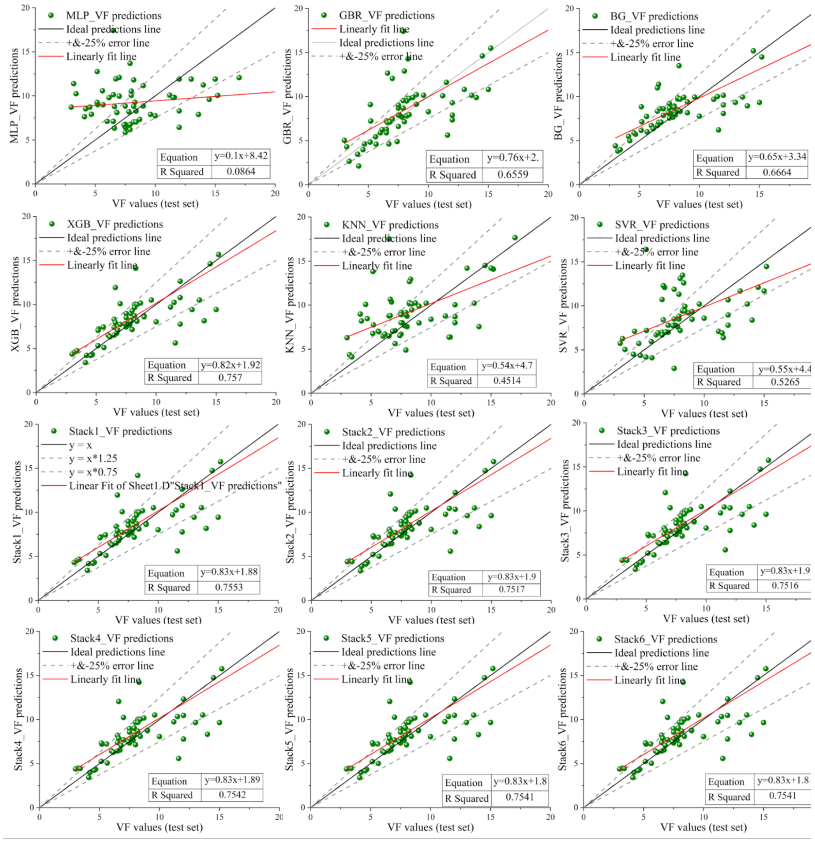


FIGURE 9. Scatter plots of the VF predictions.

Table 8. The hyperparametric details of the ML models trained for predicting SF and VF time.

Trained model	Hyperparameters
MLP_SF	Hidden layer size = 162, iterations = 361, alpha = 0.0035, beta1= 0.869, beta2=0.900, learning rate = 0.023, power t = 0.600, and tolerance = 9.9368.
GBR_SF	No. of base estimators = 289, subsample = 0.72, max depth = 8, max features = 8, min samples split = 15, min samples in leaf = 7, min weight fraction leaf = 0.023, min leaf nodes = 279, min impurity decrease for splitting = 0.023, and learning rate = 0.1653.
BG_SF	No. of base estimators = 176, max samples =200, max features =5, and random state = 177
XGB_SF	No. of base estimators = 745, learning rate = 0.2391, max depth = 5, max leaves = 57, min child weight = 6.838, and subsample =0.8509.
KNN_SF	No. of neighbours = 1, leaf size = 39, and p = 1
SVR_SF	C=100, and epsilon = 0.01
Stack1_SF	Linear stacking of BG_SF, and XGB_SF
Stack2_SF	Linear stacking of GBR_SF, BG_SF, and XGB_SF
Stack3_SF	Linear stacking of GBR_SF, BG_SF, XGB_SF, and MLP_SF
Stack4_SF	Linear stacking of GBR_SF, BG_SF, XGB_SF, MLP_SF, and KNN_SF
Stack5_SF	Linear stacking of Stack2_SF, Stack3_SF, and Stack4_SF
Stack6_SF	Linear stacking of Stack1_SF, Stack2_SF, Stack3_SF, and Stack4_SF
MLP_VF	Hidden layer size = 66, iterations = 361, alpha = 0.00158, beta1= 0.898, beta2=0.965, learning rate = 0.09266, power t = 0.8298, and tolerance = 6.415.
GBR_VF	No. of base estimators = 209, subsample = 0.69, max depth = 12, max features = 2, min samples split = 12, min samples in leaf = 7, min weight fraction leaf = 0.037, min leaf nodes = 145, min impurity decrease for splitting = 0.0289, and learning rate = 0.1673.
BG_VF	No. of base estimators = 22, max samples =195, max features =3, and random state = 367
XGB_VF	No. of base estimators = 51, learning rate = 0.495, max depth = 9, max leaves = 24, min child weight = 1, and subsample =0.7286.
KNN_VF	No. of neighbours = 5, leaf size = 5, and p = 1
SVR_VF	C=63.48, and epsilon = 0.2
Stack1_VF	Linear stacking of BG_VF, and XGB_VF
Stack2_VF	Linear stacking of GBR_VF, BG_VF, and XGB_VF
Stack3_VF	Linear stacking of GBR_VF, BG_VF, XGB_VF, and MLP_VF
Stack4_VF	Linear stacking of GBR_VF, BG_VF, XGB_VF, MLP_VF, and KNN_VF
Stack5_VF	Linear stacking of Stack2_VF, Stack3_VF, and Stack4_VF
Stack6_VF	Linear stacking of Stack1_VF, Stack2_VF, Stack3_VF, and Stack4_VF

It can be observed in the rankings calculated in Table 9 that the best prediction performance was exhibited by Stack6_SF among SF prediction models and Stack4_VF among VF prediction models. Indeed, the stacked models performed better than the ML models. This was expected since the stacked models fit the results of already fitted models to the target variable. However, the patterns of stacked models were similar to the CS prediction stacked models. Whilst increasing the number of models as input to stacking process led to slight improvement in performance of Stack2_SF Stack2_VF as compared to Stack1_SF and Stack1_VF, the performance stagnates after further incorporation of more models into stacking for Stack3_SF, Stack4_SF, Stack3_VF and Stack4_VF. This stagnation sustained even when second level stacking was done in models Stack5_SF, Stack6_SF, Stack5_VF and Stack6_VF. This indicates that there is an optimum level of improvement that could be achieved by stacking even in case of SF and VF predictions.

However, among the ML models, the tree based models have performed substantially better than other ML models SVR, KNN and MLP, both for SF and VF predictions. This can be attributed to the additive learning approach which establishes an adequate bias variance trade off in the predictions as compared to the neighbour averaging approach of KNN, hyperplane soft margin (ε) maximisation tack of the SVR and gradient descent approach of MLP. The neighbour averaging shall exhibit low bias and high variance whereas the hyperplane fitting principally exhibits low variance and high bias. Nonetheless, although the PSO was adopted for hyperparameter tuning, the MLP did not perform well in prediction. This points towards the requirement of better hyperparameter tuning or use of deep neural networks for prediction of SF and VF, just as in the case of CS prediction.

It is further evident from the evaluation metrics that the prediction performance of most models declined on the test set as compared to the training set both for the SF and the VF predictions. In case of SF predictions, minimum decline in R^2 was exhibited by KNN_SF (8.54%) and maximum decline was exhibited by SVR_SF (21.75%). The stacked models exhibited a decline in the range of 16-17%. Likewise, for VF predictions, the minimum decline in R^2 was exhibited by BG_VF (17.6%) and maximum decline by GBR_VF (31.1%). The stacked models exhibited a decline in range of 21-22%. This decline is not too high for the developed models, corroborating the reliability and robustness of their prediction performances.

Further, a decline was also apparent in performance of all models upon K fold cross validation as compared to training set predictions. For SF predictions, the minimum decline in R^2 was exhibited by MLP_SF (7.33%) and maximum decline was exhibited by KNN_SF (51%). The stacked models exhibited a R^2 decline ranging from 27% to 30.5%. However, for VF predictions, the decline in R^2 upon K fold cross validation as compared to training set predictions was high. The best performing stacked models exhibited a decline ranging from 66% to 76%. This owes to the complex nature of the confounded relations between flow times and the mix characteristics. Therefore, the VF time often surfaces as a highly volatile and unpredictable fresh property of SCCs, and is often complex to model. Nevertheless, the best performing models Stack6_SF and Stack4_VF were further used in the GUI for giving a fair idea about the properties of SCRAC.

Table 9. Scores and ranking of the SF and VF prediction models, respectively.

Model	Test set predictions								Training set predictions								K fold cross validation								Total score	Ranking
	R^2		KGE		MAE		MSE/100		R^2		KGE		MAE		MSE/100		R^2		KGE		MAE		MSE/100			
	V	S	V	S	V	S	V	S	V	S	V	S	V	S	V	S	V	S	V	S	V	S	V	S		
MLP_SF	0.21	1	0.31	1	53	1	65	1	0.19	1	0.27	1	59	1	68.6	1	0.17	1	0.29	1	59.0	1	69.0	1	12	12
GBR_SF	0.77	4	0.82	8	26	5	19.0	4	0.93	7	0.95	7	10	11	6.0	7	0.68	11	0.80	7	29.6	12	26.2	9	89	6
BG_SF	0.70	3	0.68	4	29	3	24.8	3	0.85	3	0.79	3	22	3	12.3	3	0.69	12	0.69	3	32.8	4	26.2	6	49	10
XGB_SF	0.77	5	0.81	11	25	12	18.7	5	0.92	5	0.94	5	12	4	6.3	5	0.66	5	0.78	5	30.8	6	28.3	5	66	9
KNN_SF	0.79	12	0.89	3	28	4	17.5	12	0.86	4	0.92	4	9	12	11.8	4	0.42	2	0.72	4	38.6	3	47.2	2	75	8
SVR_SF	0.46	2	0.48	2	41	2	44.1	2	0.59	2	0.53	2	33	2	34.5	2	0.46	3	0.45	2	43.5	2	45.8	3	26	11
Stack1_SF	0.77	8	0.82	12	26	11	18.6	8	0.93	6	0.95	6	12	5	6.3	6	0.65	4	0.78	6	31.3	5	29.2	4	77	7
Stack2_SF	0.77	6	0.82	9	26	7	18.7	6	0.93	8	0.95	8	10	10	5.9	8	0.68	6	0.80	8	29.8	10	26.2	7	90	5
Stack3_SF	0.77	7	0.82	10	26	6	18.7	7	0.93	11	0.95	11	10	6	5.9	11	0.68	7	0.80	9	29.8	11	26.2	8	101	4
Stack4_SF	0.77	10	0.82	5	26	9	18.6	10	0.93	10	0.95	10	10	8	5.9	10	0.68	9	0.80	10	29.8	8	26.2	11	115	2
Stack5_SF	0.77	9	0.82	6	26	8	18.6	9	0.93	9	0.95	9	10	9	5.9	9	0.68	8	0.80	11	29.8	7	26.2	10	107	3
Stack6_SF	0.77	11	0.82	7	26	10	18.6	11	0.93	12	0.95	12	10	7	5.9	12	0.68	10	0.80	12	29.8	9	69.0	12	129	**1**

Model	Test set predictions								Training set predictions								K fold cross validation								Total score	Ranking
	R^2		KGE		MAE		MSE		R^2		KGE		MAE		MSE		R^2		KGE		MAE		MSE			
	V	S	V	S	V	S	V	S	V	S	V	S	V	S	V	S	V	S	V	S	V	S	V	S		
MLP_VF	0.09	1	0.03	1	4.1	1	39.9	1	0.07	1	0.02	1	4.0	1	61.4	1	0.07	1	0.10	1	4.1	1	57.2	1	12	12
GBR_VF	0.66	4	0.80	5	2.4	4	15.0	4	0.95	5	0.92	5	0.9	5	3.2	5	0.47	12	0.64	12	2.7	11	37.8	12	84	8
BG_VF	0.67	5	0.73	4	2.3	5	14.6	5	0.81	4	0.72	4	1.5	4	12.7	4	0.44	11	0.57	6	2.6	12	38.8	11	75	9
XGB_VF	0.76	12	0.86	6	1.9	12	10.6	12	0.96	6	0.97	6	0.7	12	2.5	6	0.24	3	0.54	5	2.9	4	46.0	4	88	7
KNN_VF	0.45	2	0.62	2	2.9	2	24.0	2	0.45	2	0.51	3	2.5	2	36.7	2	0.25	4	0.47	2	3.2	2	48.5	2	27	11
SVR_VF	0.53	3	0.63	3	2.9	3	20.7	3	0.45	3	0.44	2	1.8	3	36.4	3	0.27	5	0.47	3	3.1	3	47.0	3	37	10
Stack1_VF	0.76	11	0.86	12	1.9	11	10.7	11	0.96	7	0.97	7	0.7	11	2.4	7	0.23	2	0.54	4	2.8	5	45.9	5	93	5
Stack2_VF	0.75	7	0.86	8	1.9	7	10.8	7	0.96	8	0.97	8	0.7	9	2.4	8	0.32	6	0.58	7	2.8	7	43.0	7	89	6
Stack3_VF	0.75	6	0.86	7	1.9	6	10.8	6	0.96	9	0.97	9	0.7	10	2.4	9	0.32	8	0.58	9	2.8	10	42.9	10	99	4
Stack4_VF	0.75	10	0.86	9	1.9	10	10.7	10	0.96	12	0.97	12	0.7	7	2.4	12	0.32	9	0.58	10	2.8	9	42.9	8	118	**1**
Stack5_VF	0.75	8	0.86	11	1.9	8	10.7	8	0.96	10	0.97	10	0.7	6	2.4	10	0.32	10	0.58	11	2.8	8	42.9	9	109	2
Stack6_VF	0.75	9	0.86	10	1.9	9	10.7	9	0.96	11	0.97	11	0.7	8	2.4	11	0.32	7	0.58	8	2.8	6	43.0	6	105	3

Where* **V *represents the value of evaluation metric and* ***S*** *represents the score of models from 1 to 12, based on that evaluation metric.*

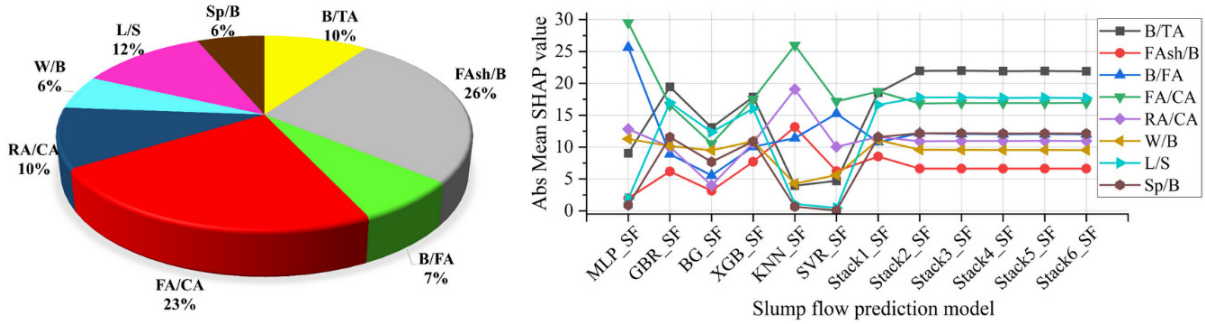
3.2.1. Sensitivity and SHAP Analysis

The impact of every input feature on the actual SF and VF calculated as per the Equation [15], considering p=50, has been depicted in Figure 10 (a) and Figure 10 (c) respectively. Additionally, the absolute mean SHAP values of all

the eight input features have been plotted for all the SF and VF prediction models, in Figure 10 (b) and Figure 10 (d) respectively. Although there is a slight difference in the patterns of relative importance of input features on the SF and VF, the parallel inferences that can be drawn from the sensitivity analysis and the SHAP analysis are discussed as follows.

The FA/CA and L/S ratios exhibit significant impact of 23% and 12% on the SF, respectively which is in line with the high absolute mean SHAP values relative to the other input features. Moreover, this is in line with the existing literature. The high FA leads to increased paste volume and easy flowability. Similarly, the high liquid content directly affects the flowability and SF of concrete as well.

The RA/CA, B/FA, B/TA and Sp/B showed moderate impact of 9.9%, 7%, 9.8% and 6.3% on the SF respectively which is consistent with the lower SHAP positioning in Figure 10 (b) relative to other input features. The detrimental effect of RA on the SF is subject to the rough nature of the surface of RA. Whereas the increase in binder content (B/FA and B/TA) leads to moderate enhancement of the SF due to paste lubrication effect. Similarly, the increase Sp/B majorly leads to reduction in yield stress of the paste and the concrete thus enhancing the SF.



(a) (b)

Figure 10. (a) Percentage impact as per sensitivity analysis for SF, (b) Absolute mean SHAP values of mix ratios for all SF prediction models, (c) Percentage impact as per sensitivity analysis for VF, (d) Absolute mean SHAP values of mix ratios for all VF prediction models.

The W/B exhibits low impact of 5.65% on the SF which is consistent with the low position of SHAP trend relative to other input features. It signifies that the effect of total liquid, the role of Sp and the secondary effect of the binder constituents is more predominant on the SF as compared to the W/B ratio.

Nonetheless, the impact of FAsh/B on SF as per sensitivity analysis shows inconsistency with the SHAP trend. The impact of FAsh/B was 26.2% on SF as per sensitivity analysis whereas the lowest absolute mean SHAP values were observed. This points towards the interference of the effects of other mix ratios and further the varying compositions of the different mix constituents.

For the case of VF time, FA/CA and RA/CA ratios exhibit high impact of 22.8% and 13% on the VF respectively, which is in line with the high absolute mean SHAP values relative to the other input features. Parallelly, the W/B and B/FA exhibited low impact of 1.3% and 7.31% on VF which matches with the moderately low absolute mean SHAP trend as compared to other input features such as L/S and RA/CA (shown in fig. 20). However, the sensitivity analysis impact and the absolute mean SHAP trend of other mix ratios on VF prediction, shows a contradictory trend. The FAsh/B and Sp/B exhibit high impact values of 31.2% and 22.6% on VF but low position in absolute mean SHAP trends. On the other hand, the B/TA and L/S ratios exhibit low impact of 0.5% and 1.2% on VF but high position in SHAP trends. This points towards the interference of the effects of other mix ratios and further the varying compositions of the different mix constituents. In fact, this effect is even more pronounced than for SF prediction case.

3.3. Multi Objective Pareto Optimisation

The four objectives chosen for optimisation of the SCRAC concrete included, maximisation of CS, maximisation of SF, minimisation of CO_2 emissions and minimisation of Cost. The Figure 11 (a) shows multiple objective pareto fronts of the whole dataset, considering CS-SF-CO_2 emission and CS-SF-Cost. The cost and the CO_2 emissions of the SCRAC mixes were calculated using the typical values mentioned in Table 5. Further, the Figure 11(b), Figure 11(c), and Figure 11(d) present the trade off and partial pareto fronts between CS-SF, CS-CO_2 emission, and CS-Cost respectively. Few outliers were ignored for fitting the partial pareto front curves, subject to exceptional compositions of some mixes in the literature referred for preparing the dataset.

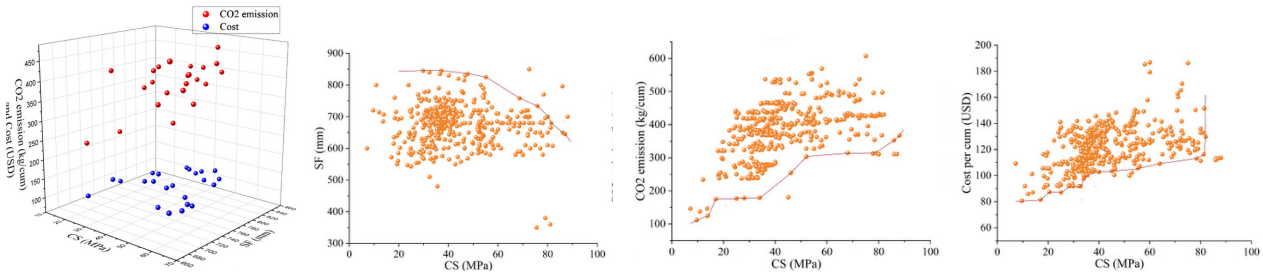


Figure 11. The figure shows (a) Multi objective pareto front, (b) SF-CS trade off and partial pareto front, (c) CS-CO_2 emissiontrade off and partial pareto front, and (d) CS-Cost trade off and partial pareto front.

A smooth curve is observed for the CS-SF trade off. In the CS-SF partial pareto front, a SF value of 650 mm is observed for the maximum CS of 88MPa, whereas the rise in SF value stagnates to 850 mm when the CS is sacrificed below 38 MPa. On the other hand, an irregular partial pareto front curve is seen for CS-CO_2 emissions trade off. The minimum CO_2 emission of 109 kg/cum corresponds to a CS of 9.8 MPa and maximum CO_2 emission of 375 kg/cum corresponds to a CS of 88 MPa in the optimals of CS-CO_2 emissions partial pareto front. Further, the lowest pareto optimal of CS-Cost trade off exhibits a minimum cost of about 81 USD for CS of 10 MPa. However, a drastic rise in cost is seen in curve beyond the CS of 80 MPa primarily due to increased content of binder. Since these curves represent a wide variety of SCRAC mixes reported in the relevant research, the patterns of these curves can be considered for mix design optimisation in engineering and construction practice.

3.4. Graphical User Interface

A visual of the GUI prepared for the prediction of CS, SF and VF has been presented in Figure 12. The GUI was prepared using the streamlit platform. A Github repository was created to ensure continuous access of the python script and model files to the streamlit GUI. The model files were prepared in .joblib format and stored in the Github codespace.

In the GUI, the contents of SCRAC mix constituents such as cement, FAsh, FA, NA, RA, water and Sp were set as the input instead of the mix ratios. This was done to ascertain ease of use with respect to the engineering and industrial applications. The mix ratios are calculated from the values of mix constituents and fed to the models GBR_CS, Stack6_SF and Stack4_VF models for prediction of respective properties. These models predict the value of CS in MPa, EFNARC SF class (as per the predicted SF value), EFNARC VF class (as per the predicted VF value). The GUI can be accessed on request, through the link given in the caption of Figure 12.

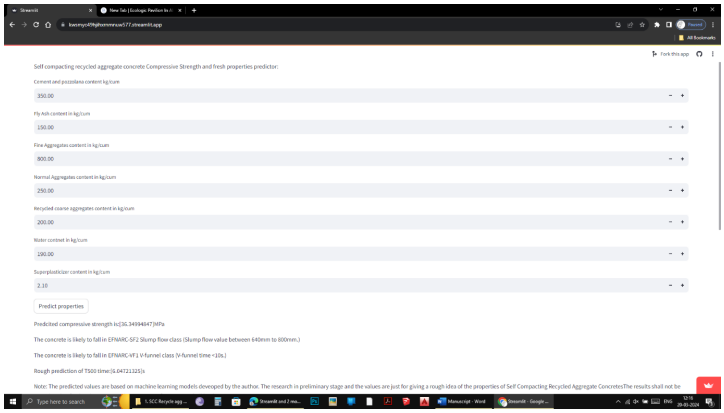


Figure 12. The streamlit GUI prepared for prediction of CS and fresh properties, accessible on request through https://kwsmyc49hjihxrnmnuw577.streamlit.app/

CONCLUSIONS

The research aimed at the development of ML assisted framework for prediction of CS and two fresh properties (SF and VF) of SCRAC. SCRAC is an important advancement in self compacting concretes, which incorporates recycled aggregates as a replacement to natural aggregates. The key findings of the research are summarised as follows.

- The tree based models performed better in CS prediction, as compared to other ML models such as SVR_CS, MLP_CS and KNN_CS. The best CS prediction performance was exhibited by GBR_CS (R^2 = 0.8684 on test set), which was marginally better than the prediction performance of even the stacked models. The worst prediction performance was exhibited by KNN_CS (R^2 = 0.4817 on test set).
- The tree based models performed better even in the SF prediction, as compared to the other ML models such as SVR_SF, MLP_SF and KNN_SF. By and large, the six stacked models performed even better than the tree based models. The best prediction performance was exhibited by Stack6_SF (R^2 = 0.7764 on test set). However, the prediction performance metrics exhibited lower accuracy as compared to CS predictions. This is attributable to high bias and volatile nature of SF. SF depends hugely on inherent characteristics of constituent minerals and Sp type, and hence the bias. Poorest SF prediction performance was exhibited by MLP_SF (R^2 = 0.2094 on test set).
- The best VF prediction was exhibited by Stack4_VF (R^2 = 0.7542 on test set). Again, the stacked models performed better than the ML models and tree based models performed better than the SVR_VF, KNN_VF and MLP_VF. The flow times of SCRAC are even more volatile and unpredictable than SF due to number of compositional and environmental factors. The poorest prediction performance was exhibited by MLP_VF (R^2 = 0.0864 on test set).
- The maximum impact on CS was exhibited by the FAsh/B (33%) as per the sensitivity analysis. However, the FA/CA exhibited maximum impact of 23% on both SF and VF. This was coherent with the trends of absolute mean SHAP values.
- The SC-SF partial pareto front exhibits maximum SF of 650mm for the CS of 88 MPa, whereas the rise in SF stagnates the range of 850mm when the CS is sacrificed below 38 MPa.
- The lowest pareto optimal of CS-Cost trade off exhibits a minimum cost of about 81 USD for CS of 10 MPa. However, a drastic rise in cost is seen in curve beyond the CS of 80 MPa primarily due to increased content of binder.
- The minimum CO_2 emission of 109 kg/cum corresponds to a CS of 9.8 MPa and maximum CO_2 emission of 375 kg/cum corresponds to a CS of 88 MPa in the optimals of CS-CO_2 emissions partial pareto front.
- The GUI performs the prediction of CS, SF and VF by calculating mix ratios from the input values of mix constituents. Keeping the values in line with the descriptive statistics of the datasets, the GUI prepared for SCRAC properties prediction, gives a lead towards application of AI in SCRAC mix design.

Acknowledgements

The authors are pleased to express sincere gratitude towards the department of civil engineering and department of computer science and engineering, of Dr B R Ambedkar National Institute of Technology, Jalandhar, India for their constant support in this research.

Authorship Contribution Statement

Tejinderpal Singh: Write up-original draft, conceptualisation, data cleansing, formal analysis, research methodology and software.

Kanish Kapoor: Write up-review & editing, conceptualisation, supervision, validation and resources.

Surinder Pal Singh: Write up-review & editing, conceptualisation, supervision and validation.

Data Availability

The data and access to GUI will be made available on request.

Declaration of competing interest

The authors of this article declare that they have no financial, professional or personal conflicts of in-terest that could have inappropriately influenced this work.

REFERENCES

1. Zambon I, Colantoni A, Salvati L. 2019. Horizontal vs vertical growth: Understanding latent patterns of urban expansion in large metropolitan regions. Sci. Total Environ. 654:778–785. https://doi.org/10.1016/j.scitotenv.2018.11.182
2. Akristiniy VA, Boriskina YI. 2018. Vertical cities-the new form of high-rise construction evolution. E3S Web Conf. 33(2018):01041. https://doi.org/10.1051/e3sconf/20183301041
3. Felekoğlu B, Türkel S, Baradan B. 2007. Effect of water/cement ratio on the fresh and hardened properties of self-compacting concrete. Build. Environ. 42(4):1795–1802. https://doi.org/10.1016/j.buildenv.2006.01.012

4. Aslam MS, Huang B, Cui L. 2020. Review of construction and demolition waste management in China and USA. J. Environ. Manage. 264:110445. https://doi.org/10.1016/j.jenvman.2020.110445
5. Moschen-Schimek J, Kasper T, Huber-Humer M. 2023. Critical review of the recovery rates of construction and demolition waste in the European Union – An analysis of influencing factors in selected EU countries. Waste Manage. 167:150–164. https://doi.org/10.1016/j.wasman.2023.05.020
6. Jain MS. 2021. A mini review on generation, handling, and initiatives to tackle construction and demolition waste in India. Environ. Technol. Innov. 22:101490 https://doi.org/10.1016/j.eti.2021.101490
7. Huang P, Dai K, Yu X. 2023. Machine learning approach for investigating compressive strength of self-compacting concrete containing supplementary cementitious materials and recycled aggregate. J. Build. Eng. 79:107904. https://doi.org/10.1016/j.jobe.2023.107904
8. Wang Z, Wu B. 2023. A mix design method for self-compacting recycled aggregate concrete targeting slump-flow and compressive strength. Constr. Build. Mater. 404:133309. https://doi.org/10.1016/j.conbuildmat.2023.133309
9. Abed M, Nemes R, Tayeh BA. 2020. Properties of self-compacting high-strength concrete containing multiple use of recycled aggregate. Journal of King Saud University - Engineering Sciences. 32(2):108–114. https://doi.org/10.1016/j.jksues.2018.12.002
10. Martínez-García R, Guerra-Romero IM, Morán-del Pozo JM, de Brito J, Juan-Valdés A. 2020. Recycling aggregates for self-compacting concrete production: A feasible option. Materials. 13(4):868. https://doi.org/10.3390/ma13040868
11. Abed M, Rashid K, Rehman MU, Ju M. 2022. Performance keys on self-compacting concrete using recycled aggregate with fly ash by multi-criteria analysis. J. Clean. Prod. 378:134398. https://doi.org/10.1016/j.jclepro.2022.134398
12. Wang Z, Wu B. 2023. A mix design method for self-compacting recycled aggregate concrete targeting slump-flow and compressive strength. Constr. Build. Mater. 404:133309. https://doi.org/10.1016/j.conbuildmat.2023.133309
13. Chen X, Lv L, Wu Q, Cheng S, Zhou Q, Zhao C, Fan T, Zhao R. 2023. Optimization on the mix design method of self-compacting concrete with recycled coarse aggregate based on paste rheological threshold theory and material packing characteristics. Constr. Build. Mater. 407:133509. https://doi.org/10.1016/j.conbuildmat.2023.133509
14. Long W, Cheng B, Luo S, Li L, Mei L. 2023. Interpretable auto-tune machine learning prediction of strength and flow properties for self-compacting concrete. Constr. Build. Mater. 393:132101. https://doi.org/10.1016/j.conbuildmat.2023.132101
15. Mahmood MS, Elahi A, Zaid O, Alashker Y, Şerbănoiu AA, Grădinaru CM, Ullah K, Ali T. 2023. Enhancing compressive strength prediction in self-compacting concrete using machine learning and deep learning techniques with incorporation of rice husk ash and marble powder. Case Stud. Constr. Mater. 19:e02557. https://doi.org/10.1016/j.cscm.2023.e02557
16. Alarfaj M, Qureshi HJ, Shahab MZ, Javed MF, Arifuzzaman M, Gamil Y. 2024. Machine learning based prediction models for spilt tensile strength of fiber reinforced recycled aggregate concrete. Case Stud. Constr. Mater. 20:e02836. https://doi.org/10.1016/j.cscm.2023.e02836
17. Alyaseen A, Poddar A, Kumar N, Tajjour S, Prasad CVSR, Alahmad H, Sihag P. 2023. High-performance self-compacting concrete with recycled coarse aggregate: Soft-computing analysis of compressive strength. J. Build. Eng. 77:107527. https://doi.org/10.1016/j.jobe.2023.107527
18. Alyaseen A, Poddar A, Kumar N, Sihag P, Lee D, kumar R, Singh T. 2023. Assessing the compressive and splitting tensile strength of self-compacting recycled coarse aggregate concrete using machine learning and statistical techniques. Mater. Today Commun. 38:107970. https://doi.org/10.1016/j.mtcomm.2023.107970
19. de-Prado-Gil J, Martínez-García R, Jagadesh P, Juan-Valdés A, Gónzalez-Alonso M-I, Palencia C. 2023. To determine the compressive strength of self-compacting recycled aggregate concrete using artificial neural network ANN. Ain Shams Eng. J. 15(2):102548. https://doi.org/10.1016/j.asej.2023.102548
20. De-Prado-gil J, Zaid O, Palencia C, Martínez-García R. 2022. Prediction of splitting tensile strength of self-compacting recycled aggregate concrete using novel deep learning methods. Mathematics 10(13):2245. https://doi.org/10.3390/math10132245
21. Huang P, Dai K, Yu X. 2023. Machine learning approach for investigating compressive strength of self-compacting concrete containing supplementary cementitious materials and recycled aggregate. J. Build. Eng. 79:107904. https://doi.org/10.1016/j.jobe.2023.107904
22. Jagadesh P, de Prado-Gil J, Silva-Monteiro N, Martínez-García R. 2023. Assessing the compressive strength of self-compacting concrete with recycled aggregates from mix ratio using machine learning approach. J. Mater. Res. Technol. 24:1483-1498. https://doi.org/10.1016/j.jmrt.2023.03.037
23. Ghahdarijani AM, Hormozi F, Asl AH. 2017. Convective heat transfer and pressure drop study on nanofluids in double-walled reactor by developing an optimal multilayer perceptron artificial neural network. Int. Commun. Heat Mass Transf. 84:11–19. https://doi.org/10.1016/j.icheatmasstransfer.2017.03.014
24. Yao J, Huang S, Xu Y, Gu C, Liu J, Yang Y, Ni T, Kong D. 2023. Mix design of equal strength high volume fly ash concrete with artificial neural network. Case Stud. Constr. Mater. 19:e02294. https://doi.org/10.1016/j.cscm.2023.e02294
25. Friedman JH. 2001. Reitz lecture greedy function approximation: a gradient boosting machine 1:1189–1232.
26. Louk MHL, Tama BA. 2023. Dual-IDS: A bagging-based gradient boosting decision tree model for network anomaly intrusion detection system. Expert Syst. Appl. 213(Part B):119030. https://doi.org/10.1016/j.eswa.2022.119030

27. Chen T, Guestrin C. 2016. XGBoost: A scalable tree boosting system. Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Association for Computing Machinery. 785–794. https://doi.org/10.1145/2939672.2939785
28. Wang AX, Chukova SS, Nguyen BP. 2023. Ensemble k-nearest neighbors based on centroid displacement. Inf. Sci. N.Y. 629:313–323. https://doi.org/10.1016/j.ins.2023.02.004
29. El Bilali A, Abdeslam T, Ayoub N, Lamane H, Ezzaouini MA, Elbeltagi A. 2023. An interpretable machine learning approach based on DNN, SVR, Extra Tree, and XGBoost models for predicting daily pan evaporation. J. Environ. Manage. 327:116890. https://doi.org/10.1016/j.jenvman.2022.116890
30. Mirjalili S. 2019. Particle swarm optimisation. In: Evolutionary algorithms and neural networks. Studies in Computational Intelligence, vol 780. Springer, Cham. https://doi.org/10.1007/978-3-319-93025-1_2
31. Freitas D, Lopes LG, Morgado-Dias F. 2020. Particle swarm optimisation: A historical review up to the current developments. Entropy. 22(3):362. https://doi.org/10.3390/E22030362
32. Zhang C, Shao H. 2000. Particle swarm optimisation for evolving artificial neural network. SMC 2000 conference proceedings. 4:2487-2490. https://doi.org/10.1109/ICSMC.2000.884366.
33. Wong TT, Yeh PY. 2020. Reliable accuracy estimates from k-fold cross validation. IEEE Trans. Knowl. Data Eng. 32(8):1586–1594. https://doi.org/10.1109/TKDE.2019.2912815
34. Fushiki T. 2011. Estimation of prediction error by using K-fold cross-validation. Stat Comput 21:137–146. https://doi.org/10.1007/s11222-009-9153-8
35. Yao Y, Pirš G, Vehtari A, Gelman A. 2022. Bayesian hierarchical stacking: some models are somewhere useful. Bayesian Anal. 17(4):1043–1071. https://doi.org/10.1214/21-BA1287
36. Louhichi M, Nesmaoui R, Mbarek M, Lazaar M. 2023. Shapley values for explaining the black box nature of machine learning model clustering. Procedia Comput. Sci. 220:806–811. https://doi.org/10.1016/j.procs.2023.03.107
37. Singh R, Patel M. 2023. Experimental and machine learning approaches to investigate the application of sugarcane bagasse ash as a partial replacement of fine aggregate for concrete production. J. Build. Eng. 76:107168. https://doi.org/10.1016/j.jobe.2023.107168
38. Knoben WJM, Freer JE, Woods RA. Technical note: Inherent benchmark or not? comparing nash-sutcliffe and kling-gupta efficiency scores. Hydrology and Earth System Sciences. 1-7. https://doi.org/10.5194/hess-2019-327
39. Jagadesh P, de Prado-Gil J, Silva-Monteiro N, Martínez-García R. 2023. Assessing the compressive strength of self-compacting concrete with recycled aggregates from mix ratio using machine learning approach. J. Mater. Res. Technol. 24:1483-1498. https://doi.org/10.1016/j.jmrt.2023.03.037
40. Censor Y. 1977. Applied mathematics and optimization pareto optimality in multiobjective problems. Mathematics.
41. Mohammadi Golafshani E, Kim T, Behnood A, Ngo T, Kashani A. 2024. Sustainable mix design of recycled aggregate concrete using artificial intelligence. J. Clean. Prod. 442:140994. https://doi.org/10.1016/j.jclepro.2024.140994
42. Mohtasham Moein M, Saradar A, Rahmati K, Ghasemzadeh Mousavinejad SH, Bristow J, Aramali V, Karakouzian M. 2023. Predictive models for concrete properties using machine learning and deep learning approaches: A review. J. Build. Eng. 63(A):105444. https://doi.org/10.1016/j.jobe.2022.105444
43. Dafico LCM, Barreira E, Almeida RMSF, Vicente R. 2023. Machine learning models applied to moisture assessment in building materials. Constr. Build. Mater. 405:133330. https://doi.org/10.1016/j.conbuildmat.2023.133330
44. Ahmad A, Ahmad W, Chaiyasarn K, Ostrowski KA, Aslam F, Zajdel P, Joyklad P. 2021. Prediction of geopolymer concrete compressive strength using novel machine learning algorithms. Polymers Basel. 13(19):3389. https://doi.org/10.3390/polym13193389
45. Xiao C, Qiao B, Li J, Yang Z, Ding J. 2022. Prediction of transverse reinforcement of RC columns using machine learning techniques. Adv. Civ. Eng. 2022. https://doi.org/10.1155/2022/2923069
46. Vadyala SR, Betgeri SN, Matthews JC, Matthews E. 2022. A review of physics-based machine learning in civil engineering. Results Eng. 13:100316. https://doi.org/10.1016/j.rineng.2021.100316
47. Belkin M, Hsu D, Ma S, Mandal S. 2019. Reconciling modern machine-learning practice and the classical bias–variance trade-off. Proc. Natl. Acad. Sci. USA 116:15849–15854. https://doi.org/10.1073/pnas.1903070116
48. Chandra Deka P. A primer on machine learning applications in civil engineering.
49. Singh A, Mehta PK, Kumar R. 2022. Strength and microstructure analysis of sustainable self-compacting concrete with fly ash, silica fume, and recycled minerals. Mater. Today Proc. 78(1):86–98. https://doi.org/10.1016/j.matpr.2022.11.282
50. Choudhary R, Gupta R, Alomayri T, Jain A, Nagar R. 2021. Permeation, corrosion, and drying shrinkage assessment of self-compacting high strength concrete comprising waste marble slurry and fly ash, with silica fume. Struct. 33:971-985. https://doi.org/10.1016/j.istruc.2021.05.008
51. Rajkumar PRK, Jilani S, Sudha C, Jegan M, Sundararaj JB. 2021. Flexural performance of textile reinforced concrete with hybrid fabric mesh. ARPN J. Eng. Appl. Sci. 16(20):2132–2140.