Scilight

Bulletin of Computational Intelligence
https://www.sciltp.com/journals/bci

Article

Check for updates

Explainable Hybridized Machine Learning for Prediction of Compressive Strength of Fly-Ash based Geopolymer Concrete

Ajaya Subedi [1], Subodh Subedi [2], Sabin Adhikari [3], Parth Gajjar [4] and Sagar Sapkota [5,*]

[1] Department of Civil Engineering, Pashchimanchal Campus, Institute of Engineering, Tribhuvan University, Pokhara 33700, Nepal
[2] Department of Civil Engineering, Thapathali Campus, Institute of Engineering, Tribhuvan University, Kathmandu 44600 Nepal
[3] Department of Computer Science, University of South Dakota, Vermillion, SD 57069, USA
[4] Department of Civil Engineering, Institute of Technology, Nirma University, Ahmedabad, Gujarat 382481, India
[5] Department of Computer and Information Science, International Centre for Academics (IGNOU), Kathmandu 44600, Nepal
* Correspondence: sagar@s.ica.edu.np

How To Cite: Subedi, A.; Subedi, S.; Adhikari, S.; et al. Explainable Hybridized Machine Learning for Prediction of Compressive Strength of Fly-Ash based Geopolymer Concrete. *Bulletin of Computational Intelligence* **2026**, *2*(2), 213–234. https://doi.org/10.53941/bci.2026.100012

Received: 10 January 2026
Revised: 11 June 2026
Accepted: 12 June 2026
Published: 29 June 2026

Abstract: This study utilizes robust machine learning (ML) techniques to predict compressive strength of fly-ash-based geopolymer concrete (GPC), a sustainable replacement for traditional concrete. Popular ensemble models, specifically Random Forest (RF) and Extreme Gradient Boosting (XGB), were taken as base models and were hybridized using metaheuristic algorithms (Particle Swarm Optimization (PSO) and Grey Wolf Optimizer (GWO) for hyperparameter optimization. A 5 × 5 nested cross-validation (nCV) approach was adopted, where inner folds were used for hyperparameter tuning, and outer folds for unbiased performance evaluation for the limited 273-sample dataset. The findings revealed that GWO-XGB outperformed other hybridized processes, with aggregated R^2 and RMSE of 0.9661 ± 0.011 and 3.0401 ± 0.5363, respectively, in the testing phases. The performance ranking for both the training and testing phases was: GWO-XGB > PSO-XGB > GWO-RF > PSO-RF. Further, SHAP analysis was performed on models obtained from the best-tuning process, which identified Curing Period (CTP) and Curing Temperature (CTR) as the most critical parameters influencing FA-GPC strength. The best-optimized model was also used to build a graphical user interface (GUI). This work offers a reliable framework, not only for predicting CS but also for demonstrating how feature interactions contribute to strength development, providing insights into effective ML use in concrete technology.

Keywords: geopolymer concrete; compressive strength; machine learning; metaheuristic algorithms; hyperparameter tuning; SHAP

1. Introduction

Concrete is the world's second most widely consumed material, with Ordinary Portland Cement (OPC) as its primary binder [1]. The production of OPC, however, is highly energy-intensive, with a large carbon footprint that accounts for 5–7% of global CO_2 emissions. A more sustainable approach is utilizing Supplementary Cementitious Materials (SCMs) such as fly ash (FA), ground granulated blast-furnace slag, Metakaolin, silica fume, rice husk ash, sugarcane bagasse ash, etc., as a substitute [2,3]. In addition, a promising solution lies in utilizing these aluminosilicate-rich SCMs as precursor materials with alkaline solutions to form a geopolymer binder [4] which can then be combined

with aggregates to form geopolymer concrete (GPC). By completely eliminating the need for OPC, GPC reduces CO_2 emissions by 50–80% and energy demand by nearly 60% compared to Portland cement concrete [5].

In GPC, the composition generally consists of three components: precursor materials, alkaline activators, and aggregates. Here, precursors are activated with liquid alkaline solutions through a polycondensation reaction, where silica (Si) and alumina (Al) react rapidly in an alkaline environment to form three-dimensional polymeric chain and ring structures composed of Si-O-Al-O links [6]. Among various precursors, FA, an industrial by-product of coal combustion, is widely preferred because it is readily available and cost-effective and imparts comparable mechanical strength and acid resistance [7,8]. Studies have reported that FA-based geopolymer composites generate 90% less CO_2 than OPC-based composites while offering excellent strength and durability [9]. Overall, the type of precursors and concentrations of activators ($NaOH/Na_2SiO_3$ or KOH/K_2SiO_3) have a critical influence on binder reactivity, setting time, workability, and compressive strength (CS) of GPC [10].

Given several benefits of sustainability, strength, and durability, geopolymer is established as a third-generation binder, following lime/gypsum and OPC [11]. Additionally, several experimental studies have examined the mechanical and durability properties of GPC. Traditionally, the performance of GPC is primarily established in a laboratory setting, generally through compressive strength tests, as it is a key indicator of overall performance and a crucial criterion in most design standards [12,13]. Initially, predicting compressive strength was not possible given the intricate complexity of GPC, which depends on additional factors such as precursor type, Si/Al ratio, and alkaline activator. Further, the source of precursor and ratio/concentration of activator solutions ($NaOH/Na_2SiO_3$ or KOH/K_2SiO_3) influence binder reactivity, setting time, workability, and ultimately CS of GPC [14,15]. However, in recent years, the widespread adoption of ML techniques across domains has suggested its potential for efficient prediction, offering an alternative to experimental methods.

These ML models, which though solely rely on input data without considering underlying chemical reactions, have been able to capture the complex, inextricable relationship, making them popular among researchers [16]. More specifically, ML algorithms have been widely used for predicting concrete strength, interfacial bond strength, workability, crack propagation, durability assessment, as well as for mix design optimization [17,18]. For GPC mixes as well, ML regression algorithms have been increasingly employed by researchers to develop efficient and accurate models for predicting compressive strength. Compressive strength prediction of one-part alkali-activated material enabled by interpretable machine learning [19]. Several studies have used different ML algorithms for the prediction of CS of GPC, with the most common being Random Forest (RF), Extreme Gradient Boosting (XGB), Support Vector Machines (SVMs), AdaBoost (ADB), and Artificial Neural Networks (ANNs) [20–22]. Table 1 provides an overview of the relevant recent literature on CS prediction in GPC.

Table 1. Summary of recent ML approaches for compressive strength prediction in geopolymer concrete.

S.N	Study	Dataset	Algorithms [a]	SCMs [b]	Best Performance
1	Rathnayaka et al. [23]	226	ANN, RF, XGB	FA	R^2 = 0.90, RMSE = 4.7 by ANN
2	Siddiq et al. [8]	1126	ANN	FA	R^2 = 0.919, RMSE = 0.0212
3	Bypour et al. [24]	157	LR, XGB	FA, GGBFS, CBA, C-Sl	R^2 = 0.86, RMSE = 5.61 by XGB
4	Arunvivek et al. [25]	360	ANN	C-Sl, CEN	R^2 = 0.98, RMSE = 0.916
5	Bypour et al. [26]	161	ADA, ET, RF, GB, DT, XGB	FA	R^2 = 0.80, RMSE = 6.60 by ADA
6	Onyelowe et al. [27]	132	GB, CN2, NB, SVM, SGD, KNN, DT, RF	FA, GGBFS	R^2 = 0.99, RMSE = 0.916 by K-NN
7	Anwar et al. [28]	563	MLR, ANN, SVR, KNN, DT, DT-BAG, DT-BOOST	FA	R^2 = 0.89, RMSE = 4.73 by ANN
8	Dodo et al. [29]	156	LR, ANN, ANN-BAG, ANN-ADA,	FA, Slag	R^2 = 0.914, RMSE = 4.75 by ADA-ANN
9	Harmaji et al. [5]	156	MLPR, VR, XGB	FA	R^2 = 0.89, RMSE* = 0.084 by XGB
10	Golafshani et al. [30]	314	DT, RF, ET, ADA, GB, XGB, CB, CATB, LGBM	FA, Sl-C	R^2 = 0.9550, RMSE = 4.7515 by XGB
11	Eftekhar et al. [31]	235	GB, RF, DT, ANN, SVM	MK	R^2 = 0.983, RMSE = 2.218 by GB
12	Jiang et al. [32]	1136	PR, GP, ENS	FA	R^2 = 0.93, RMSE = 1.81 by ENS
13	Nazar et al. [33]	245	ANN, GEP, ANFIS	FA	R^2 = 0.94, RMSE = 0.07 by ANN
14	Ansari et al. [12]	154	LR, ANN, ADA	FA	R^2 = 0.944, RMSE = 2.506 by ADA
15	Sharma et al. [34]	147	LR, Lasso, RR	FA, GGBFS	R^2 = 0.80, RMSE = 2.3.97 by LR
16	Cao et al. [6]	151	SVM, MLP, XGB	FA	R^2 = 0.98, RMSE = 1.78 by XGB

[a] LR—Linear Regression; Lasso—Lasso Regression; RR—Ridge Regression; VR—Voting Regressor; BAG—Bagging Regressor; MLPR—Multilayer Perceptron Regressor; DT—Decision tree; DT-BAG—Decision Tree with Bagging; DT-

BOOST—Decision Tree with Boosting; ET—Extremely Randomized Trees; RF—Radom Forest; ADA—AdaBoost; CATB—Cat Boost; GB—Gradient Boosting; XGB—Extreme Gradient Boosting; LGBM—Light Gradient Boosting Machine; SVM—Support Vector Machine; SVR—support vector regression; ANN—Artificial Neural Network; ANN-BAG—ANN with Bagging; ANN-ADA—ANN with AdaBoost; CN2—Rule-based induction algorithm; NB—Naive Bayes; ANFIS—adaptive neuro-fuzzy inference system; GP—Genetic Programming; PR—Polynomial Regression; ENS—Ensemble of PR and GP; GEP—gene expression programming; [b] SCM—Supplementary Cementitious Materials; FA—Fly Ash; GGBFS—Ground Granulated Blast Furnace Slag; CGA—Coal Bottom Ash; C-Sl—Copper Slag; RHA—Rice Husk Ash; MK—Metakaolin; Sl-C—Slag Cement, CEN—Cenosphere.

RF and XGB, though they are conceptually simple tree-based algorithms, are capable of handling datasets of all sizes and model both linear and nonlinear relationships effectively, making them widely used in literature [20]. Notably, recent studies have focused on developing hybrid models, particularly by utilizing nature-inspired MHA to increase the prediction accuracy [35]. Asteris et al. [36] demonstrated that combining outputs of multiple surrogate ML models through hybrid ensembling significantly improved CS prediction accuracy. The advantage of using MHA over traditional approaches, such as grid or random search, for optimization is that these algorithms intelligently explore the hyperparameter space by mimicking population-based search behavior, potentially discovering optimal configurations that traditional methods might miss [37]. Various MHAs like Particle Swarm Optimization (PSO), Grey Wolf Optimizers (GWO), Whale Optimization [38], Leopard Seal Optimization [39], fuzzy-based optimization approaches [40], and Imperialist Competitive Algorithm [41] have been used by researchers for different concrete types. Sapkota et al. [42] investigated algorithms like water strider optimization and the Aquila optimizer algorithm for FA-GPC concrete with an R^2 value of 0.943. Khatoon et al. [43] utilized the Giant Armadillo algorithm on RF and XGB base models to predict the strength of fiber-reinforced GPC. They discovered that the hybridized XGB model exhibited high accuracy of $R^2 = 0.985$, RMSE = 0.029 MPA, closely matching experimental results. Philip et al. [44] utilized PSO with an XGB model on CS prediction of GPC concrete and achieved a superior performance with an R^2 value of 0.97. Similarly, Paudel et al. [45] demonstrated an improvement in CS prediction through a hybrid XGB-Walrus Optimization Algorithm. These studies have all highlighted that using metaheuristic optimizers to base models significantly improves their predictive performance.

Despite the growing use of ML in CS prediction for geopolymer concrete, several limitations persist in current knowledge. Existing works on fly ash-based geopolymer concrete (FA-GPC) often rely on default ML model configuration, with limited hyperparameter optimization and model explainability. While MHA models have been applied to concrete strength prediction, there is relatively little literature on FA-GPC, and even fewer studies with comprehensive explainability analysis. To address these gaps, this study proposes a hybrid predictive framework combining Random Forest (RF) and Extreme Gradient Boosting (XGB) with swarm-based metaheuristic algorithms, Particle Swarm Optimization (PSO) and Grey Wolf Optimizer (GWO) for accurate CS prediction of FA-GPC. A nCV (5 × 5) framework is considered with MHA hybridized modelling to provide optimal models that show robust, unbiased, and higher performance even in lower cardinal datasets. The paper is structured as follows in subsequent sections: The critical research gaps and the study's significance are identified in Section 2. Section 3 provides a step-by-step overview of the research methodology, covering dataset preprocessing, ML techniques, and model building and evaluation. It is followed by Section 4, which presents the model performance and results, SHAP analysis, and GUI development. Finally, Section 5 concludes with a discussion of the conclusions, limitations, and recommendations.

2. Research Significance

Several studies have demonstrated the potential of ML approaches; however, limited experimental datasets, biased evaluation procedures, and variation in the constituents in the mixes facilitate further investigation on using robust ML in concrete domains. Regarding previous studies, most of the papers only emphasize prediction using default models, where the tuning process for an optimal hyperparameter is rarely discussed in existing literature. Further, previous studies only focus on the average performance of models, excluding model explainability to interpret how a model works to achieve such performance, and the roles of each feature to reach a prediction to the targeted value. Additionally, studies with small or literature-based datasets do not provide a strong evaluation protocol of data splitting, and random train-test split is performed to show a biased and optimistic performance of models.

Such existing research voids are to be addressed by producing an unbiased model with a stronger evaluation protocol to authenticate the performance of the final model, and also explaining how such models work to reach a prediction through interacting with available input features. This research utilizes a nCV (5 × 5) framework and uses hybridized ensemble models tuned with metaheuristic optimizers to build a strong framework for authenticating the tuning process. The best tuning process is selected to perform SHAP analysis, which uses

cooperative game theory to understand the interactive and contributive role of each input feature to reach a prediction. Further, the models obtained from the best tuning process are integrated to build a deployable model using a GUI as an interface.

3. Research Methodology

The process flow diagram of the present study is depicted in Figure 1. The entire process is subdivided into different stages: Data Collection and Preprocessing, data splitting, Model Building and Hybridization, Performance Evaluation, Model Explanation and Interpretation and GUI Development.

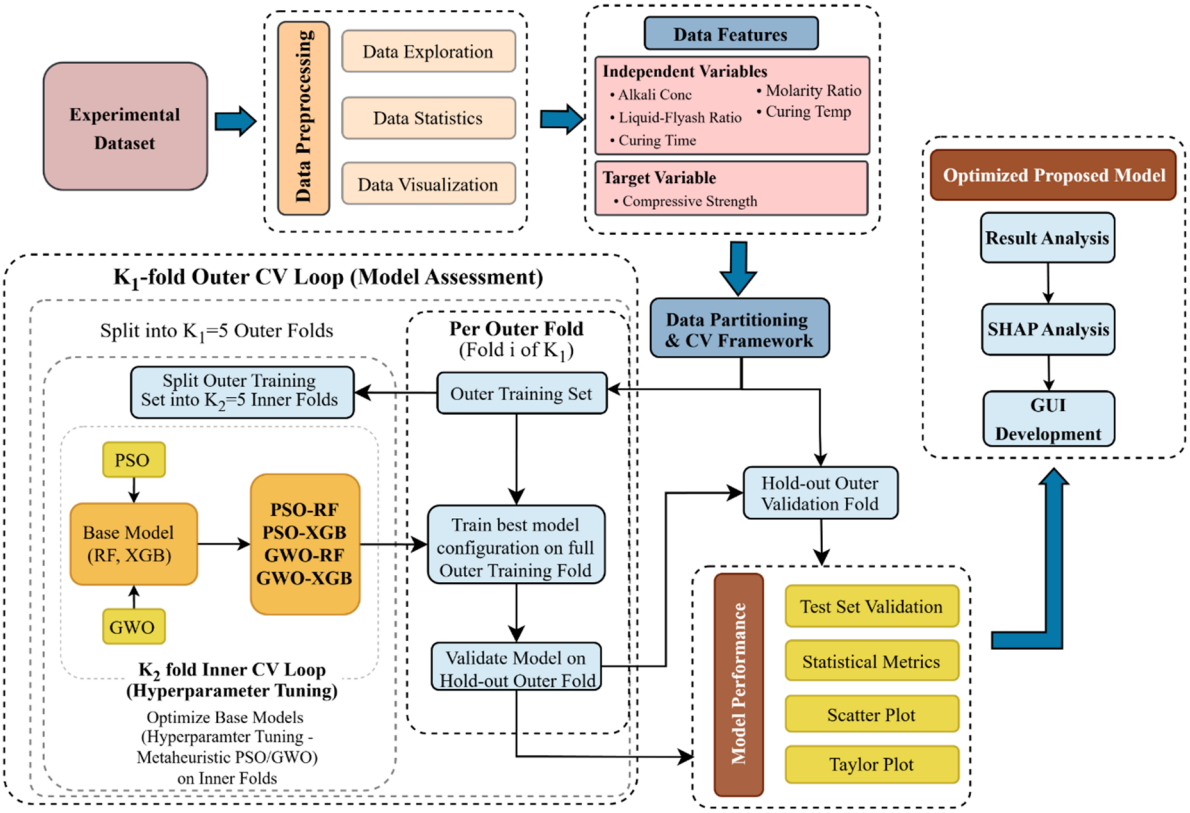

Figure 1. Overview of Methodology.

3.1. Data Description and Analysis

The initial stage of the modeling process starts with data collection. Appropriate and balanced datasets are crucial for developing models, resulting in optimal output prediction. In the current research, 273 data are collected from similar experimental studies performed by same authors (Ling et al. [46,47]). The dataset is comprised of a total of five input parameters: Alkaline concentration (AC, kg), Molarity Ratio (MR, mol/L), Liquid-to-Fly Ash Mass Ratio (LFR), Curing Temperature (CTR, °C), Curing Period (CTP, days). Compressive strength (CS) is taken as targeted variable in this study. Table 2 shows a summary of statistical analysis of all input datasets, with indicators such as: Minimum value (Min), First quartile (Q_1), Median, Mean, Third quartile, Maximum value (Max), and Standard deviation (Std). Prior to model development, data preprocessing was performed systematically, where data were initially inspected for duplicates, missing values, outliers, and other inconsistencies. It was confirmed that there were no missing or duplicate observations, and outliers remained within an acceptable range. Following this, univariate and multivariate exploratory data analysis were performed through different visualization techniques to examine relationships and distributions within the dataset.

Figure 2 illustrates histogram plots with kernel density estimation (KDE) overlays with normalized density distribution of each individual feature in the respective plots. Histogram stores wide ranges of data into various bins, which indicate the concentration of data, and KDE estimates the probability density function and the probable shape of the density of given dataset. It is observed that AC, MR, LFR, CTR, and CTP possess a discrete multi-modal distribution with low cardinality data, while CTP exhibits right-skewed distribution, suggesting concentration of data in lower ranges. The targeted variable (CS) also exhibits right-skewed distribution, with a mean of 20.24 MPa and a median of 15.30 MPa indicating majority of samples in lower ranges.

Table 2. Statistical description of dataset adopted in current study.

Parameter	Unit	Min	Q_1	Median	Mean	Q_3	Max	Std
AC	Kg	15	15	20	19.96	25	25	4.11
MR	mol/L	1	1	1.5	1.5	2	2	0.41
LFR	-	0.33	0.33	0.4	0.45	0.5	0.6	0.1
CTR	°C	23	23	50	37.24	50	50	13.5
CTP	days	1	3	7	10.22	28	28	10.86
CS	MPa	1.1	6.3	15.3	20.24	28.9	76	16.95

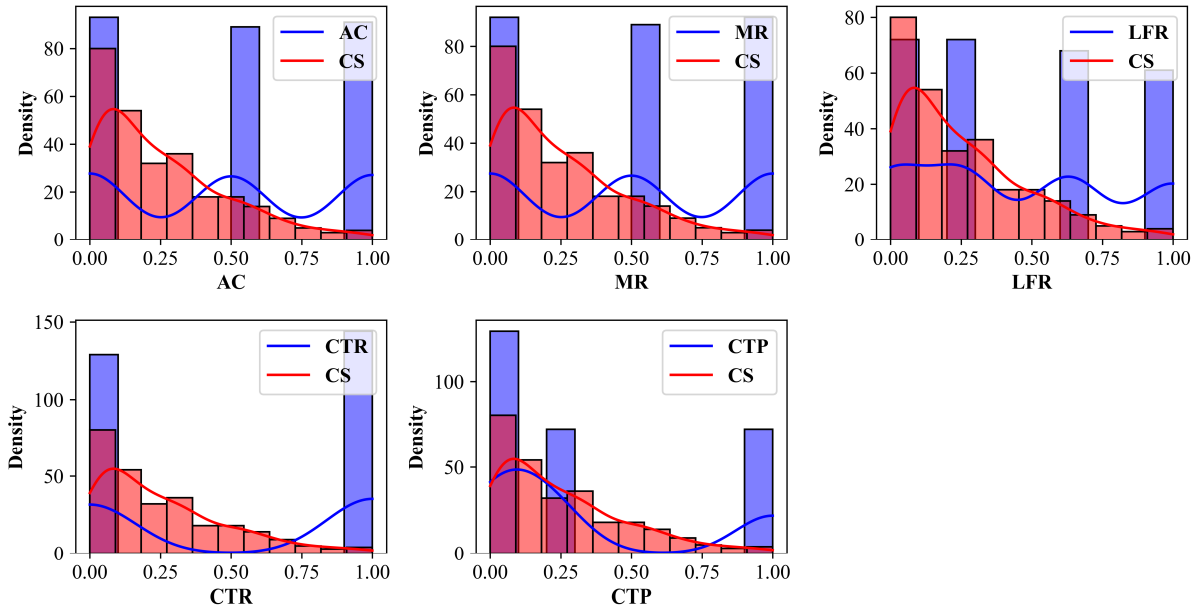

Figure 2. Histogram plots with KDE overlays.

Similarly, a heatmap in Figure 3 was generated to visualize Pearson correlation coefficient (r) among the variables. It highlights the possibility of linear dependency among the input features (AC, MR, LFR, CTR, and CTP) and target variable (CS) towards each other. Values of r lie in interval [−1, 1]; with positive, negative and zero values suggesting the direct, inverse and no correlation respectively, between variables. The heatmap points out that there is no linear dependency among the input features. AC (0.41), CTR (0.40) and CTP (0.40) show moderate positive correlation with CS, while MR (−0.38) and LFR (−0.31) indicate moderate negative correlation with CS. Additionally, Figure 4 illustrates the hexbin plot, which helps visualize the distribution pattern of input features and their contributing impact on determining output. It visualizes the data through hexagonal binning, where color intensity of bins highlights the concentration of data with darker and lighter shades indicating higher and sparse density of data. The distribution pattern of hexagonal bins of all subplots is discernible. Plots of all input features (AC, MR, LFR, CTR, and CTP) exhibit discrete clusters, which validate the lower cardinality in the dataset. Low input values of AC, CTR, and CTP show higher concentration of data at lower CS values, while the higher concentration shifts towards higher ranges of CS for higher input values, which hints the enhancement of CS at higher contents of AC, CTR, and CTP. In contradiction to aforementioned features, MR and LFR show increment in CS in lower input values, suggesting a minimum amount of these contents suffices the enhancement of the strength.

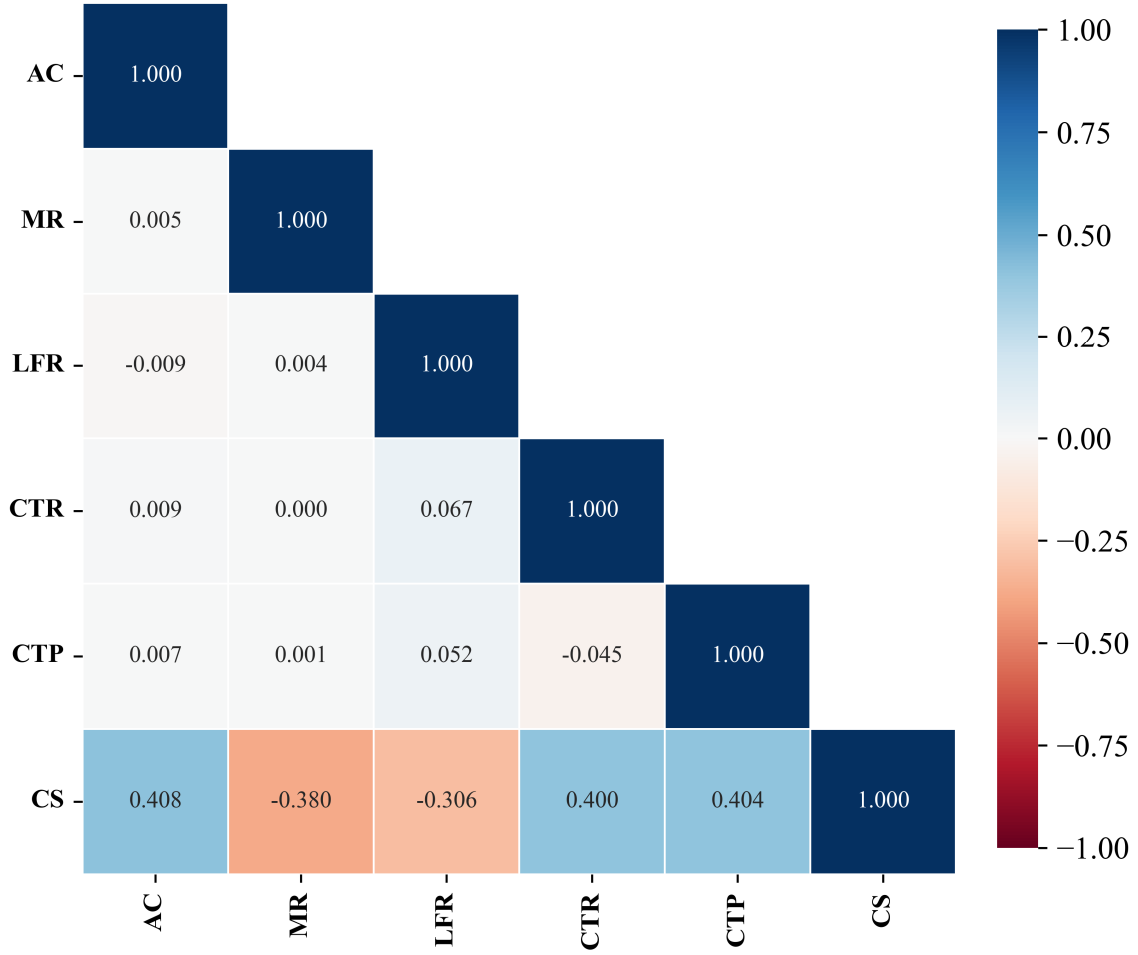

Figure 3. Heatmap matrix of Pearson Correlation coefficient between input and output features.

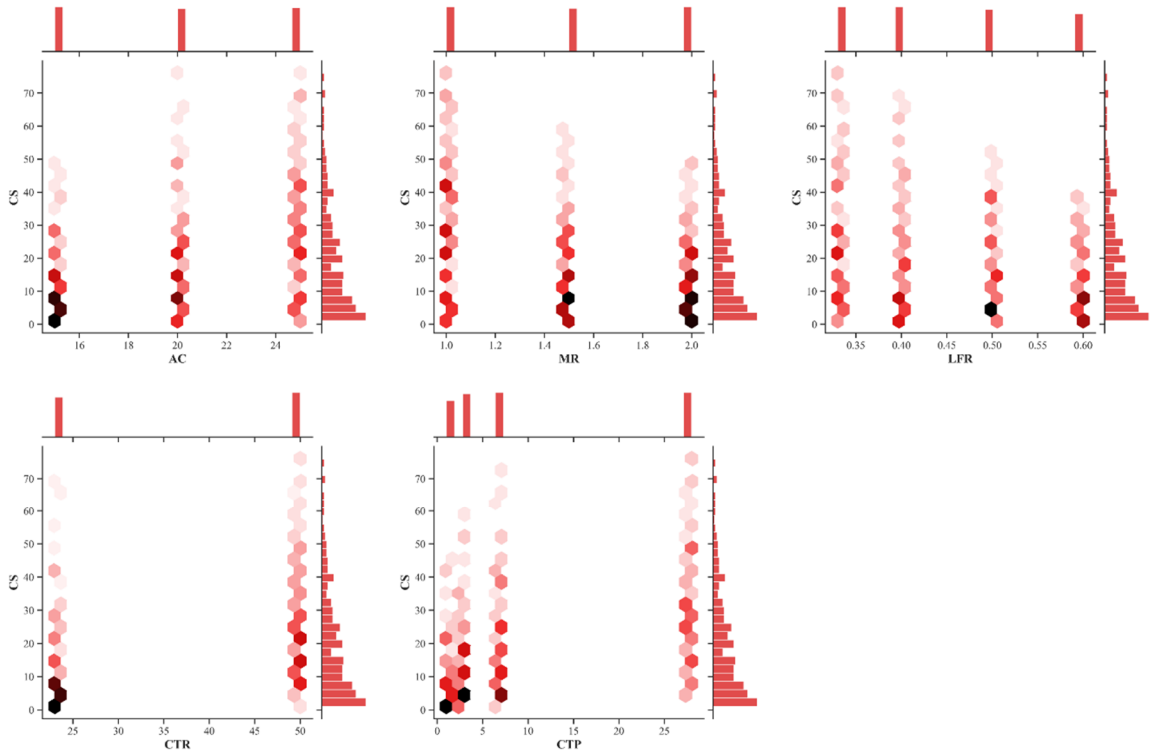

Figure 4. Hexbin plots of input features with an output variable.

3.2. Machine Learning Techniques

3.2.1. Random Forest

Random Forest (RF), a bagging-based ensemble model that combines the concepts of bootstrapping and aggregation [48], is a popular model utilized in confronting both classification and regression-based problems. It is highly recognized for its predictive power, flexibility, and robustness to noise. RF operates on three key phases during the construction of the model [49]. First, the model randomly bootstraps the original dataset D to form sample datasets D_i, while excluding one-third of the data as out-of-bag (OOB) data. Each decision tree h_i is then trained on its corresponding D_i, where data are recursively split at each node using randomly selected features that best reduce prediction error. The predicted results of all decision trees are aggregated to build a unified model, whose performance is finally evaluated on OOB data for accuracy assessment. This final aggregated output of all decision trees is taken as the final prediction, which can be expressed mathematically as in Equation (1) [50]. Because of employing various numbers of trees on particular sets of data, this model shows better performance on larger datasets. The model architecture of the RF model is demonstrated in Figure 5a.

$$H(x) = \frac{1}{K}\sum_{i=1}^{K}\left(h_i(x, \theta_k)\right) \tag{1}$$

where, $H(x)$ denotes the predicted value, h_i represents an individual decision tree, x is the independent variable, θ_k refers to an independent and identically distributed random vector, and K indicates the total number of decision trees.

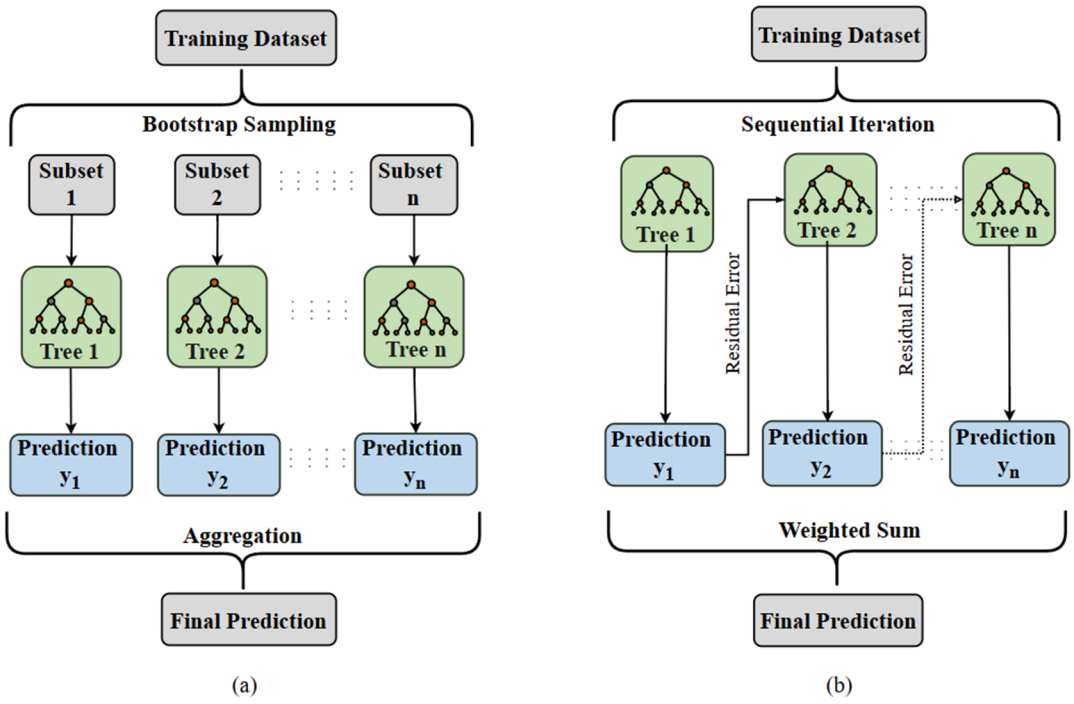

Figure 5. Model architecture (**a**) RF model (Bagging) and (**b**) XGB model (Boosting).

3.2.2. Extreme Gradient Boosting

Extreme Gradient Boosting (XGB) is well established ensemble model, which works on the concepts of gradient boosting's architecture, and performs sequential iteration for optimizing [49]; thus, making it efficient in managing larger datasets, and in handling complex prediction tasks. XGB builds the model in a sequential manner, where errors of weak preceding decision trees are considered during the construction of subsequent additive trees; this process is termed gradient boosting. The main objective of gradient boosting is to rectify the errors obtained from preceding models for maximizing the objective function of additive decision trees [51]. This results in higher accuracy with minimized risks of overfitting. Mathematically, the optimal tree at each step is found by minimizing the regularized objective function, as shown in Equations (2) and (3) [52]. The framework of the XGB model is demonstrated in Figure 5b.

$$L(\theta) = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{j=1}^{k} \Omega(f_i) \tag{2}$$

$$\Omega(\mathrm{f}) = \gamma T + \frac{1}{2}\lambda \|w\|^2 \tag{3}$$

Here, l represents a differentiable convex loss function which evaluates the difference of predicted ($\hat{y}_i$) and real (y_i) values, Ω is used as regression tree functions for penalizing complexity of model, T denotes total number of leaves within the tree, w is the associated scores of those leaves, and g evaluates the minimization of gain obtained in loss reduction for splitting an internal node.

3.2.3. Metaheuristic Algorithm

Grey Wolf Optimizer (GWO)

Grey Wolf optimizer (GWO) is a widely acclaimed metaheuristic algorithm, which is inspired by the social hierarchy and hunting techniques of grey wolves [53].The social dominancy of the grey wolves in the pack starting from alpha (α), beta (β), gamma (γ), and omega (ω) are emulated into fittest, second best, third best and remainder candidate, respectively, in GWO. The hunting strategy of the pack is embedded into GWO in three distinguished stages: identifying, encircling, hunting, and attacking of the prey. Foremost, the position of candidates of a population is initialized into a dimensional vector matrix. Then, the strategic hunting procedure is followed for updating values, which is mathematically expressed from Equation (4).

$$\bar{X}_{(m+1)} = \frac{1}{3}\sum_{i}^{3} \vec{X}_i - \vec{A} \cdot \left|\vec{C} \cdot \vec{X}_{i(m)} - \vec{X}_{(m)}\right|, \;\; i = \{\alpha, \beta, and\ \gamma\} \tag{4}$$

where $\bar{X}$ is the vector of the position, m refers to the current iteration, and $\vec{A}$ and $\vec{C}$ are the coefficient vectors, elaborated further in Equation (5), along with $\vec{a}$ (linearly decreased component from 2 to 0). $\vec{r}_1$ and $\vec{r}_2$ are random vectors. The value of $\vec{A}$ signifies both exploration ($|A|$ >1) as well as exploitation ($|A|$ <1) of search space by GWO algorithm.

$$\vec{A} = 2\vec{a} \cdot \vec{r}_1 - \vec{r}_1 \;, \qquad \vec{C} = 2 \cdot \vec{r}_2 \tag{5}$$

Particle Swarm Optimization (PSO)

Particle swarm optimization (PSO) [54] is a popular stochastic nature-inspired optimizer that is based on the movement of an individual specimen in bird flocking or fish school. The position of each of the population (X_i) is initialized in D dimension with random values, along with the velocities (v_i), as expressed in Equation (6). The newer position is determined after updating values of velocities [55,56] from a preceding population, as expressed in Equation (7). The velocity update of each particle is influenced by the global best solution (G_{best}) and the individual's best solution (P_{best}), that are also needed to be updated in each iteration.

$$X_i = \{x_{i1}, x_{i2}, \cdots, x_{id}, \cdots x_{iD}\}, \qquad v_i = \{x_{i1}, x_{i2}, \cdots, x_{id}, \cdots x_{iD}\} \tag{6}$$

$$v_{i(t+1)} = w_t \cdot v_{i(t)} + c_1 \cdot rand \cdot (P_{best} - x_{iD}) + c_2 \cdot (G_{best} - x_{iD}) \tag{7}$$

$$w_t = \left\{\frac{(t_{max} - t)^n}{(t_{max})^n}\right\}(w_i - w_f) + w_f \tag{8}$$

where, t is the current iteration, (t + 1) implies the succeeding iteration, t_{max} denotes max iteration, w_t denotes the inertial weight factor, $w_f$and w_i are final initial inertial weight factor, n ranges from 0.9 to 1.3, c_1 denotes cognitive coefficient factor, and c_2 represents the social coefficient factor. The values of c_1 and c_2 are responsible for proportioning the exploration and exploitation in the given search space.

3.3. Model Building and Hybridization

A model exhibiting commendable performance in training but fails to generalize unseen datasets are considered an overfitted model [57]. To mitigate overfitting issues and enhance the generalization of the model, k-fold cross validation [58] technique is often implemented, where the dataset is divided into k subsets, k-1 of which are used for training and the remaining subset for validation. However, standard k-fold CV can still

introduce bias in model selection. Nested cross-validation (nCV) addresses this limitation by using two loop level (inner and outer loop), where inner CV is performed for choosing optimal model through hyperparameter tuning, and outer CV is essentially used for assessing performance of the selected models. Hyperparameter tuning, the process of optimizing model hyperparameters for faster convergence and higher predictive power [59], combined with nCV, effectively reduces overfitting and ensures robust generalization. This study employed both techniques for developing effective optimized ML models for CS prediction.

This study performs two-level loops of nCV, with both the inner and outer loops comprising five folds. First, the entire dataset was split into unique five outer folds, with each fold having 20% data from the total dataset. The folds were allocated to a separate and unique locations to avoid any data leakage within the procedure. At the first loop, the first fold is assigned as a testing set (20%), and the remaining four folds are accumulated into a training set (80%). Further, on the training set, an inner 5-fold cross-validation is performed to tune the hyperparameters using MHA. In the current study, hybridized models were developed by integrating MHA (PSO and GWO) algorithms with base models (RF and XGB) for optimizing their model hyperparameters. The optimal model is then selected from the iterative process. The same process is followed for all outer folds, resulting in five sets of 80–20 split of training and testing data, with each set having a distinct split of data as the testing set. Since performance is measured in all outer folds, uncertainty of model performance due to random splitting of testing data is eliminated in this procedure. The mean performance scores and their deviations from models of all five sets are investigated to validate the authenticity of the models. Such scores play a decisive role in choosing the best hybridized process, which can produce a robust model resulting in true and unbiased estimation of error from the provided lower cardinal dataset. It is crucial to understand that nCV is not performed for selecting the optimal tuned model for final production, but for assessing the unbiased performance that a hybridized model can produce from the given tuning process.

Figure 6 demonstrates the visual representation of the nCV procedure used in this paper, and Figure 7 depicts the hybridization process of base models through GWO and PSO algorithms on a single outer loop, as applied in this study for hyperparameter tuning. For optimization, parameters; n_estimators, max_depth, max_features, and min_samples_split were used for RF. Similarly, parameters such as n_estimators, max_depth, learning_rate, subsample, reg_alpha, and reg_lambda were used for tuning in XGB. Both MHA (GWO and PSO) comprise common parameters such as n_pop, epoch, and dim, which determine the size of population, maximum number of iterations, and dimensional search space, respectively. Furthermore, inherent MHA coefficients of PSO, such as initial weight factor (w_i = 0.8), final weight factor (w_f = 0.4), n = 1, cognitive coefficient (c1 = 1.7), and social coefficient (c2 = 1.7), were also implemented. It is also imperative to select the best sets of hyperparameters for the tuning process to minimize the dimensionality of the searching space for achieving computational efficiency without compromising the model efficacy and convergence. Table 3 summarizes all the hyperparameters used, their search range, and the best sets of optimized values obtained from the MHA framework for each hybridized process**.**

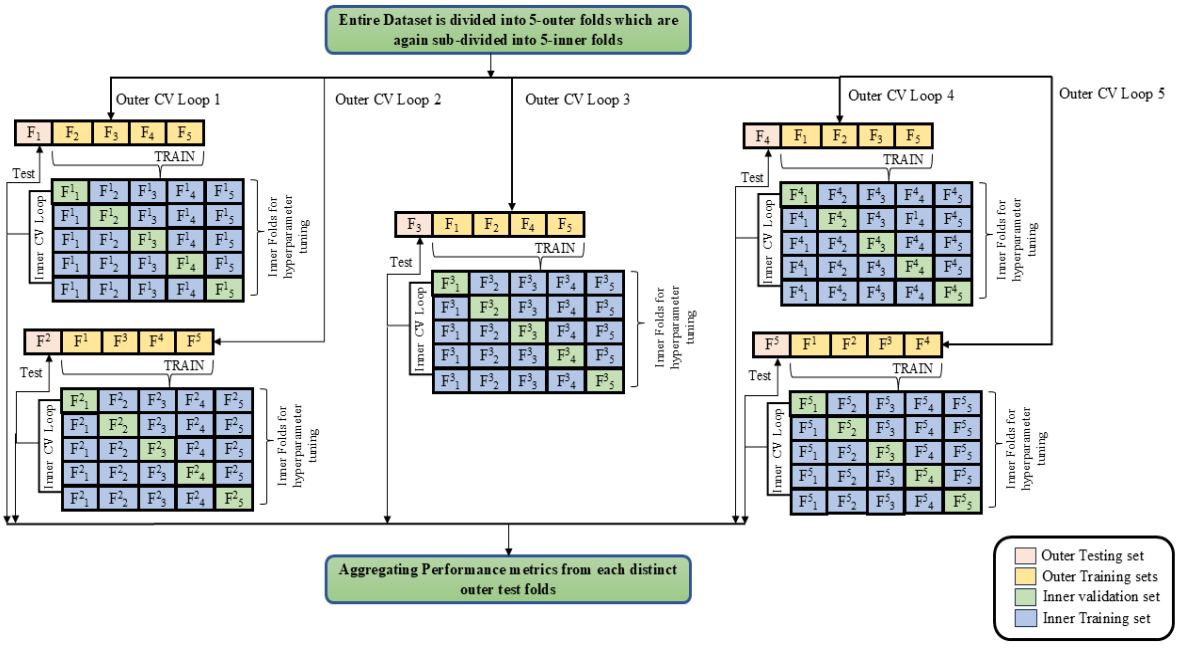

Figure 6. Visual Representation of nested Cross fold validation procedure.

Table 3. Optimized hyperparameters adopted by hybridized ML models.

Models	Hyperparameters	Bests Set of Optimization Value			MHA Parameter
		Range	From GWO	From PSO	
RF	n_estimators	[50–250]	229	221	n_pop = 25 epochs = 50 dim = 4
	max_depth	[2–25]	22	9	
	max_features	[0.1–1.0]	0.98	0.95	
	min_samples_split	[2–10]	2	2	
XGB	n_estimators	[100–500]	318	478	n_pop = 25 epochs = 50 dim = 6
	max_depth	[2–30]	18	24	
	learning_rate	[0.01–1]	0.10853	0.02446	
	subsample	[0.1–1]	0.41283	0.34048	
	reg_alpha	[0.001–1]	0.10185	0.24700	
	reg_lambda	[0.001–1]	0.27081	0.11294	

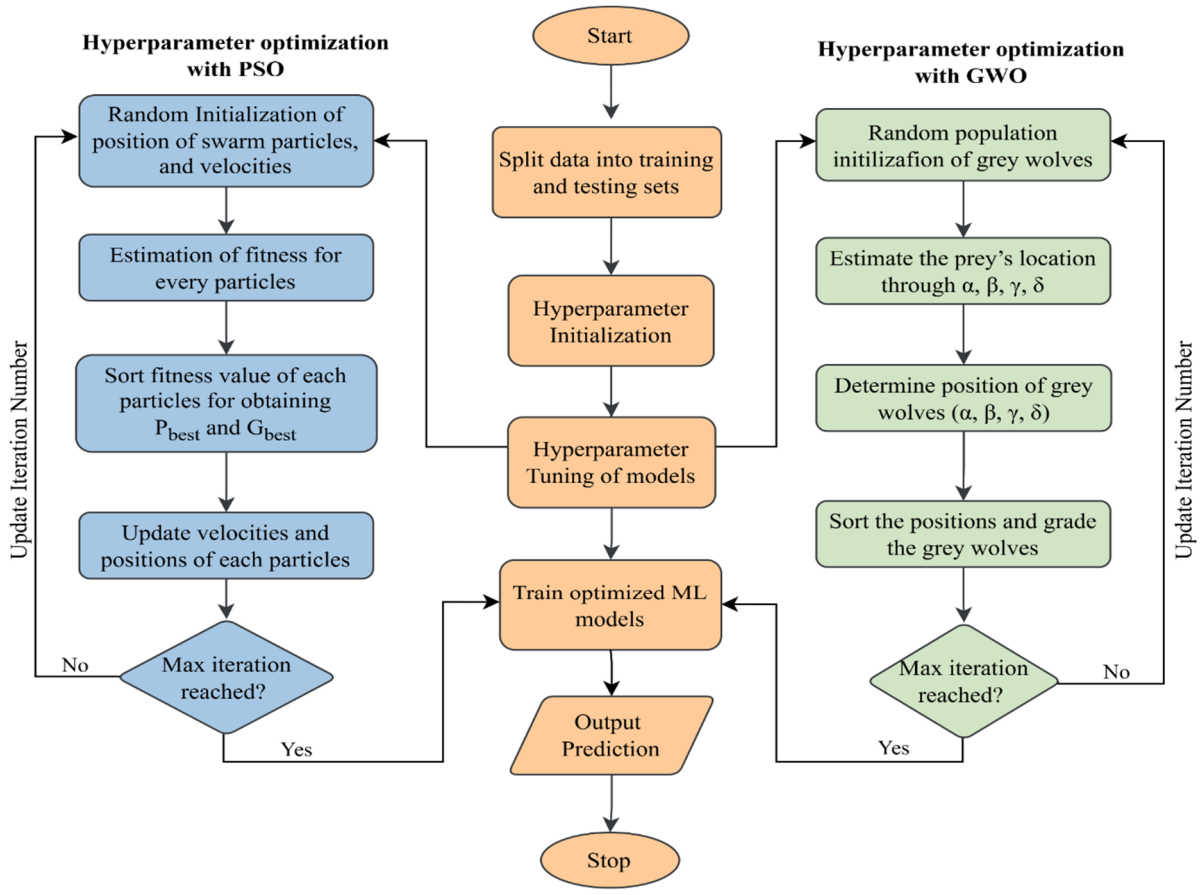

Figure 7. Framework of hybridization of base models.

3.4. Model Assessment Criteria

3.4.1. Performance Metrics

It is crucial for every model to validate its performance on both training and testing sets to substantiate the practicality and scientific validity of the outcome [60]. A model is constructed on the given sets of training datasets, and then the generated model is tested on the unseen data to ascertain its generalization ability. The performance of the model is validated through the performance metrics, which emphasize estimating the performance based on the model's generalization skills, predictive power, and robustness [61]. The current study utilizes eight metrics, as shown in Equations (9)–(18), for comprehending the efficacy of the model. The selected statistical parameters are Coefficient of Determination (R^2), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Performance Index factor (PIF), Willmott's Index of Agreement (WI) and Kling–Gupta efficiency (KGE). Each metric has its own significance. For instance, R^2 determines the proportion of variance in the model, whereas RMSE and MAE indicate the error in the model's performance. Also, if MAE is larger than RMSE, it suggests the model has predicted with a few larger errors, which leads to a plummeting of the predictive power of the model. Further, indicators like PIF, WI, and KGE diagnose the possibility and region of model failure. For instance, KGE,

which includes correlation ratio, bias and variability ratio as determining factors showing how real data differs from predicted result, and indicates the possibility of performance inconsistencies by the model. Further, WI and PIF are more sensitive than statistical parameters like RMSE, MAE and $R^{2,}$ which delivers more authentic and true nature of the model's performance. The ideal value of R^2 (1), RMSE (0), MAE (0), PIF (0), WI (1) and KGE (1) are the targeted values of each metric, which are compared with the results obtained from each model to determine how well the model performed in both training and generalization phases.

$$R^2 = 1 - \frac{\sum_{i=1}^{m}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{m}\left(y_i - \mu_y\right)^2} \tag{9}$$

$$\text{RMSE} = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(y_i - \hat{y}_i)^2} \tag{10}$$

$$\text{MAE} = \frac{1}{m}\sum_{i=1}^{m}|y_i - \hat{y}_i| \tag{11}$$

$$\text{PIF} = \frac{\text{RRMSE}}{R + 1} \tag{12}$$

$$\text{WI} = 1 - \frac{\sum_{i=1}^{m}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{m}\left(|\hat{y}_i - \mu_y| + |y_i - \mu_y|\right)^2} \tag{13}$$

$$\text{RRMSE} = \frac{1}{|\mu_y|}\sqrt{\frac{1}{m}\sum_{i=1}^{m}(y_i - \hat{y}_i)^2} \tag{14}$$

$$\text{R} = \frac{\sum_{i=1}^{m}\left(y_i - \mu_y\right)\left(\hat{y}_i - \mu_{\hat{y}}\right)}{\sqrt{\sum_{i=1}^{m}\left(y_i - \mu_y\right)^2 \sum_{i=1}^{m}\left(\hat{y}_i - \mu_{\hat{y}}\right)^2}} \tag{15}$$

$$\text{KGE} = 1 - \sqrt{(r-1)^2 + (\beta - 1)^2 + (\gamma - 1)^2} \tag{16}$$

$$\beta = \ bias\ ratio\ = \frac{\mu_{\hat{y}}}{\mu_y} \tag{17}$$

$$\gamma = \ variability\ ratio\ = \frac{CV_{\hat{y}}}{CV_y} = \frac{\left(\frac{\sigma_{\hat{y}}}{\mu_{\hat{y}}}\right)}{\left(\frac{\sigma_y}{\mu_y}\right)} \tag{18}$$

where, m is the total number of data samples, r is correlation coefficient, CV is coefficient of variation, y_i denotes true value, $\hat{y}_i$ refers to predicted value, μ_y and $\mu_{\hat{y}}$ are the mean of actual values and predicted values, and σ_y and $\sigma_{\hat{y}}$ represent standard deviation of the actual and predicted values, respectively.

3.4.2. SHapley Additives eXplanations (SHAP)

SHAP is a versatile post-hoc interpretive tool that is based on feature explainability and provides local and global explanations. Shapley value, which serves as a foundation for SHAP analysis, is a conceptualized collaborative game theory, where criteria for measuring justifiable contribution of each players include efficiency, symmetry, additivity and null player [62]. Integration of Shapely values in ML was propagated by Lundberg and Lee [63] in the form of SHAP. It is widely used for sensitivity analysis to reveal how each input feature impacts model outputs, particularly for complex ensemble models that are hard to interpret [64]. Model interpretation is represented by SHAP values, which can be positive or negative, suggesting whether input features increase or decrease the output prediction. These values can be mathematically expressed for model function $h(y')$, as in Equation (19).

$$h(y') = \phi_0 + \sum_{i=1}^{N} \phi_i y' \quad (19)$$

where, $h(y')$ is predicted output, ϕ_0 refers the mean prediction for training datasets, and $\phi_i y'$ denotes the shapely value for a given feature '*j*'. ML models can be interpreted in SHAP using both local and global explanations; Local explanations interpret the effect of every feature for a singular instance, while overall impact of every feature is summarized by global explanations.

4. Result and Discussion

4.1. Performance Assessment of Models

The mean performance of all employed tuning processes (GWO-RF, GWO-XGB, PSO-RF, PSO-XGB) in both training and testing phases is presented in Table 4. The performance of all outer training loops and testing folds corresponding to their tuned models is shown as an averaged value with variance. The results from the training phase indicate how well the different models have learned, whereas the metrics of the testing phase give insights into the generalization capability and risks of overfitting when compared with the training phase. GWO-XGB showed higher predictive accuracy (R^2 score) in the testing (0.9662 ± 0.0111) phases. GWO-XGB is followed by PSO-XGB with R^2 of 0.9997 ± 0.0004 and 0.9652 ± 0.0116 in the training and testing phases, respectively. Though PSO-XGB outshine GWO-XGB (0.9995 ± 0.0005) in the training phase, it showed poor generalization in testing sets, suggesting higher overfitting in the respective model. Overall, the performance shown by XGB models in the training phase is commendable. The small datasets with low cardinality or the hyperparameters used in the study could be the reasons for the well-trained XGB models. On the other hand, hybridized RF models had exhibited comparatively weaker performance, with lower R^2 value in both training (GWO-RF = 0.9894 ± 0.0006 and PSO-RF = 0.9891 ± 0.0004), as well as in the testing phase (GWO-RF = 0.9226 ± 0.0262 and PSO-RF = 0.9212 ± 0.028). RMSE values, which show the mean error between actual and predicted value, were found to be lower in XGB models (GWO-XGB= 3.0402 ± 0.5364 and PSO-XGB= 3.1022 ± 0.5975) if compared with RF models (GWO-RF= 4.5580 ± 0.3945 and PSO-RF= 4.5898 ± 0.3986) in the testing phase. RMSE suggests the better predictive accuracy of GWO-XGB when compared with other models. MAE, which evaluates the magnitude of absolute average error, was lowest in GWO-XGB in the testing phase (2.2312 ± 0.3504). In both RMSE and MAE, the values of testing phases are comparatively larger than training phases, suggesting the imperfection of models to nullify the absolute error, which might be caused by overfitting of the model, or by the fluctuations and noise contained in the data. Pif, which measures the variability and correlation between predicted and actual data, was minimum for GWO-XGB (0.0760 ± 0.0119), suggesting higher correlation between prediction and real data with lesser error between them, hinting at the better performance of the XGB model on new unseen data. Similarly, WI, which evaluates the agreement of true values with predicted values, are found to be higher in XGB models (GWO-XGB = 0.9911 ± 0.003, PSO-XGB = 0.9908 ± 0.0031), pointing out that the predictions given by XGB models can follow a similar pattern to that of true values. Further, the KGE value, which combines bias, variability, and correlation into a single evaluation metric, also identifies GWO-XGB (0.9589 ± 0.0169) as an efficient model for prediction tasks.

All the mean performance metrics of each model are further illustrated in Figure 8 in the form of radar plot for visual assessment. Here, the error metrics (MAE, RMSE, Pif) are arranged in increasing order, while accuracy metrics (R^2, WI) are plotted in decreasing order. This way, the metrics were scaled in such a manner that the model closest to the center of the web represents the best-performing model. It is visually evident from the respective figure that in the testing set, GWO-XGB has the smallest polygon and is consequently the best-performing model.

Overall, the performance metrics provide insights into the performance of both hybridized XGB and RF models. The nCV framework has provided authentic results with average scores and probable variance, which offer an unbiased and real estimate of model performance. XGB models showing lesser variance and better mean scores validate them as robust predictive models. Further, XGB parameters such as subsample, reg_alpha, and reg_lambda used for mitigating overfitting risks, have been found to contributing in regularizing models for providing better generalization, which is highlighted when comparing performance scores obtained from XGB models in training and testing sets. On the other hand, RF models had shown weaker performance in both the training and testing phases. Though RF are equipped with a powerful bootstrapping technique, it failed to learn the patterns of data, leading to a weaker training model. This failure might be caused by a lack of diversity in data sampling due to a smaller dataset. However, RF models were found to be insensitive to hyperparameters; minimal changes in hyperparameters were suffice for RF models to achieve a noteworthy performance.

Table 4. Performance analysis of models during training and testing phase.

Performance Metrics	Training Set				Testing Set			
	GWO-RF	PSO-RF	GWO-XGB	PSO-XGB	GWO-RF	PSO-RF	GWO-XGB	PSO-XGB
R^2	0.9894 ± 0.0006	0.9891 ± 0.0004	0.9995 ± 0.0005	0.9997 ± 0.0004	0.9226 ± 0.0262	0.9212 ± 0.028	0.9662 ± 0.0111	0.9652 ± 0.0116
RMSE	1.7436 ± 0.0396	1.7661 ± 0.0463	0.3290 ± 0.2026	0.2421 ± 0.1701	4.5580 ± 0.3945	4.5898 ± 0.3986	3.0402 ± 0.5364	3.1022 ± 0.5975
MAE	1.2497 ± 0.0583	1.2653 ± 0.0635	0.2411 ± 0.1480	0.1872 ± 0.1242	3.4176 ± 0.3197	3.4488 ± 0.3137	2.2312 ± 0.3504	2.2364 ± 0.3959
Pif	0.0432 ± 0.0011	0.0437 ± 0.0004	0.0081 ± 0.005	0.006 ± 0.0043	0.1162 ± 0.021	0.117 ± 0.0215	0.0760 ± 0.0119	0.0772 ± 0.0117
WI	0.9972 ± 0.0002	0.9971 ± 0.0001	0.9999 ± 0.0001	0.9999 ± 0.0001	0.9774 ± 0.0076	0.9770 ± 0.0081	0.9911 ± 0.003	0.9908 ± 0.0031
KGE	0.9370 ± 0.0043	0.9370 ± 0.0032	0.9986 ± 0.0011	0.9988 ± 0.0012	0.8528 ± 0.0202	0.8520 ± 0.0177	0.9589 ± 0.0169	0.9541 ± 0.0145

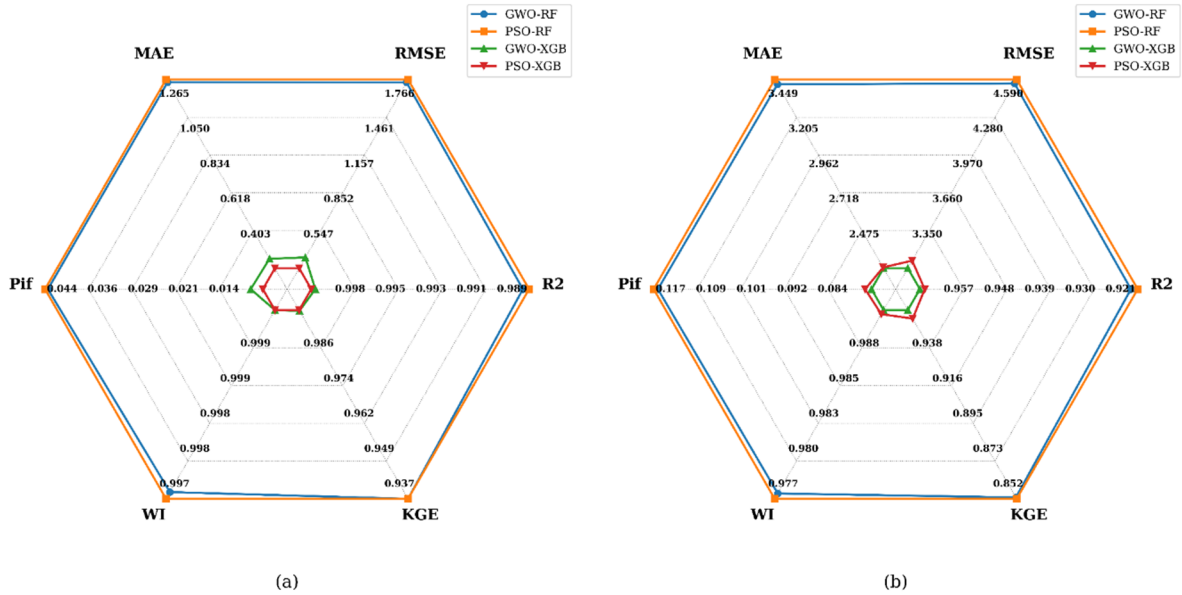

Figure 8. Radar plot showing model performance of (**a**) Training Set (**b**) Testing Set.

4.2. Comparative Performance of Hybridized Models

Visualization plots facilitate the comparison of the performance of different models. To visually assess the performance of the hybridized tuning process, a scatter plot was generated (Figure 9) to compare the predicted CS against the actual value of outer test folds. The best models obtained from the tuning process on the outer training loop are used in predicting values of their corresponding outer test folds. Under ideal conditions, the residual between actual and predicted values must be zero. However, in practicality, variability in prediction occurs and such variation can be visually assessed using 20% deviation line. The respective plot illustrates the distribution of datapoints of testing folds along the best fitted line; the distribution highlights the model's effectiveness in predicting values for new datapoints that had never been encountered by the model during the training phase. GWO-XGB shows denser clustering of data alongside the fitted line, especially in lower CS values, and only a little scattering within the deviation line, suggesting a decent generalization of a model during the testing phase. PSO-XGB also follows a similar pattern to that of GWO-XGB. On the other hand, RF models show sparse distribution, mostly in the higher regions, pointing out the lower predictive accuracy of RF models. It is clearly visible that XGB models performed better than RF models in predicting CS values, for both higher and lower values, which could be one of the reasons for XGB to achieve better performance than RF.

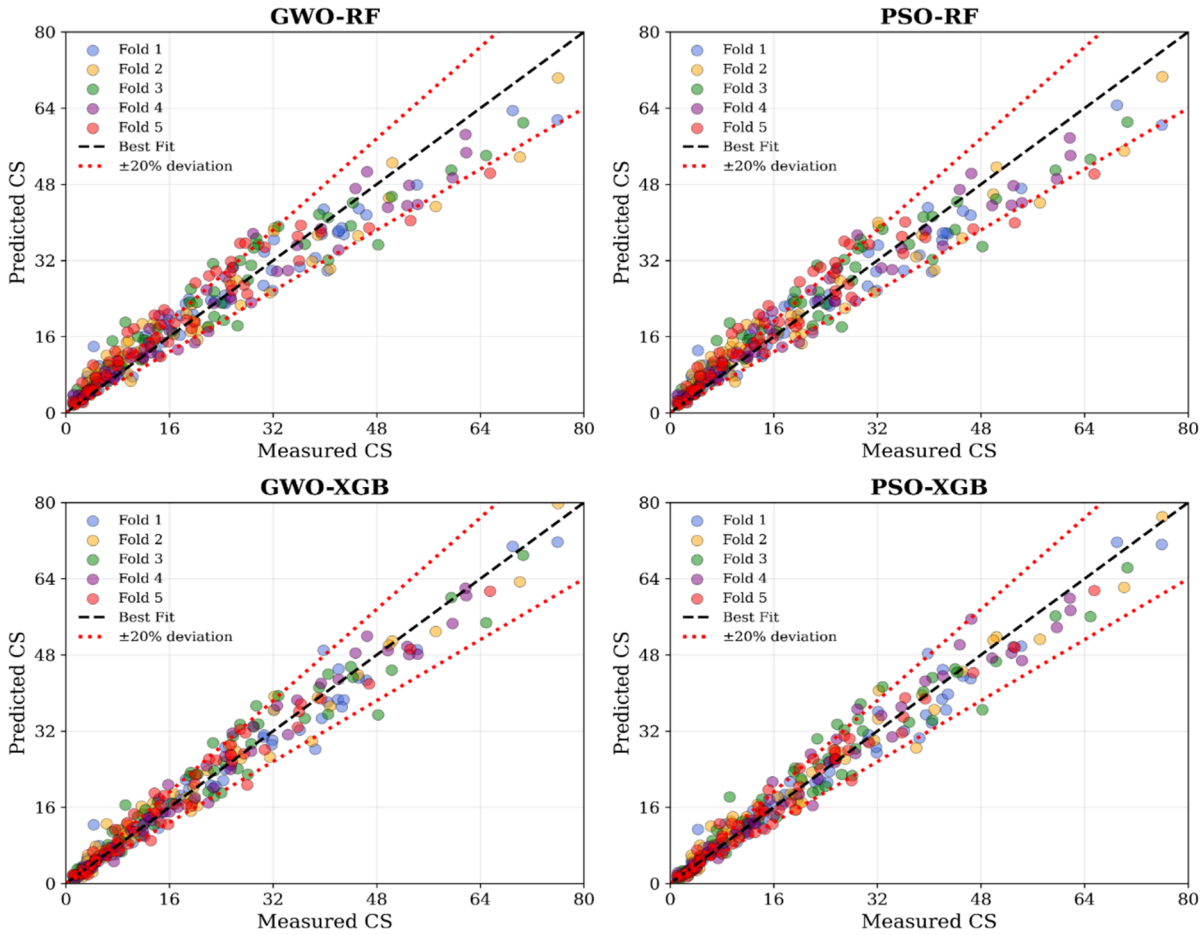

Figure 9. Scatter plot in both training and testing phase.

Further, the Taylor diagram, as presented in Figure 10, is another visualizing tool that provides a better insight on predictive accuracy of the model. Since this diagram displays three complementary metrics (RMSE, R and Std) all on a single polar coordinate plot, it serves as a better indicator to evaluate the performance of models compared to other performance indicators [65]. Here, observed data serve as the reference point, and the model predictions are plotted relative to it. The closer a point denoting model lies to this reference, the more accurate model is said to match the pattern of the real observations [66]. For the nCV framework used in this study, metrics values (RMSE, R and Std) of all training loops of an individual tuning process are averaged and are plotted. This similar averaging is also done in metrics of testing folds of each tuning process. In training phases, XGB exhibit higher training capability, as they consist of higher R, reside nearer to the reference point on the Std arc, and have minimum RMSE, compared to RF models. Similarly, in the testing phase, a noticeable decrease in correlation score and an increment in RMSE values are experienced in all processes, suggesting a significant loss of performance due to the generalization of models. However, XGB models consistently demonstrate higher predictive power as they lie in close proximity to the reference than RF models. Overall, Taylor plot and scatter plot visually confirm GWO-XGB as better model among all the tuning process suggesting the models obtained from GWO-XGB align more precisely with the real values; thus, they can encounter any unseen data with precise prediction capability. Thus, GWO-XGB models are employed for subsequent SHAP analysis, as well as for GUI program development.

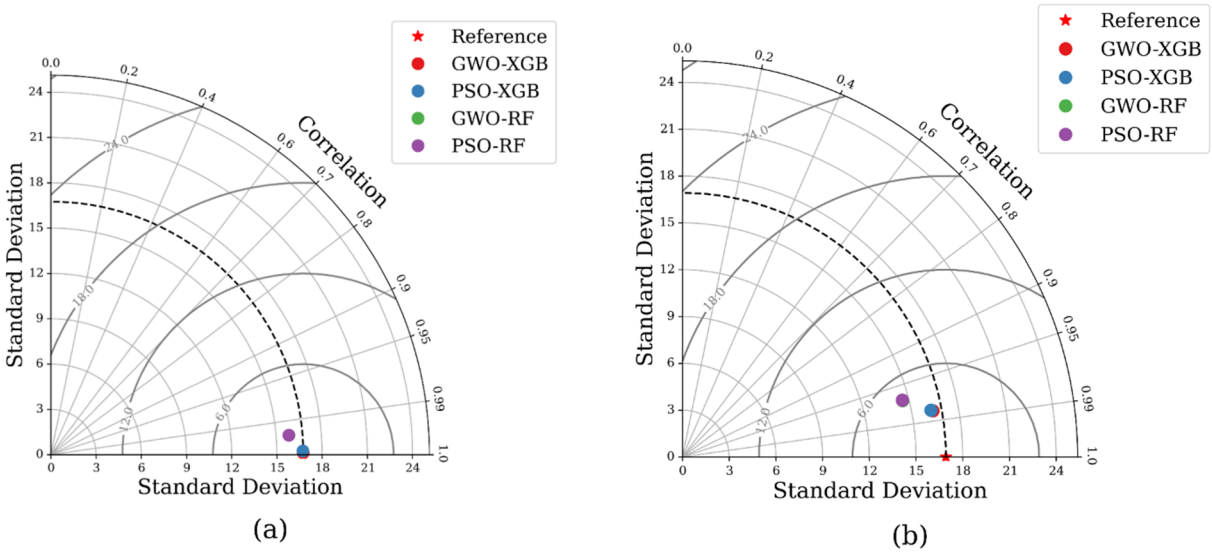

Figure 10. Taylor Diagram (**a**) Training set (**b**) Testing set.

4.3. Model Explainability Using SHAP

Tree-based models such as XGB and RF are black-box models—the inner mechanisms of such models are difficult to interpret. Therefore, different explanation and interpretation techniques are used to determine the model's interpretation of the features and their interaction, of which SHAP analysis is one. The current paper integrates SHAP analysis as a model explainer, which is performed on the best hybridized tuning process (GWO-XGB) by using outer test folds as input. The prediction of each outer fold is done using the tuned models obtained from their corresponding outer training loops. The obtained SHAP values are further visualized using different plots, as shown in Figures 11–13. Figure 11 depicts absolute mean SHAP values of each feature, which helps to sort the features based on their contribution to the output prediction. Here, CTR (8) is observed to be the most crucial feature, followed by CTP, with a mean SHAP value of 7. The values corroborate the impact of curing environment and curing age on the mechanical strength of concrete [67,68]. MR (6.4) and AC (5.8) are found to have a moderate impact on predicting CS. LFR (5.2) shows the least significance in the determination of strength in GPC. SHAP values of LFR possibly imply the significance of the selection of optimal contents of liquid and aluminosilicate precursor, as it alters the concrete matrix, retards polymerization rate, workability, and mechanical strength of the concrete [69].

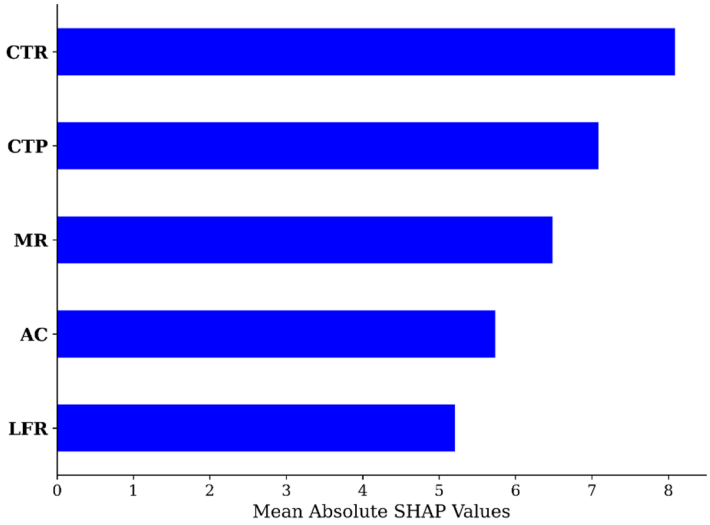

Figure 11. Mean Absolute SHAP plot.

The summary plot, presented in Figure 12 further explains the contribution of features with proper insights on the influence of each feature at a certain range. For instance, higher values of CTP show larger SHAP values,

up to +20 MPa, suggesting that a greater curing period enhances the compressive strength. Similarly, lower MR values are found to be more significant in improving the strength of concrete.

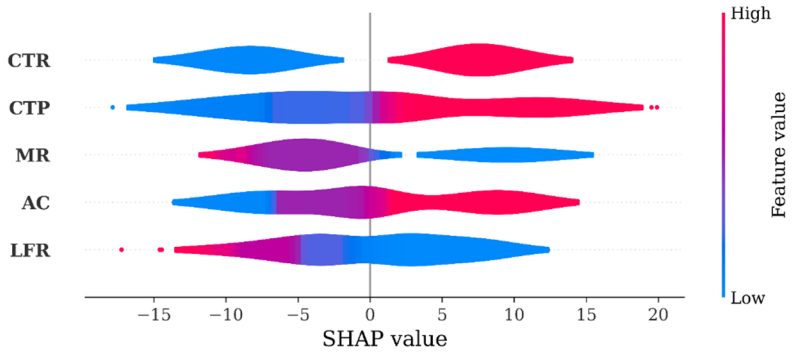

Figure 12. Summary Plot.

Furthermore, SHAP Dependence is plotted in Figure 13, which explores the feature values and their contribution in generating the final output in the form of a scatter plot. For instance, CTP exhibits nonlinear growth in SHAP values, which suggests the strength of concrete increases at extended periods, highlighting the positive impact of the longer curing duration on the strength. A longer curing period is essential in accelerating dissolution, polymerization, and reprecipitation processes of geopolymerization of concrete, which ultimately leads to enhancement in CS [70]. Similarly, CTR shows negative SHAP values (−5 to −15) at lower values and demonstrates significantly larger values at higher temperatures, suggesting the strength of the concrete improves when curing temperature increases. A higher curing temperature indeed plays a vital role in accelerating the polymerization reaction in the mix [28], which results to extensive gel formation, causing a dense matrix structure of a concrete [71] that ultimately leads to higher concrete strength. The linear trend shown by AC also suggests the improvement of concrete strength at higher alkaline concentrations. Higher AC promotes rapid formation of geopolymer gels by accelerating the breakdown of aluminosilicates, which improves the strength of the geopolymer concrete [15,72]. On the other hand, LFR and MR show a decline in SHAP values at higher feature values, clearly demonstrating the negative influence of such features in determining the strength of the concrete, if they are available in higher values. The trend followed by features shown in the SHAP Dependence plot also follow the similar pattern to that of the actual dataset, which can be visualized from the hexbin plot (Figure 4), and hints that a model was able to comprehend the behaviour shown by the real datasets. With the help of SHAP, the role of each available feature can not only be comprehended but also compared with real-life experiments. SHAP further determines how well models understand the effect of each feature to predict a desirable output [73,74]. Overall, SHAP provides a good insight into feature optimization as well as model interpretation to reach a desirable outcome.

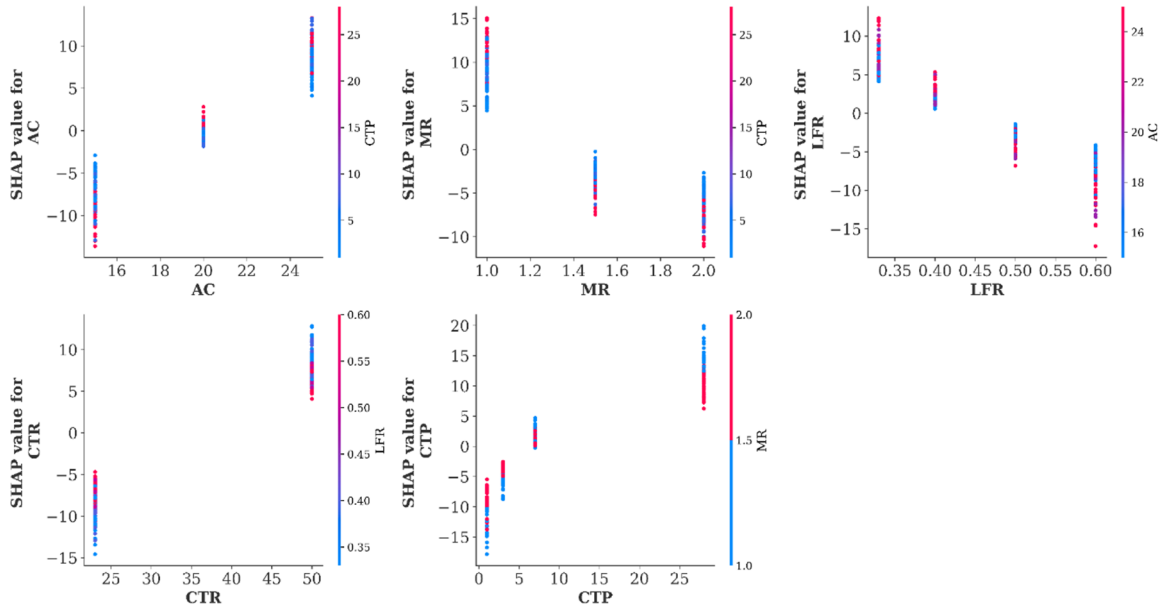

Figure 13. SHAP Dependence Plot.

5. Graphical User interface (GUI) Development

An interactive graphical user interface (GUI) helps create a visual platform for users to interact with computational ML models without requiring programming expertise. The present study also utilized a GUI tool, facilitating users in predicting CS of FA-GPC by varying input constituents. The interface was developed utilizing Python modules such as tkinter, customtkinter and ttkbootstrap, and was built on a pipeline integrating models obtained from the best tuning process (GWO-XGB) among the four models. The visual interface of the application as presented in Figure 14, has labeled input fields for predicting CS, along with their range of values it takes. Furthermore, utilizing the same pipeline, the GUI also suggests parameters using the Optuna module, where users can input a desirable strength to generate a possible value of constituents that could reach a targeted strength. Deploying GUI enables researchers and practitioners to leverage the advanced AI/ML model efficiently, thereby bridging AI/ML techniques with construction practice.

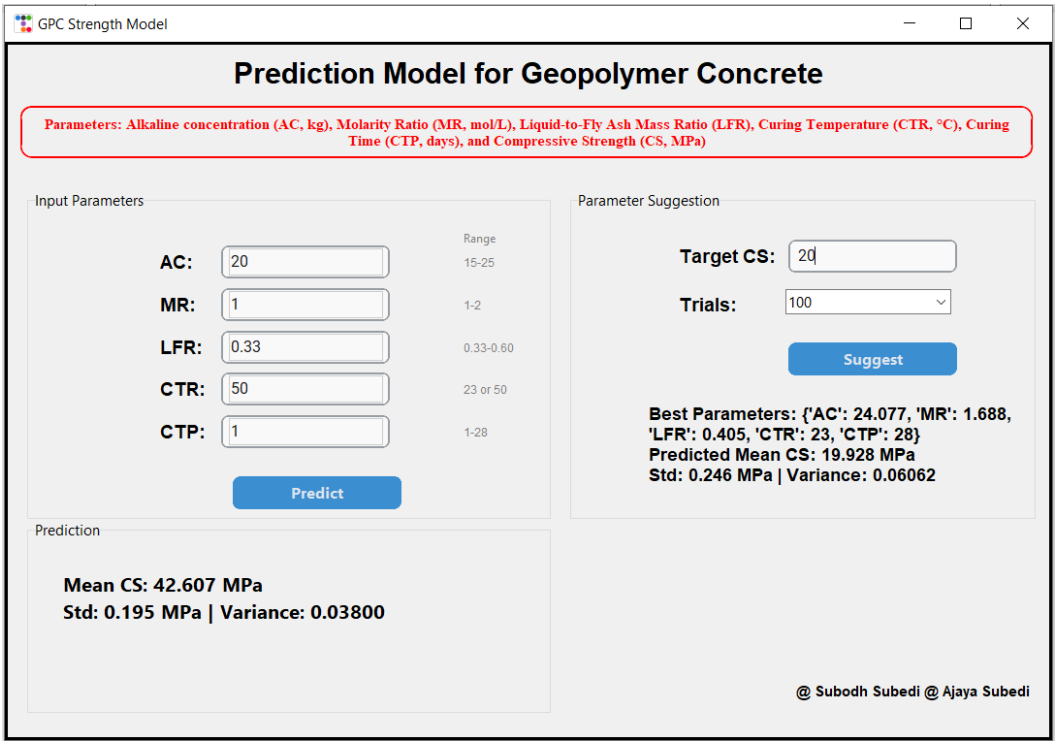

Figure 14. Graphical User Interface (GUI) development.

6. Conclusions, Limitations and Recommendations

The core of this study involved utilizing ML techniques to predict the concrete strength of FA-GPC. Following the development of ML models, performance assessment and comparative analysis of all models were performed, followed by SHAP analysis and GUI development on the best-tuned models. The key findings drawn from the study are summarized as below:

1. Among all models, GWO-XGB shows better performance, with testing accuracy of R^2 and RMSE of 0.9662 ± 0.0111 and 3.0402 ± 0.5364 respectively. PSO-XGB subsequently follows with R^2 of 0.9652 ± 0.0116 and RMSE of 3.1022 ± 0.5975. The performance of both XGB models was also consistent among all performance metrics like MAE, Pif, and WI. With WI around 0.99, Pif less than 0.08, and a lower MAE of around 2.2, both XGB models exhibited higher accuracy and reliability in prediction.
2. For this dataset, all XGB models consistently outperformed RF models on both the training and testing phases, across all statistical indices and visual assessments, highlighting their ability to generalize well to new data and higher accuracy in CS prediction.
3. SHAP-based analysis, when performed on the best tuning process (GWO-XGB), identified Curing conditions (CTP, CTR) as the most influential parameter that contributes to CS in FA-GPC. It is followed by MR, AC, and LFR, respectively. The best models (GWO-XGB) were integrated and used in developing a GUI interface for allowing constructive use of the developed models among researchers and field experts.

Though the study had achieved commendable results by incorporating hybridized ML techniques, a few noticeable limitations should also be mentioned. To begin with, a small sample size (273) is used in this study,

which does not provide comprehensive knowledge about the domain. Another insufficiency in the study is the lack of utilization of important input parameters such as the chemical and mineralogical composition of FA, aggregates, and additives. Future studies should focus on expanding the dataset for more standardized and diverse experimental results, thus incorporating additional input features of GPC mixes. Additionally, ML techniques can be employed for producing multiple outputs like CS, Direct Tensile strength, Split Tensile strength, workability, and durability to make it more comprehensive and valuable to researchers. This study is also focused on a specific domain represented by the provided datasets, which might produce unrealistic output in the developed GUI if the parameter ranges are out of the bounds of the provided datasets. Utilizing the optimized ML models for an independent external validation from other literature of FA-GPC can be done in subsequent research, which would increase the reliability of the model. Apart from this, newer and more advanced optimization algorithms can also be incorporated with various ML-based models to produce more robust hybridized models. These developments would produce more robust and novel ML techniques, making the research more valuable and promising to both researchers and industry practitioners.

Author Contributions

A.S.: Investigation, Visualization, Software, Writing—Original draft, Methodology, Writing—Review and Editing; S.S. (Subodh Subedi): Conceptualization, Investigation, ML modeling, Visualization, Software, Writing—Original draft, Methodology, ML modeling, Writing—Review and Editing. S.A.: Data Curation, Software, Writing—Review and Editing, Methodology; P.G.: Investigation, Methodology, Software, Writing—Review and Editing; S.S. (Sagar Sapkota): Conceptualization, Methodology, Supervision, Writing—Review and Editing; All authors have read and agreed to the published version of the manuscript.

Funding

This research received no external funding.

Institutional Review Board Statement

Not Applicable.

Informed Consent Statement

Not Applicable.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Conflicts of Interest

The authors declare no conflict of interest.

Use of AI and AI-Assisted Technologies

AI tools such as MS-Editor and Grammarly AI were minimally used, solely for fine-tuning the readability of the original manuscript, mainly focusing on grammatical errors, clarity and tone adjustments.

References

1. Jamaludin, N.F.A.; Muthusamy, K.; Md Jaafar, M.F.; et al. Preliminary Investigation on Spent Garnet as a Novel Supplementary Cementitious Material. *Constr. Build. Mater.* **2024**, *451*, 138789. https://doi.org/10.1016/j.conbuildmat.2024.138789.
2. Subedi, A.; Thapa, B.B.; Poudel, A.; et al. Exploring the Potential of Himalayan Giant Nettle Fiber and Supplementary Cementitious Materials for Sustainable Concrete Development. *Asian J. Civ. Eng.* **2025**, *26*, 635–648. https://doi.org/10.1007/s42107-024-01211-9.
3. Poudel, S.; Subedi, A.; Khadka, B.; et al. Study on the Mechanical Properties of Concrete with Recycled Aggregates, Glass Fibers, and Fly Ash: Exploring Environmental Waste Utilization. *Innov. Infrastruct. Solut.* **2025**, *10*, 377. https://doi.org/10.1007/s41062-025-02183-8.
4. Tayeh, B.A.; Hamada, H.M.; Almeshal, I.; et al. Durability and Mechanical Properties of Cement Concrete Comprising

Pozzolanic Materials with Alkali-Activated Binder: A Comprehensive Review. *Case Stud. Constr. Mater.* **2022**, *17*, e01429. https://doi.org/10.1016/j.cscm.2022.e01429.

5. Harmaji, A.; Kirana, M.C.; Jafari, R. Machine Learning to Predict Workability and Compressive Strength of Low- and High-Calcium Fly Ash–Based Geopolymers. *Crystals* **2024**, *14*, 830. https://doi.org/10.3390/cryst14100830.
6. Cao, R.; Fang, Z.; Jin, M.; et al. Application of Machine Learning Approaches to Predict the Strength Property of Geopolymer Concrete. *Materials* **2022**, *15*, 2400. https://doi.org/10.3390/ma15072400.
7. Chen, K.; Wu, D.; Xia, L.; et al. Geopolymer Concrete Durability Subjected to Aggressive Environments—A Review of Influence Factors and Comparison with Ordinary Portland Cement. *Constr. Build. Mater.* **2021**, *279*, 122496. https://doi.org/10.1016/j.conbuildmat.2021.122496.
8. Siddiq, M.U.; Anwar, M.K.; Almansour, F.H.; et al. AI-Driven Optimization of Fly Ash-Based Geopolymer Concrete for Sustainable High Strength and CO2 Reduction: An Application of Hybrid Taguchi–Grey–ANN Approach. *Buildings* **2025**, *15*, 2081. https://doi.org/10.3390/buildings15122081.
9. Luhar, I.; Luhar, S. A Comprehensive Review on Fly Ash-Based Geopolymer. *J. Compos. Sci.* **2022**, *6*, 219. https://doi.org/10.3390/jcs6080219.
10. Rihan, M.A.M.; Onchiri, R.O.; Gathimba, N.; et al. Mechanical and Microstructural Properties of Geopolymer Concrete Containing Fly Ash and Sugarcane Bagasse Ash. *Civ. Eng. J.* **2024**, *10*, 1292–1309. https://doi.org/10.28991/CEJ-2024-010-04-018.
11. Danish, A.; Ozbakkaloglu, T.; Ali Mosaberpanah, M.; et al. Sustainability Benefits and Commercialization Challenges and Strategies of Geopolymer Concrete: A Review. *J. Build. Eng.* **2022**, *58*, 105005. https://doi.org/10.1016/j.jobe.2022.105005.
12. Shamim Ansari, S.; Muhammad Ibrahim, S.; Danish Hasan, S. Conventional and Ensemble Machine Learning Models to Predict the Compressive Strength of Fly Ash Based Geopolymer Concrete. *Mater. Today Proc.* **2023**, *in press*. https://doi.org/10.1016/j.matpr.2023.04.393.
13. Alyami, M.; Nassar, R.U.D.; Khan, M.; et al. Estimating Compressive Strength of Concrete Containing Rice Husk Ash Using Interpretable Machine Learning-Based Models. *Case Stud. Constr. Mater.* **2024**, *20*, e02901. https://doi.org/10.1016/j.cscm.2024.e02901.
14. Assi, L.N.; Eddie Deaver, E.; Ziehl, P. Effect of Source and Particle Size Distribution on the Mechanical and Microstructural Properties of Fly Ash-Based Geopolymer Concrete. *Constr. Build. Mater.* **2018**, *167*, 372–380. https://doi.org/10.1016/j.conbuildmat.2018.01.193.
15. Rihan, M.A.M.; Alahmari, T.S.; Onchiri, R.O.; et al. Impact of Alkaline Concentration on the Mechanical Properties of Geopolymer Concrete Made up of Fly Ash and Sugarcane Bagasse Ash. *Sustainability* **2024**, *16*, 2841. https://doi.org/10.3390/su16072841.
16. Sapkota, S.C.; Shrestha, S.; Haq, M.; et al. Explainable Hybridized Ensemble Model with a Metaheuristics Algorithm for the Prediction of Compressive Strength of Ultrahigh-Performance Concrete. *J. Struct. Des. Constr. Pract.* **2026**, *31*, 04026023. https://doi.org/10.1061/JSDCCC.SCENG-1948.
17. Ji, H.; Lyu, Y.; Ying, W.; et al. Machine Learning Guided Iterative Mix Design of Geopolymer Concrete. *J. Build. Eng.* **2024**, *91*, 109710. https://doi.org/10.1016/j.jobe.2024.109710.
18. Sapkota, S.C.; Sapkota, S.; Bansal, T.; et al. Leveraging Interpretable Ensemble Machine Learning for Predicting Interfacial Bond Strength Between Normal-Strength Concrete Substrate and UHPC Overlays. *Arab. J. Sci. Eng.* **2024**, *50*, 8621–8645. https://doi.org/10.1007/S13369-024-09690-3.
19. Shah, S.F.A.; Chen, B.; Zahid, M.; et al. Compressive Strength Prediction of One-Part Alkali Activated Material Enabled by Interpretable Machine Learning. *Constr. Build. Mater.* **2022**, *360*, 129534. https://doi.org/10.1016/j.conbuildmat.2022.129534.
20. Sapkota, S.C.; Sapkota, S.; Saini, G.; et al. Computational intelligence enhances compressive strength estimation of sustainable concrete using industrial waste. *Struct. Concr.* **2026**, *27*, 1160–1183. https://doi.org/10.1002/suco.70271.
21. Mazumder, E.A.; Sapkota, S.C.; Das, S.; et al. A Generalized Explainable Approach to Predict the Hardened Properties of Self-Compacting Geopolymer Concrete Using Machine Learning Techniques. *Comput. Concr.* **2024**, *34*, 279–296. https://doi.org/10.12989/cac.2024.34.3.279.
22. Sapkota, S.C.; Saha, P.; Das, S.; et al. Prediction of the Compressive Strength of Normal Concrete Using Ensemble Machine Learning Approach. *Asian J. Civ. Eng.* **2023**, *25*, 583–596. https://doi.org/10.1007/S42107-023-00796-X.
23. Rathnayaka, M.; Wijesundara, K.; Gunasekara, C.; et al. Integrating Advanced Data Imputation Techniques and Machine Learning Models to Develop a Predictive Model for Compressive Strength of Geopolymer Concrete. *Arch. Civ. Mech. Eng.* **2025**, *25*, 201. https://doi.org/10.1007/s43452-025-01254-y.
24. Bypour, M.; Yekrangnia, M.; Kioumarsi, M. Evaluation of the Compressive Strength of Fly Ash-Based Geopolymer Concrete Using Machine Learning. In *The International Conference on Net-Zero Civil Infrastructures: Innovations in Materials, Structures, and Management Practices*; Kioumarsi, M., Shafei, B., Eds.; Springer Nature: Cham, Switzerland,

2025; Volume 237, pp. 801–811.

25. Arunvivek, G.K.; Anandaraj, S.; Kumar, P.; et al. Compressive Strength Modelling of Cenosphere and Copper Slag-Based Geopolymer Concrete Using Deep Learning Model. *Sci. Rep.* **2025**, *15*, 27849. https://doi.org/10.1038/s41598-025-13176-z.
26. Bypour, M.; Yekrangnia, M.; Kioumarsi, M. Machine Learning-Driven Optimization for Predicting Compressive Strength in Fly Ash Geopolymer Concrete. *Clean. Eng. Technol.* **2025**, *25*, 100899. https://doi.org/10.1016/j.clet.2025.100899.
27. Onyelowe, K.C.; Ebid, A.M.; Awoyera, P.; et al. Prediction and Validation of Mechanical Properties of Self-Compacting Geopolymer Concrete Using Combined Machine Learning Methods a Comparative and Suitability Assessment of the Best Analysis. *Sci. Rep.* **2025**, *15*, 6361. https://doi.org/10.1038/s41598-025-90468-4.
28. Anwar, M.K.; Qurashi, M.A.; Zhu, X.; et al. A Comparative Performance Analysis of Machine Learning Models for Compressive Strength Prediction in Fly Ash-Based Geopolymers Concrete Using Reference Data. *Case Stud. Constr. Mater.* **2025**, *22*, e04207. https://doi.org/10.1016/j.cscm.2025.e04207.
29. Dodo, Y.; Arif, K.; Alyami, M.; et al. Estimation of Compressive Strength of Waste Concrete Utilizing Fly Ash/Slag in Concrete with Interpretable Approaches: Optimization and Graphical User Interface (GUI). *Sci. Rep.* **2024**, *14*, 4598. https://doi.org/10.1038/s41598-024-54513-y.
30. Golafshani, E.; Khodadadi, N.; Ngo, T.; et al. Modelling the Compressive Strength of Geopolymer Recycled Aggregate Concrete Using Ensemble Machine Learning. *Adv. Eng. Softw.* **2024**, *191*, 103611. https://doi.org/10.1016/j.advengsoft.2024.103611.
31. Eftekhar Afzali, S.A.; Shayanfar, M.A.; Ghanooni-Bagha, M.; et al. The Use of Machine Learning Techniques to Investigate the Properties of Metakaolin-Based Geopolymer Concrete. *J. Clean. Prod.* **2024**, *446*, 141305. https://doi.org/10.1016/j.jclepro.2024.141305.
32. Jiang, P.; Zhao, D.; Jin, C.; et al. Compressive Strength Prediction and Low-Carbon Optimization of Fly Ash Geopolymer Concrete Based on Big Data and Ensemble Learning. *PLoS ONE* **2024**, *19*, e0310422. https://doi.org/10.1371/journal.pone.0310422.
33. Nazar, S.; Yang, J.; Amin, M.N.; et al. Machine Learning Interpretable-Prediction Models to Evaluate the Slump and Strength of Fly Ash-Based Geopolymer. *J. Mater. Res. Technol.* **2023**, *24*, 100–124. https://doi.org/10.1016/j.jmrt.2023.02.180.
34. Sharma, U.; Gupta, N.; Verma, M. Prediction of Compressive Strength of GGBFS and Flyash-Based Geopolymer Composite by Linear Regression, Lasso Regression, and Ridge Regression. *Asian J. Civ. Eng.* **2023**, *24*, 3399–3411. https://doi.org/10.1007/s42107-023-00721-2.
35. Hoque, M.A.; Shrestha, A.; Sapkota, S.C.; et al. Prediction of Autogenous Shrinkage in Ultra-High-Performance Concrete (UHPC) Using Hybridized Machine Learning. *Asian J. Civ. Eng.* **2025**, *26*, 649–665. https://doi.org/10.1007/s42107-024-01212-8.
36. Asteris, P.G.; Skentou, A.D.; Bardhan, A.; et al. Predicting Concrete Compressive Strength Using Hybrid Ensembling of Surrogate Machine Learning Models. *Cem. Concr. Res.* **2021**, *145*, 106449. https://doi.org/10.1016/j.cemconres.2021.106449.
37. Li, L.; Sun, W.; Gómez-Zamorano, L.Y.; et al. From Research Trend to Performance Prediction: Metaheuristic-Driven Machine Learning Optimization for Cement Pastes Containing Bio-Based Phase Change Materials. *Polymers* **2025**, *17*, 2541. https://doi.org/10.3390/polym17182541.
38. Zhao, J.; Nguyen, H.; Nguyen-Thoi, T.; et al. Improved Levenberg–Marquardt Backpropagation Neural Network by Particle Swarm and Whale Optimization Algorithms to Predict the Deflection of RC Beams. *Eng. Comput.* **2021**, *38*, 3847–3869. https://doi.org/10.1007/S00366-020-01267-6.
39. Sapkota, S.C.; Shrestha, A.; Tang, W. Metaheuristic-Optimized Machine Learning Models for Accurate Prediction of Limestone Calcined Clay Cement (LC3) Compressive Strength. *Case Stud. Constr. Mater.* **2025**, *23*, e05303. https://doi.org/10.1016/j.cscm.2025.e05303.
40. Liao, J.; Asteris, P.G.; Cavaleri, L.; et al. Novel Fuzzy-Based Optimization Approaches for the Prediction of Ultimate Axial Load of Circular Concrete-Filled Steel Tubes. *Buildings* **2021**, *11*, 629. https://doi.org/10.3390/BUILDINGS11120629.
41. Duan, J.; Asteris, P.G.; Nguyen, H.; et al. A Novel Artificial Intelligence Technique to Predict Compressive Strength of Recycled Aggregate Concrete Using ICA-XGBoost Model. *Eng. Comput.* **2020**, *37*, 3329–3346. https://doi.org/10.1007/s00366-020-01003-0.
42. Sapkota, S.C.; Dahal, D.; Yadav, A.; et al. Analyzing the Behavior of Geopolymer Concrete with Different Novel Machine-Learning Algorithms. *J. Struct. Des. Constr. Pract.* **2025**, *30*, 4025027. https://doi.org/10.1061/JSDCCC.SCENG-1724.
43. Khatoon, S.; K, K.A.; Sapkota, S.C. Experimental Insights and Hybridized Ensemble Machine Learning Validation of Fiber Reinforced Geopolymer Concrete Strength. *Asian J. Civ. Eng.* **2025**, *27*, 1289–1312. https://doi.org/10.1007/s42107-025-01562-x.

44. Philip, S.; Marakkath, N. Development and Optimization of Geopolymer Concrete with Compressive Strength Prediction Using Particle Swarm-Optimized Extreme Gradient Boosting. *Appl. Soft Comput.* **2025**, *176*, 113149. https://doi.org/10.1016/j.asoc.2025.113149.
45. Paudel, S.; Shrestha, A.; Sapkota, S.C.; et al. Explainable Machine Learning Integrated with Bio-Inspired Optimization for Predicting and Interpreting the Compressive Strength of Metakaolin-Based Geopolymers. *Asian J. Civ. Eng.* **2026**, *27*, 2981–3001. https://doi.org/10.1007/S42107-026-01652-4.
46. Ling, Y.; Wang, K.; Wang, X.; et al. Prediction of engineering properties of fly ash-based geopolymer using artificial neural networks. *Neural Comput. Appl.* **2021**, *33*, 85–105. https://doi.org/10.1007/s00521-019-04662-3.
47. Ling, Y.; Wang, K.; Wang, X.; et al. Effects of Mix Design Parameters on Heat of Geopolymerization, Set Time, and Compressive Strength of High Calcium Fly Ash Geopolymer. *Constr. Build. Mater*. **2019**, *228*, 116763. https://doi.org/10.1016/j.conbuildmat.2019.116763.
48. Mai, H.V.T.; Nguyen, T.A.; Ly, H.B.; et al. Prediction Compressive Strength of Concrete Containing GGBFS Using Random Forest Model. *Adv. Civ. Eng.* **2021**, *2021*, 6671448. https://doi.org/10.1155/2021/6671448.
49. Alsaadawi, M.M.; Elshaarawy, M.K.; Hamed, A.K. Concrete Compressive Strength Classification Using Hybrid Machine Learning Models and Interactive GUI. *Innov. Infrastruct. Solut.* **2025**, *10*, 198. https://doi.org/10.1007/s41062-025-01983-2.
50. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. https://doi.org/10.1023/A:1010933404324.
51. Sapkota, S.C.; Sapkota, S.; Saini, G. Prediction of Split Tensile Strength of Recycled Aggregate Concrete Leveraging Explainable Hybrid XGB with Optimization Algorithm. *Multiscale Multidiscip.Model. Exp. Des.* **2024**, *7*, 4343–4359. https://doi.org/10.1007/s41939-024-00480-w.
52. Sapkota, S.C.; Shrestha, A.; Haq, M.; et al. Enhancing Shear Strength Predictions of UHPC Beams through Hybrid Machine Learning Approaches. *Sci. Rep.* **2025**, *15*, 28259. https://doi.org/10.1038/s41598-025-13444-y.
53. Mirjalili, S.; Mirjalili, S.M.; Lewis, A. Grey Wolf Optimizer. *Adv. Eng. Softw.* **2014**, *69*, 46–61. https://doi.org/10.1016/j.advengsoft.2013.12.007.
54. Eberhart, R.; Kennedy, J. A New Optimizer Using Particle Swarm Theory. In Proceedings of the International Symposium on Micro Machine and Human Science, Nagoya, Japan, 4–6 October 1995; pp. 39–43.
55. Chatterjee, A.; Siarry, P. Nonlinear Inertia Weight Variation for Dynamic Adaptation in Particle Swarm Optimization. *Comput. Oper. Res.* **2006**, *33*, 859–871. https://doi.org/10.1016/j.cor.2004.08.012.
56. Shi, Y.; Eberhart, R. A Modified Particle Swarm Optimizer. In Proceedings of the IEEE Conference on Evolutionary Computation, Anchorage, AK, USA, 4–9 May 1998; pp. 69–73.
57. Lyngdoh, G.A.; Zaki, M.; Krishnan, N.M.A.; et al. Prediction of Concrete Strengths Enabled by Missing Data Imputation and Interpretable Machine Learning. *Cem. Concr. Compos.* **2022**, *128*, 104414. https://doi.org/10.1016/j.cemconcomp.2022.104414.
58. Yadav, S.; Shukla, S. Analysis of K-Fold Cross-Validation over Hold-Out Validation on Colossal Datasets for Quality Classification. In Proceedings of the IEEE 6th International Conference on Advanced Computing (IACC), Bhimavaram, India, 27–28 February 2016; pp. 78–83.
59. Kim, J.; Lee, D. Comparative Study on Hyperparameter Tuning for Predicting Concrete Compressive Strength. *Buildings* **2025**, *15*, 2173. https://doi.org/10.3390/buildings15132173.
60. Sapkota, S.C.; Haq, M.; Thapa, B.; et al. Predicting the Strength of Sustainable Graphene-Enhanced Cementitious Composites Using Novel Machine Learning and Explainable AI Techniques. *Infrastructures* **2026**, *11*, 146. https://doi.org/10.3390/infrastructures11050146.
61. Shrestha, A.; Sapkota, S.C. Hybrid Machine Learning Model to Predict the Mechanical Properties of Ultra-High-Performance Concrete (UHPC) with Experimental Validation. *Asian J. Civ. Eng.* **2024**, *25*, 5227–5244. https://doi.org/10.1007/S42107-024-01109-6.
62. Shapley, L.S. A Value for N-Person Games. In *Contributions to the Theory of Games (AM-28)*; Princeton University Press: Princeton, NJ, USA, 1953; Volume II, pp. 307–318.
63. Lundberg, S.M.; Lee, S.I. A Unified Approach to Interpreting Model Predictions. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 4766–4775. https://doi.org/10.5555/3295222.3295230.
64. Sapkota, S.C.; Panagiotakopoulou, C.; Dahal, D.; et al. Optimizing High-Strength Concrete Compressive Strength with Explainable Machine Learning. *Multiscale Multidiscip. Model. Exp. Des.* **2025**, *8*, 160. https://doi.org/10.1007/s41939-025-00737-y.
65. Nikoopayan Tak, M.S.; Feng, Y.; Mahgoub, M. Advanced Machine Learning Techniques for Predicting Concrete Compressive Strength. *Infrastructures* **2025**, *10*, 26. https://doi.org/10.3390/INFRASTRUCTURES10020026.
66. Taylor, K.E. Summarizing Multiple Aspects of Model Performance in a Single Diagram. *J. Geophys. Res. Atmos.* **2001**, *106*, 7183–7192. https://doi.org/10.1029/2000JD900719.
67. Rathnayaka, M.; Karunasinghe, D.; Gunasekara, C.; et al. Machine Learning Approaches to Predict Compressive

Strength of Fly Ash-Based Geopolymer Concrete: A Comprehensive Review. *Constr. Build. Mater.* **2024**, *419*, 135519. https://doi.org/10.1016/j.conbuildmat.2024.135519.

68. Chu, H.H.; Khan, M.A.; Javed, M.; et al. Sustainable Use of Fly-Ash: Use of Gene-Expression Programming (GEP) and Multi-Expression Programming (MEP) for Forecasting the Compressive Strength Geopolymer Concrete. *Ain Shams Eng. J.* **2021**, *12*, 3603–3617. https://doi.org/10.1016/j.asej.2021.03.018.
69. Kwek, S.Y.; Awang, H.; Cheah, C.B. Influence of Liquid-to-Solid and Alkaline Activator (Sodium Silicate to Sodium Hydroxide) Ratios on Fresh and Hardened Properties of Alkali-Activated Palm Oil Fuel Ash Geopolymer. *Materials* **2021**, *14*, 4253. https://doi.org/10.3390/ma14154253.
70. Tran, N.T.; Nguyen, D.H.; Tran, Q.T.; et al. Experimental Study and Machine Learning Based Study of Compressive Strength of Geopolymer Concrete. *Mag. Concr. Res.* **2024**, *76*, 723–737. https://doi.org/10.1680/jmacr.23.00144.
71. Sathiparan, N.; Jeyananthan, P.; Subramaniam, D.N. A Comparative Study of Machine Learning Techniques and Data Processing for Predicting the Compressive Strength of Pervious Concrete with Supplementary Cementitious Materials and Chemical Composition Influence. *Next Mater.* **2025**, *9*, 100947. https://doi.org/10.1016/j.nxmate.2025.100947.
72. Xu, Z.; Yue, J.; Pang, G.; et al. Influence of the Activator Concentration and Solid/Liquid Ratio on the Strength and Shrinkage Characteristics of Alkali-Activated Slag Geopolymer Pastes. *Adv. Civ. Eng.* **2021**, *2021*, 6631316. https://doi.org/10.1155/2021/6631316.
73. Raut, A.N.; Murmu, A.L.; Sapkota, S.C.; et al. Experimental and Machine Learning-Based Analysis of Oxide Ratios in Alumina-Silicate Geopolymer Formation. *Iran. J. Sci. Technol. Trans. Civ. Eng.* **2025**, *50*, 1199–1222. https://doi.org/10.1007/S40996-025-01875-6.
74. Khatoon, S.; Kaliluthin, A.K.; Sapkota, S.C. Experimental Investigation with Machine Learning Validation on the Mechanical Strength of Natural Fiber Reinforced Geopolymer Concrete. *Innov. Infrastruct. Solut.* **2026**, *11*, 218. https://doi.org/10.1007/s41062-026-02616-y.