scientific reports

OPEN

Interpretable machine learning approaches to assess the compressive strength of metakaolin blended sustainable cement mortar

Naseer Muhammad Khan[1,2,3,4,5✉], Liqiang Ma[1,2,3,5✉], Waleed Bin Inqiad[6], Muhammad Saud Khan[7], Imtiaz Iqbal[6], Muhammad Zaka Emad[9] & Saad S. Alarifi[8]

The use of naturally available materials such as metakaolin (MK) can greatly reduce the utilization of emission intensive materials like cement in the construction sector. This would reduce the stress on depleting natural resources and foster a sustainable construction industry. However, the laboratory determination of 28 day compressive strength (C-S) of MK-based mortar is associated with several time and resource constraints. Thus, this study was conducted to develop reliable empirical prediction models to assess CS of MK-based mortar from its mixture proportion using machine learning algorithms like gene expression programming (GEP), extreme gradient boosting (XGB), multi expression programming (MEP), bagging regressor (BR), and AdaBoost etc. A comprehensive dataset compiled from published literature having five input parameters including water-to-binder ratio, mortar age, and maximum aggregate diameter etc. was used for this purpose. The developed models were validated by means of error metrics, residual assessment, and external validation checks which revealed that XGB is the most accurate algorithm having testing R^2 of 0.998 followed by BR having R^2 values equal to 0.946 while MEP had the lowest testing R^2 of 0.893. However, MEP and GEP algorithms expressed their output in the form of empirical equations which other black-box algorithms couldn't produce. Moreover, interpretable machine learning approaches including shapely additive explanatory analysis (SHAP), individual conditional expectation (ICE), and partial dependence plots (PDP) were conducted on the XGB model which highlighted that water-to-binder ratio and sample age are some of the most significant variables to predict the C-S of MK-based cement mortars. Finally, a graphical user interface (GUI) was made for implementation of findings of this study in the civil engineering industry.

Keywords Cement mortar, Metakaolin, Interpretable machine learning, Gene expression programming, Compressive strength

Abbreviations
ML Machine learning
RMSE Root mean squared error

[1]Xinjiang Key Laboratory of Coal-bearing Resources Exploration and Exploitation, Xinjiang Institute of Engineering, Urumqi 830023, China. [2]Key Laboratory of Xinjiang Coal Resources Green Mining (Xinjiang Institute of Engineering), Ministry of Education, Urumqi 830023, China. [3]Xinjiang Engineering Research Center of Green Intelligent Coal Mining, Xinjiang Institute of Engineering, Urumqi 830023, China. [4]Sustainable Advanced Geomechanical Engineering, Military College of Engineering, National University of Sciences and Technology, Risalpur 23200, Pakistan. [5]School of Mines, China University of Mining and Technology, Xuzhou 221116, Jiangsu, China. [6]Department of Civil Engineering, College of Engineering & Physical Sciences, Aston University, Aston St, Birmingham B4 7ET, UK. [7]Department of Civil Engineering, Price Faculty of Engineering, University of Manitoba, Winnipeg, MB R3T 5V6, Canada. [8]Department of Geology and Geophysics, College of Science, King Saud University, P.O. Box 2455, Riyadh 11451, Saudi Arabia. [9]Department of Petroleum Engineering , King Fahd University of Petroleum and Minerals, Dhahran, Eastern Province, Saudi Arabia. ✉email: nmkhan@mce.nust.edu.pk; ckma@cumt.edu.cn

AI	Artificial intelligence
GP	Genetic programming
XGB	Extreme gradient boosting
PI	Performance index
SVM	Support vector machine
GEP	Gene expression programming
GBR	Gradient boosting regressor
MK	Metakaolin
GB	Gradient boosting
LR	Linear regression
ICE	Individual conditional expectation
ET	Expression tree
R^2	Coefficient of determination
DL	Deep learning
C-S	Compressive strength
RF	Random forest
MEP	Multi expression programming
R	Correlation coefficient
MAE	Mean absolute error
DT	Decision tree
BR	Bagging regressor
SHAP	Shapely additive explanatory analysis
MLP	Multi-layer perceptron
GUI	Graphical user interface
PDP	Partial dependence plots
OF	Objective function
ANN	Artificial neural networks
GA	Genetic algorithm
KNN	K-nearest neighbor

Mortars hold paramount importance in masonry construction since they are used to join bricks/stones in a masonry structure. Generally, mortars are made up of binding materials, aggregates, admixtures, and water, etc. Cement is one of the most significant contributors of CO_2 emissions from the construction industry, being the most widely used binder worldwide[1]. It has been reported that more than 7% of CO_2 emissions worldwide originate directly from the cement industry[2]. The current annual production is 4,000 million tons, with projections indicating a rise to 6,000 million tons by 2060[3]. The release of greenhouse gas CO_2 in the atmosphere by virtue of cement production results in climate change and global warming. This leads to melting of icecaps, increased global temperatures, and brings about extreme weather changes. Thus, efforts are being made to discover new materials to replace cement in the mortars to reduce the negative effects of cement production on the environment and to make the mortars more durable, cost-effective, and eco-friendly[4,5]. The use of alternative materials to replace cement in concrete and mortar will help to lessen the utilization of depleting natural resources such as limestone and gypsum since these the production of cement calls for huge amounts of these precious natural resources. This will also lead to a decrease in carbon emissions within the construction sector and reduce the adverse effects of human development on the natural environment. Moreover, the increased utilization of natural materials in cement mortars will help reduce the costs of construction. In this context, metakaolin (MK) is a promising ingredient that may be added to mortars in place of cement to enhance their cementitious qualities. It is an aluminosilicate that is thermally activated and occurs naturally as a white mineral. It is produced by kaolin clay heating to elevated temperatures between 650 ºC and 800ºC[6,7]. As a pozzolanic material, MK combines with calcium hydroxide Ca(OH)$_2$ in water forming compounds having cementitious qualities[8]. Figure 1 illustrates how the use of MK in place of cement improves the rheological properties, reduces the permeability and increases the consistency of mortars[9]. Although there has been much research conducted about the utilization of MK and similar materials as a replacement of cement in mortar mixes, there is no explicit method available for optimal design of MK-based cement mortars. This is because the fresh and hardened mortar materials mechanical properties containing mineral admixtures like MK, fly ash, or limestone powder etc. exhibit a highly non-linear behaviour related to their mixture constituents. Due to this non-linear behaviour, it is particularly difficult to develop an empirical formula for predicting mechanical properties using traditional statistical and mathematical models[10]. In addition to MK, several other materials have been used to make sustainable mortars. For instance, Elfadaly et al.[11] developed high-performance geopolymer mortar incorporating pumice powder and rice straw ash as sustainable filler materials. The authors replaced rice husk ash and pumice powder with blast furnace slag at 0%, 15%, 25%, and 30% by volume and revealed that compression strength generally decreases with an increase in the pumice and rice husk ash content.

Accurately determining various mechanical characteristics is necessary for effective utilization of MK-based cement mortars in the construction industry. Out of these mechanical properties, compressive strength (C-S) is one of the most important properties since it is an indicator of overall mortar quality and durability. The C-S test is one of the most common ways for investigating mortar durability, performance, and reliability[12,13]. However, it has been highlighted time and again that an exhaustive experimental scheme is inevitable to get an idea about the relationship of C-S with the mixture components of admixture containing mortars. The mortar samples casted using different types and sizes of aggregates, variable water-to-cement ratios, and different dosages of superplasticizer (SP) etc. must be tested for compression strength to get a clear idea about the relationship of
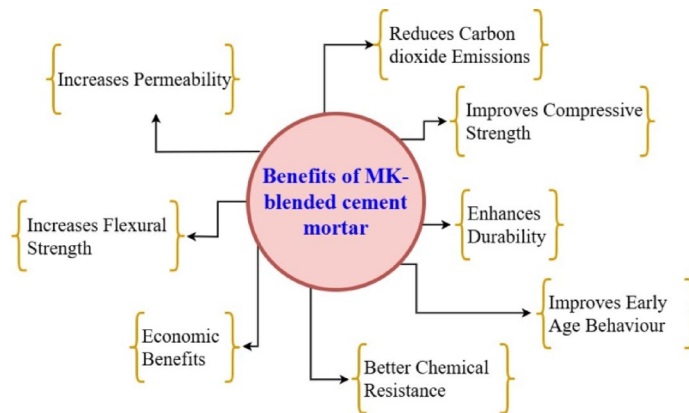
Fig. 1. Benefits of sustainable cement mortar blended with metakaolin clay.

different mixture components with the C-S. Despite the importance of such an inclusive experimental program, it has several limitations. First of all, a combination of several loading conditions and other physical conditions is necessary for getting a good idea about the relationship of mixture composition and other variables with CS[14]. In addition, these types of experiments must be carried out in controlled environments to ensure minimum distractions and interferences on the testing procedures. However, ensuring these would result in wastage of significant time and precious resources[15]. The underlying complexities of the testing procedures and chances of human errors in conduction of such comprehensive experiments further adds to the problem. Furthermore, the most commonly used tests to access mechanical properties of mortars such as C-S are destructive in nature meaning that the testing sample gets destroyed during the experiment providing another way of generating construction waste. In conclusion, the conventional testing methodologies have severe limitations. In this regard, scientists have come up with the innovative solution to use machine learning (ML) techniques for forecasting different properties of construction materials to overcome the limitations highlighted above[16]. The use of these ML methodologies for this purpose will save time, effort, and cost which is otherwise required in case of laboratory testing[17]. In Sect. 1.2, a thorough review of previously conducted research on utilization of contemporary ML techniques for prediction of different mortar and concrete properties is given.

Recent decades have seen several attempts to create reliable empirical predictive models to forecast different mechanical properties of cementitious composites using ML techniques to overcome limitations of traditional testing techniques. ML is an area of Artificial Intelligence (AI) that utilized statistics and mathematics to enable machines (computers) to identify patterns and correlations in data. Predictions for new data are then made using the knowledge gained from the data, without the need for direct human involvement. Recently, ML algorithms, with their capability of modelling complex relations between various inputs and outputs, have become the most promising way to predict complex properties of different materials[18,19]. In civil engineering, several machine learning techniques, including as random forests (RF), XGB, GEP, and artificial neural networks (ANN), have been extensively used to forecast different soil compaction and categorization characteristics[20–23], load carrying capacity of composite concrete columns[24–27], mechanical properties of innovative concrete composites[28,29], slope failure and landslide susceptibility etc[30–33]. These studies help accurately estimate material attributes, including compressive strength, which necessitates labour- and resource-intensive laboratory studies otherwise. When carefully trained, ML models can predict outcomes far more efficiently, saving a great deal of money, labour, expertise, and time in the lab. The engineering community has seen a substantial rise in the use of ML for the resolution of structural and civil engineering issues[34]. One major benefit of ML models is their capacity to automatically analyze and discover relationships between many parameters that arise in complicated engineering situations, such as predicting corrosion, assessing the impact of fiber additions, and predicting strength. ML has the potential to revolutionize the civil engineering industry by providing a powerful tool for analyzing and understanding material properties and construction processes by replacing traditional practices and increasing efficiency, accuracy and efficiency. By identifying hidden patterns and trends in large datasets, ML algorithms provide valuable insights that enhance decision-making and promote efficiency and sustainability. A significant application of ML in civil engineering is creating predictive models to estimate the compressive strength of concrete and mortar. Proven effective in optimizing material mixtures, these models minimize the need for extensive lab testing. Additionally, ML in civil engineering supports data-driven decision-making, fostering more sustainable and eco-friendly construction methods. ML is a powerful tool for advancing environmental sustainability in civil engineering through improved material selection and waste reduction. In the field of concrete technology, ML provides a powerful advantage, allowing the development of predictive models that accurately anticipate concrete properties given specified input parameter.

Literature search

The effectiveness of ML algorithms like ANN, RF, decision tree (DT), artificial neuro fuzzy inference systems (ANFIS) etc. to predict different mechanical properties of construction materials has been thoroughly reported in the literature. For example, Handousa et al.[35] utilized two algorithms including ANN and Gaussian Process Regression (GPR) to predict compression strength of concrete-filled double-skin steel tubular columns due

to their widespread uses in the construction industry. The authors used experimental data from their own laboratory experimentation and highlighted that ANN outperformed its counterpart to predict CS by achieving lower MAE and higher correlation with actual values. Also, the authors employed shapely additive analysis which revealed that the thickness of the external plate and compressive strength of concrete are some of the most predominant factors for determining the output. Also, Abdellatief et al.[36] utilized ML for determining the strength in compression and porosity of innovative foam glass for practical applications. The authors used various algorithms including s gradient boosting (GB), random forest (RF), Gaussian process regression (GPR), and linear regression (LR) etc. in a comparative manner and concluded that GPR model was the most accurate with R-values of 0.91 and 0.82 for porosity and compressive strength respectively. Regarding the utilization of ML to determine properties of sustainable mortar, Faraj et al.[37] utilized ANN, linear and non-linear regression, and M5P-tree algorithms to predict tensile strain capacity of cement mortars having fibre reinforcement which are commonly known as engineered cementitious composites (ECC). The authors collected 115 experimental results of tensile strain capacity of ECC from punished literature and used them to train the algorithms for tensile strain capacity prediction. The authors assessed the accuracy of developed predictive models using various error evaluation metrices and highlighted that ANN predicted tensile strain capacity with highest accuracy ($R^2 = 0.98$). Also, Joseph Mwiti Marangu[38] utilized a support vector machine (SVM) and ANN algorithms for the purpose of predicting compressive strength of cement mortars having calcined clay as a replacement of cement in various ratios ranging from 30 to 50%. The study utilized laboratory prepared samples to calculate the C-S of cement mortar at the age of 2,7, and 28 days and utilized the resulting data to build SVM and ANN models. The results revealed that ANN predicted mortar strength with greater accuracy having testing R^2 of 0.95 compared to SVM R^2 of 0.88. Similarly, Çalışkan et al.[39] developed cement mortar by blending fly ash and nano calcite with cement. The samples were cured and tested for C-S after 1,3,7, 28, and 90 days and the data were used to develop predictive models for C-S using SVM and extreme learning machine (ELM) algorithms. The authors compared actual and predicted results using mean squared error (MSE) and R^2 and concluded that SVM predicted C-S with greater accuracy having R^2 and MSE values equal to 0.982 and 3.882 compared to ELM values of 0.91 and 20.39. In the same way, Ozcan et al.[40] made use of RF, SVM, and AdaBoost techniques on a dataset of more than 250 specimens of cement mortar made with waste rubber tire and blast furnace slag. The ML models were developed for prediction of a single output parameter i.e., C-S by using four input features including age of specimen, cement content, and the quantities of slag and rubber tire waste. The comparison of performance of ML algorithms revealed that AdaBoost predicted C-S of cement mortar with highest accuracy ($R^2 = 0.9831$). Moreover, in a recently conducted study by Zhiqiang Chen[41], the flexural strength and compressive strength of eco-friendly cement mortar made by replacing cement with industrial by-products were predicted using XGB and bagging regressor. The author casted mortar samples by replacing cement with fly ash and rice husk ash in various ratios ranging from 20 to 40%. The study considered four input features containing cement content, specimen age, rice husk ash content, and fly ash content to predict flexural and C-S. The contrast of both algorithms showed that XGB proved to be more accurate than bagging regressor having R^2 value equal to 0.98 for C-S and 0.94 for flexural strength. In another research conducted by Wang et al.[42], the C-S prediction of cement mortar using dynamic elasticity data was investigated. The authors developed a predictive model based on back propagation neural network (BPNN) which is further based on the Bayesian theorem which helped to overcome the problem of overfitting significantly. The authors gathered a dataset of almost 160 experimental results of cement mortar compressive strength having slag as a replacement of cement. The error evaluation of BPNN model revealed that there exists a very strong correlation between actual and anticipated C-S values having R^2 equal to 0.95. A summary of previous studies regarding different properties prediction of cement mortars is given in (Table 1).

Reference	ML algorithms used	Year	Material	Predicted output	Best model	Best model R^2
Shah et al.[43]	RF, XGB, Ridge regression and light gradient boosting Machine	2022	Fly ash and slag-based alkali-activated mortar	C-S	XGB	0.94
Dehestani et al.[44]	Bagging regressor, ANN, XGB, expression tree regressor	2022	Fibre-reinforced concrete and mortar	Fracture toughness	XGB	0.92
Chen et al.[41]	XGB and bagging regressor	2024	Cement mortar blended with rice husk ash	Flexural strength	XGB	0.98
Ali et al.[45]	XGB	2021	Asphalt concrete mixture	Dynamic modulus	XGB	0.96
Xu et al.[46]	DT, ANN, RF, XGB	2021	Industrial waste mortar		XGB	0.94
Asteris et al.[47]	SVM, RF, AdaBoost, KNN, DT	2021	Cement mortar	C-S	AdaBoost	0.996
Gayathri et al.[48]	DT, XGB, GBR, MLP, SVM, AdaBoost	2022	Clay-based mortar	C-S	XGB	1
Izadgoshasb et al.[49]	ANN with multi-objective grasshopper optimization algorithm	2021	3D printed cement mortar	C-S		0.98
Fei et al.[50]	XGB, RF, AdaBoost	2023	Recycled powder mortar	C-S	XGB	0.842
Nakkeeran et al.[51]	ANN	2023	Glass fiber reinforced mortar	C-S		0.99
Kurt et al.[52]	LR, RF, DT, KNN, DL	2023	Geopolymer mortar	C-S and flexural strength	RF	0.983

Table 1. Summary of previously conducted studies.

Research significance

As discussed in Sect. 1, the use of alternative cementitious materials like MK in place of cement in concrete and mortar is crucial to foster a sustainable civil engineering industry. Since cement is responsible for the most emissions from the construction industry, the reduction in its usage will reduce the associated CO_2 emissions from the industry and consequently reduce the ill effects of human development on nature. It will also help conserve the natural environment by reducing the utilization of limited natural resources like gypsum and limestone etc. Also, MK utilization in mortar will serve as a feasible avenue for utilization of huge amounts of MK available. As evident from the literature summary in (Table 1), much research has been done regarding prediction of strength properties of cement mortars having mineral admixtures like fly ash, marble powder etc. in various percentages in the mortar mix. However, the subject of C-S prediction of mortar containing MK as a replacement of cement is comparatively unexplored. It is largely attributed to the non-linear behaviour of C-S of MK containing mortar in relation to its mixture composition. Also, the experimental determination of strength of MK blended mortar requires extensive time and resources. To bridge this gap in the literature, this study intends to utilize novel ML algorithms including boosting-based techniques (XGB and AdaBoost), bagging regressor, and evolutionary techniques (GEP and MEP) to predict C-S of MK-based cement mortars in a comparative manner. The use of MEP and GEP for prediction purpose of C-S of MK-based cement mortar is attributed to the fact that these two are grey-box models because since they give an empirical equation as their output in contrast to other famous techniques like ANN and XGB which don't possess the ability to do so. The equations yielded by MEP and GEP will help the professionals to effectively use the developed models and impart transparency to the underlying algorithm methodology[53]. Similarly, the rationale behind using algorithms like GBR, AdaBoost, and XGB stems from their superior accuracy and easy implementation. These algorithms do not need to specify any optimal model architecture before model development[54]. Moreover, a particular novelty of this study is that it gives a comparison of various bagging, boosting, and evolutionary techniques to predict C-S of MK-based cement mortar. This will help to identify the type of ML techniques most feasible for utilization in civil and materials industry for prediction of different material properties. Furthermore, this study utilizes several cutting-edge interpretable ML approaches like SHAP, partial dependence plots (PDP) analysis, and individual conditional expectation (ICE) to reveal the most important variables which contribute to the strength of MK-based cement mortar. The investigation of relationships between different inputs and output by means of the explanatory analysis techniques will provide insights into the individual and combined effect of different inputs on the outcome. This will help to optimize the mixture composition of MK-based cement mortars to cater for specific site conditions. In addition, this study depicts a practical of the utilization of transformative power of ML in the field of civil engineering by means of a graphical user interface (GUI). This GUI will enable the professionals and stakeholders to effectively determine C-S of MK-based mortars based on its mixture composition without conducting extensive experimentation.

Research methodology

Figure 2 shows the complete step-by-step methodology followed in this study. These steps include (i) Collection of the data, (ii) Analysis of data, (iii) Data splitting, (iv) ML model development, (v) Assessment of error, (vi) explanatory analysis, and (vii) GUI development. Based on the extensive review of literature, it was identified
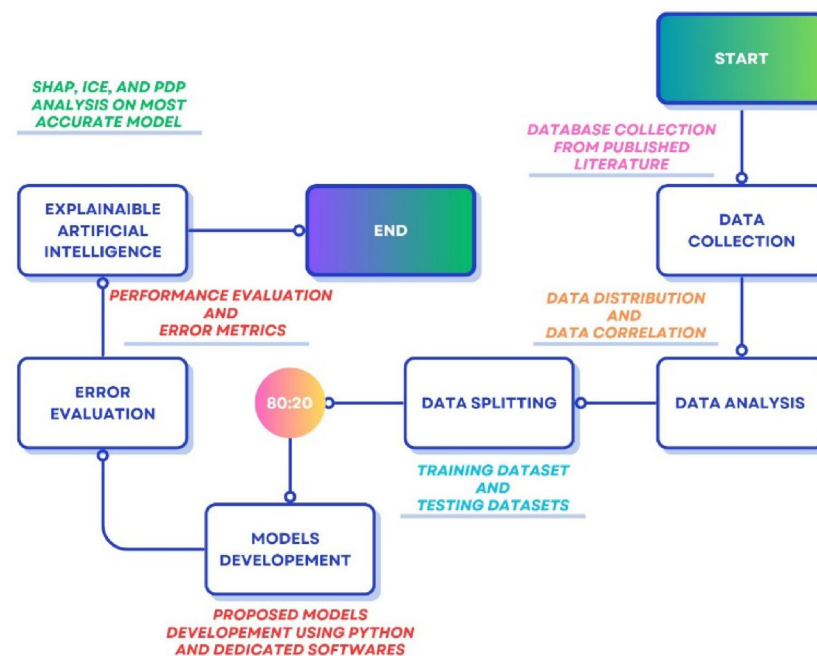

Fig. 2. Step-by-step methodology followed in this study to develop ML-based predictive models.

that there is a lack of accurate empirical predictive models for MK-based cement mortars despite its benefit as compared to normal cement mortar. Thus, this study will address the problem of predicting C-S of MK-based mortars based on different ML techniques. For this purpose, a dataset containing all influential variables to predict C-S was collected from the literature which was then processed and analysed using different statistical techniques to check for outliers and investigate the correlation and distribution of the data. After data analysis and representation, the model development process will start which encompass tunning the hyperparameters of algorithms to check which set of parameters will yield the highest accuracy. The algorithms will be trained using these parameters and then the trained algorithms will be validated by error evaluation and scatter plots etc. After training of algorithms and error evaluation, their performance will be compared to point out the most important algorithm which will then be used to conduct further explanatory analyses techniques like SHAP, PDP and ICE etc. Furthermore, the findings of the algorithms will be deployed by means of a GUI to enable the professionals to use the models that are developed in this study.

Prediction methods
Gene expression programming (GEP)

Ferreira in the year 2001 first proposed the GEP algorithm, which combines aspects of genetic algorithms (GA) and genetic programming (GP)[55]. In GEP, the concept of "survival of the fittest," inspired by Darwinian principles, plays a central role. The process initiates with the generation of random initial chromosome population, which is then decoded into mathematical expressions. Fitness evaluations determine the optimal individuals, and genetic operators drive the continuous generation of offspring. Continuous iteration refines the solution, ultimately leading to the optimal outcome, as depicted in (Fig. 3). Figure 3b,c illustrate the process through the many steps required to create a mathematical equation for GEP. There are three key parts of GEP namely: (i) Chromosome structure, (ii) Genetic operators, and (iii) Fitness evaluation. Each of the above is explained below:

Chromosome structure
Chromosomes being composed of interconnected genes by a function named linking function, with each gene featuring a head and a tail. The head is made of a set of functions such as logical and mathematical operations and terminations such as variables or constants. On the other hand, only terminations are present in the tail. Expression trees (ETs), which provide explicit expressions and are essential to the decoding of chromosomes[56].

Genetic operators
Chromosomes facilitate the emergence of the next generation through genetic operators, including mutation, transposition, and inversion[57]. Notably, mutation excels as the most efficient operator, offering substantial modification potential. The presence of mutation significantly enhances the efficiency of the individual adaptation process, leading to effective solutions across various problem domains.
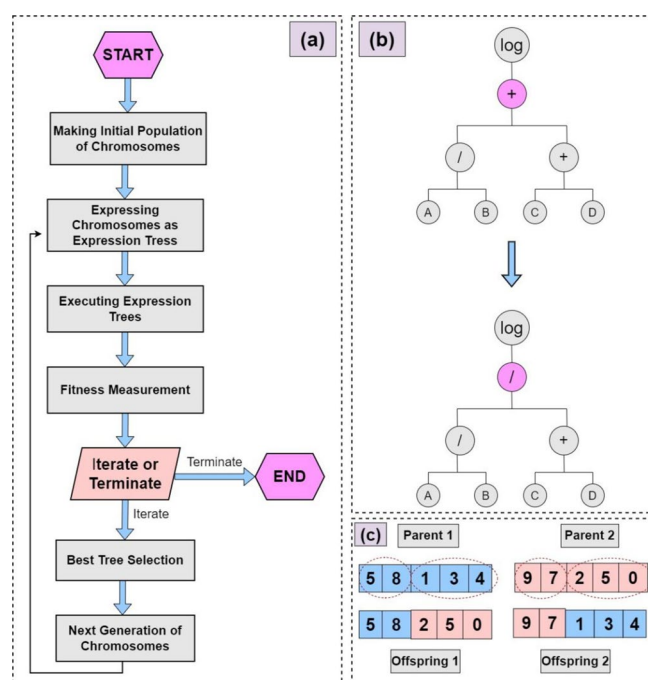

Fig. 3. Pictorial representation of: (**a**) GEP methodology; (**b**) genetic mutation; (**c**) genetic crossover of two parent chromosomes.

Fitness evaluation

The performance of chromosomes is assessed using fitness. The more evolutions there are, the more fitness there is in the population. The final fitness value is reached when the fitness value does not change after a certain number of evolutions.

Multi expression programming (MEP)

MEP is a subclass of GP that is excellent at offering several solutions, including straightforward linear solutions that fit inside a single chromosome, thereby improving the accuracy of the model[58]. The Binary Tournament is a process used to select two potential parents and recombine them to create new offspring. The genetic evolution of offspring involves both crossover and mutation, resulting in the production of novel offspring. These features allow MEP to search through a wide range of possibilities and identify the best responses. This system allows flexibility in chromosomal rearrangement during the evaluation phase by ensuring that individuals who cannot reproduce are not included in subsequent generations and eventually discarded. The first step in the MEP process is to develop the random population of computer codes. Subsequent steps continue until the initial condition is satisfied in order to formulate the most viable answers to the question. A set (or sets) of strings of formulas are used to describe the MEP genes. The size of the code (the length of the chromosome) is determined by the set of genes and is held the same throughout the process. MEP's superiority lies in its ability to create reliable models across different contexts. The optimization of the MEP algorithm involves the adjustment of parameters such as crossover probability, code length, sub-population size and mathematical functions. The selection of optimal values will improve the model accuracy but may enhance computational requirements[59]. The length of the resultant equation in MEP is exactly proportional to the length of the code. The architecture of MEP model is shown in (Fig. 4).

Extreme gradient boosting (XGB)

XGB is recently created gradient boosting tree-based model[60]. It is similar to gradient boosting methodology (GBM) but more accurate and efficient. XGB, unlike GBDT, employs a more regularized model formulation to control model complexity. XGB additionally supports column subsampling for features, a sparse technique for missing data, and multithreaded parallel computation. To optimize and enhance the XGB model, it is necessary to determine multiple parameters through data learning. XGB excels at identifying relationships within data. It employs a sequence of decision trees to progressively refine predictions. This collaborative approach results in a robust and accurate model. XGB improves its predictions over time through a process of recurrent training. It starts with the creation of a first model based on the training data. It then generates further models, each of which is specialised in correcting the errors that were missed by the previous model. This cycle is repeated until the performance of the algorithm is better than a pre-defined target or until a pre-defined maximum number of models has been reached. Figure 5 presents the model architecture to reach valid output[61]. The XGB algorithm is concisely summarized by the stages listed below:

i. Assume training and testing datasets and divide accordingly in ratios such as 2/3 and 1/3 or 3/4 and 1/4 etc. of the total dataset. Any general dataset D is represented by Eq. (1).

$$D = \{(x_i, y_i)\}_{i=1}^{n} \tag{1}$$

Then a differentiable loss function is adopted for the evaluation of performance of model. This loss function is defined as follows in Eq. (2):

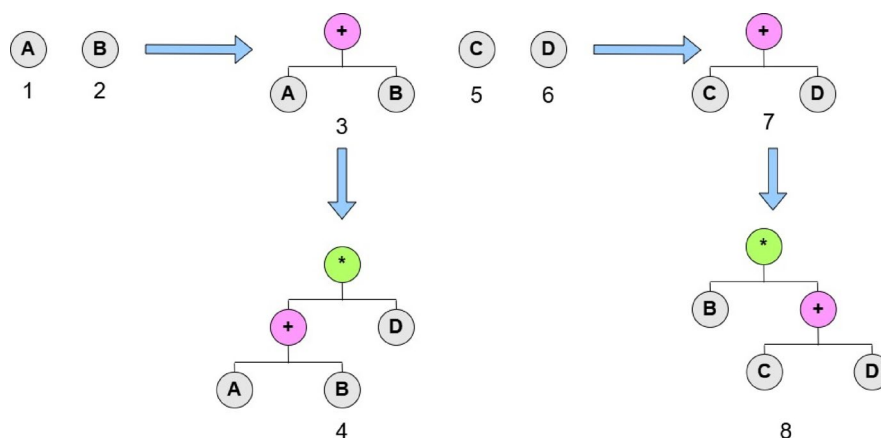$$\text{Loss function} \qquad L(y_i, \gamma) = \frac{1}{n} \sum_{i=1}^{n} (y_i - \gamma)^n \tag{2}$$

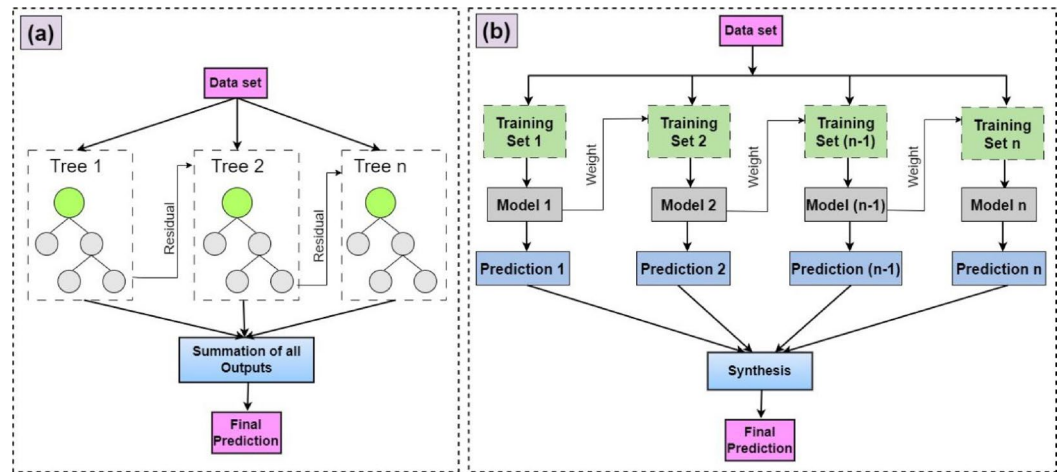**Fig. 4**. Pictorial representation of MEP model architecture.

Fig. 5. Pictorial representation of: (**a**) XGB methodology; (**b**) AdaBoost methodology.

Where predictors and predictands sets are represented by x_i and y_i respectively and i is any value assumed from 1 up to any number n which is the overall number of the datapoints. The predictions are illustrated by the symbol γ as shown in the equation of loss function. Different loss functions can be considered including mean squared error (MSE). Different hyperparameters including maximum depth of trees for base learners, number of trees, and learning rate of boosting either using default or assumed parameters or using GridsearchCV which is a library for automatic searching of best parameters for a dataset in python[62].

Adaptive gradient boosting (AdaBoost)

AdaBoost is well-known as a highly powerful boosting technique. In fact, its significance goes beyond boosting, since it is ranked among the top ten data mining algorithms. AdaBoost uses the training set to create a weak learner and modifies the distribution of data for subsequent rounds of training of the weak learner depending on the performance of the prediction. It should be noted that in the following step, more attention is given to training samples with low prediction accuracy in the previous step. Lastly, the weak learners are combined with different weights to form a strong learner[63]. The graphical illustration of the adopted methodology is represented in (Fig. 5b). Following methodology is adopted for AdaBoost predictions. Let a training dataset is consisted of 1 up to m labels ($x_{1\,to\,m}$, $y_{1\,to\,m}$) depicted as (x_1, y_1),…, (x_m, y_m) for easy understanding where x_i belongs to µ domain and y_i belongs to a range from −1 to +1. A distribution D_t is calculated for each round (t = 1 to T) of m training instances, as illustrated below in Eq. (3).

$$D_1 = \frac{1}{m} \text{ for i} = 1, \dots \dots, \text{ m} \tag{3}$$

Next, to identify a weak hypothesis, a weak learning method is used $h_t : \mu \rightarrow \{-1, +1\}$ which aims for minimizing the weighted error $?_t$ which is relative to D_t as shown in the subsequent Eqs. 4 and 5 respectively:

$$?_t = Pr_i \sim D_t[h_t(x_i) \neq y_i] \tag{4}$$

and

$$D_{t+1}(i) = \frac{D_t(i)\exp - \alpha\, y_i h_t(x_i))}{Z_t} \tag{5}$$

Where Z_t is factor for scaling selected such that the distribution is ultimately formed by D_{t+1}. A weighted sum of weak hypotheses is finally calculated to determine the sign by the combined hypothesis (H) to arrive at final prediction. This final prediction is represented by Eq. 6.

$$H(x) = sign\left(\sum_{t=1}^{T} \alpha_t h_t(x)\right) \tag{6}$$

Bagging regressor (BR)

The ensemble machine learning approach known as "bagging regressor", which is a contraction of "bootstrap aggregating", expands the predictive model with more data as it is trained, increasing the accuracy and stability of ML models by integrating multiple base learners. Regression algorithms, such as DTs, are commonly used as the basic learner in Bagging Regressor applications[64]. By using a subset of the training data to train the base learners, bagging decreases variance and overfitting. Each element of the bagging technique has the same probability of being found in the newly created data set. There is little effect on predictive ability by adding more data points to the training set. Also, the divergence could be substantially reduced by adjusting the forecast in line with the intended outcome. All these data sets are the basis for the development of new models, which are regularly updated. The ensemble models provide predictions, each of which is averaged. It is possible to
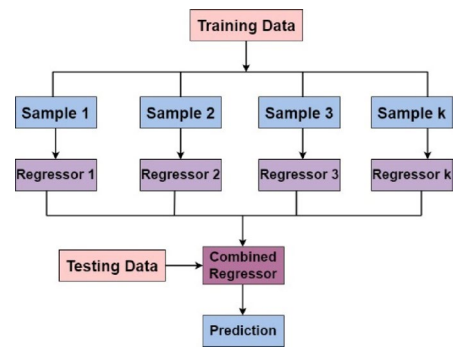
Fig. 6. Flowchart representation of bagging regressor algorithm.

No.	Reference	Samples count	Age	Compressive strength (MPa)
1	Cizer et al.[69]	5	3,7,14,56,90	47.94–66.23
2	George et al.[70]	16	7,14,28,60,90	17.60–69.70
3	Courard et al.[71]	15	1,3,7,28	27.40–71.20
4	Potgieter-Vermaak and Potgieter[72]	20	3,7,14,28,60	23.40–92.80
5	Sumasree and Sajja[73]	35	3,7,14,28,56	24.93–35.71
6	Kadri et al.[74]	12	1,3,7,14,28,56	35.21–99.177
7	Al-Chaar et al.[75]	3	3,7,28	6.06–10.73
8	Dhawan et al.[76]	143	7,28,60,90	15.45–54.90
	Total	249	1–90	6.06–99.17

Table 2. Summary of the dataset used for model development.

derive regressive forecasts by taking the average of forecasts from many models[65]. An improved and more widely applicable model is produced by this averaging procedure, which reduces the effect of over-fitting and outliers in individual base learners. Figure 6 illustrates how Bagging regressor combines training and testing data to obtain final predictions.

Cement mortar database
Experimental data collection
The establishment of a reliable database is a fundamental prerequisite for the development of ML-based predictive models. However, in the context of cement mortar, the data collection process is inherently complex due to complex mixture composition of mortar typically comprising water, binders (such as cement), aggregates, and both natural and synthetic admixtures. Consequently, the careful selection of input features is of critical importance during data collection. Moreover, proper consideration must be given to parameters such as quantity and type of cement as these characteristics can significantly influence C-S of the mortar. Another challenge lies in the acquisition of sufficient data to effectively train ML algorithms. Traditionally, cement mortars are prepared in laboratories, cast into molds, cured under controlled conditions, and tested for C-S using a compression testing machine. This experimental process is resource-intensive, time-consuming, and requires continuous oversight, which limits the feasibility of producing large numbers of specimens. As a result, most studies in the existing literature report compressive strength values up to only 28 or 90 days of curing. Data pertaining to long-term strength development is notably scarce, posing a significant limitation to the construction of a robust and generalizable database. Furthermore, during the collection of data, it is significant to check that the input parameter values cover a wide range of possible values because the accuracy of the developed ML model largely varies on the reliability of the database.

To address these constraints, an extensive database of 249 experimental results of cement mortar C-S containing MK have been aggregated from 8 different studies published in the literature whose summary is given in (Table 2). Although there were some smaller databases already available in the literature, the authors compiled a new and extensive database due to the following reasons: (1) the already existing datasets considered different set of parameters than used in current study, (2) the input variable values were not distributed evenly over a large range in existing datasets. It is an essential factor to take into account because the number of experimental instances in a dataset al.one aren't sufficient to ensure the reliability of the database, the distribution of each input and output parameter(s) also plays an important role. Thus, the mortars having maximum diameter of aggregate greater than 6 mm and C-S results for ages more than 90 days were not included in the database since such type of experimental results were less than 5% of the total collected dataset and their inclusion would have reduced the reliability of the collected dataset and in turn the accuracy of developed ML algorithms. In regard to the input features, the collected database features only those experimental results containing the following
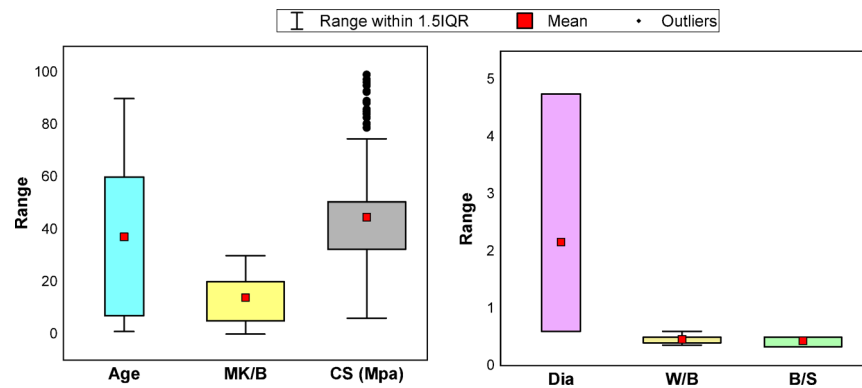
Fig. 7. Boxplots of variables used in current study to find outlier values.

	Units	Symbol	Maximum	Minimum	Mean	Standard deviation	25%	50%	75%
Age	Days	A	90	1.0	38.3	31.82	7.0	28	60
Maximum aggregate diameter	mm	Dia	4.75	0.6	2.07	1.789	0.6	0.60	4.75
Metakaolin percentage	% w/w	MK/B	30	0	13.89	10.10	5.0	15	30
Water-to-binder ratio	–	W/B	0.6	0.36	0.467	0.057	0.42	0.46	0.60
Binder-to-sand ratio	–	B/S	0.50	0.33	0.433	0.080	0.33	0.50	0.50
Compression strength	MPa	C-S	73.10	6.06	41.78	12.117	32.05	43.40	73.1

Table 3. Descriptive statistics of the processed data.

six input parameters: (i) age of specimen, (ii) water-to-binder ratio, (iii) metakaolin percentage in relation to total binder content, (iv) maximum diameter of aggregate, (v) superplasticizer percentage, and (vi) sand-to-total binder content ratio. This is due the fact that these inputs have been recognized as the most important features to predict the mortar C-S by previous related studies[39,41,44,66–68]. It is worth mentioning that in the collected database of experimental studies, OPC was used as the main binding agent and MK was added to the mortar mix in different percentages. The water-to-cement ratio was different for experimental values collected from each study, but the C-S of mortar was measured by casting cubic specimens of 50 cm x 50 cm x 50 cm, curing them at different ages (1,3,7,14,28,60,90), and crushing them using a compression testing machine to measure the C-S according ASTM C 109.

Outlier detection and removal

Regression analysis commonly involves removing outliers from the database prior to the development of regression-based models. Outliers represent anomalies in the data and are defined as are those points which deviate significantly from other data[77]. In regression analysis involving extensive data, outliers pose a significant challenge, and special care must be taken to detect and remove them from the data[78]. Therefore, the database comprising 249 experimental instances, was analysed for potential outliers using a descriptive statistical method outlined in[79]. Outlier detection was performed through hypothesis testing, where a p-test was applied at a 5% significance level. In this context, the null hypothesis assumes that the data values follow a normal distribution, meaning they fall within the expected range of variation. The alternative hypothesis, on the other hand, suggests that certain values particularly those significantly higher or lower than the rest may not belong to the same distribution and could be considered statistical outliers. Identifying and addressing such outliers is important to ensure the reliability and accuracy of the dataset used for model development. The result of this analysis in the form of bivariate boxplots for each variable involved in the study is given in (Fig. 7). The boxplots in Fig. 7 is a useful way to represent the summary of statistical characteristics of the data. The 25–75% range is represented by different coloured bars while the inter quartile range (IQR) is represented as black lines and the outliers are shown as black dots in (Fig. 6). After the identification of outliers by means of box plots as given in (Fig. 7), the strength value was capped at 75 MPa to remove the outliers from the database. Thus, a total of 234 data instances were left to be used for model development. This remaining dataset will be used from now onwards in the paper for statistical analysis and for development of the algorithms. The statistical summary of the remaining data finalized for model development is given in (Table 3). The symbols of the input and output variable(s) along with their units of measurement is given in (Table 3). Moreover, different statistical parameters like average value, maximum value, and standard deviation etc. for each variable is also given in the table.

Data description, correlation, and partitioning

Figure 8 depicts a histogram representation of the data. The graphic illustrates how the dependent and independent variables are dispersed over a broad range. The age ranges from 1 day up to 90 days encompassing
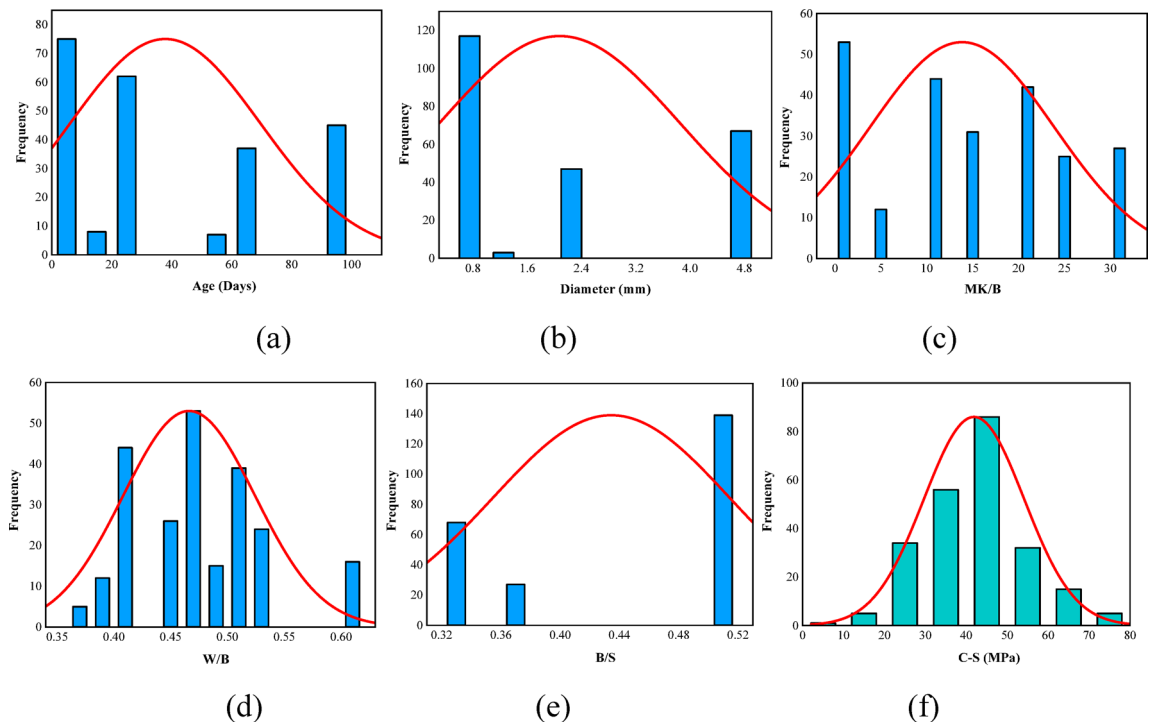
Fig. 8. Data visualization by means of histograms; (**a**) age; (**b**) diameter; (**c**) MK/B; (**d**) W/B; (**e**) B/S; (**f**) C-S.

values like 7,14, 26, 56, 60 within the range. Also, W/B ranges from 0.36 to 0.6 which is a considerably wide range for W/B ratio. In the same way, binder-to-sand ratio varies from 0.33 to 0.50. The range of maximum diameter of aggregate is from 0.6 to 4.75 which is also a very large range of values. It is because different experimental results collected from variable studies used different sized aggregate to make cement mortars. Furthermore, the percentage of MK in relation to total binder content has range from 0 to 30% which means that the dataset contains experimental results for MK-based cement mortars in which percentage of MK varies from 0 to 30%. This type of data distribution makes sure that the developed predictive models can be used for wide range of input configurations to predict the desired outcome[80]. The strength of mortar in the same way varies from 6.06 MPa to more than 70 MPa. This wider range of output variable also allows the models trained using this data to be implemented practically in a wide variety of conditions to predict C-S of Mk-based cement mortars ranging from 6 MPa to more than 70 MPa.

In addition to investigating the distribution of the data, the examination of inter-dependency between different features (both input and output) is crucial before starting the formal training process of the algorithms. This is due to the fact during developing ML-based predictive models, the dependency of input variables upon each has a significant role in addition to the correlation between inputs and outputs. If the variables are correlated more with each other rather than with the output, there is a risk of encountering the problem of "multi-collinearity". This term simply implies that the inputs are correlated greatly with each other rather than with the output and the performance of resulting ML algorithms can be affected if this problem is present in the data[81]. Thus, to investigate the inter-dependency between all possible feature combinations involved in this study, the Pearson's correlation matrix or commonly referred to as heat map is given in (Fig. 9). The values written in Fig. 9 are the correlation which range between −1 and 1 and the strength of relationship between two variables is directly related to the value of R with 0 indicating no correlation at all and 1 showing a 100% perfectly linear correlation. Whereas the direction of relationship (positive or negative) between two variables is depicted by the sign of R. Correlation values above 0.8 often indicate a fair correlation among considered two variables, whereas values above 0.9 typically indicate a superior correlation[82]. Thus, to make a ML model resistant to multi-collinearity, it is important that the R value among any two variables does not exceed 0.8[81]. Figure 8 clearly shows that the correlation of all variable combinations is below 0.8, confirming that multi-collinearity will not arise to impact the model's performance. It is evident that most of the features have a minimal correlation between them. Only maximum aggregate dia and age has a negative 0.44 R value between them.

The final step after data description and data correlation analysis is the partitioning of the data where the data is split into separate sets to be used to train and test the algorithms. The major portion of the data is utilized in training while the relatively small testing set remains unexposed to the algorithm during training phase, and it is used to check the performance of the trained algorithms. The dataset was divided randomly into 80% for training and 20% for testing, with the training set used to build the algorithms and the testing set for performance assessment. The data splitting was done in this particular ratio to make sure that the algorithms have significant amounts of data to be trained properly and at the same time reserving enough data for the algorithms to be tested[83].
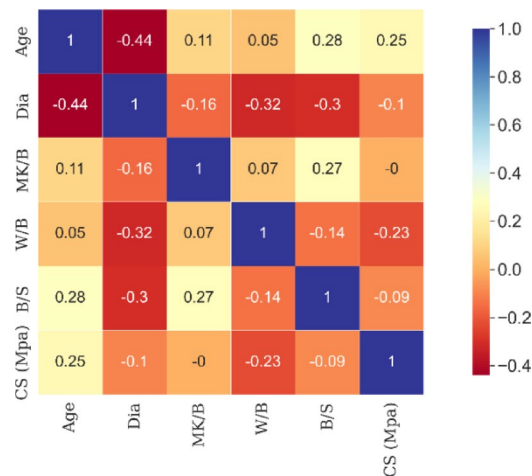
Fig. 9. Heat map showing correlation between variables used in this study.

Algorithm	Parameter	Search space	Optimal value
GEP parameters	Constants per gene	2 to 20	10
	No. of genes	2 to 10	5
	Linking function	Addition, subtraction, multiplication division	Addition
	Head size	5 to 50	20
	No. of chromosomes	5 to 150	80
	Functions	$+, -, \times, \div$, sqrt, ln, x^2	$+, -, \times, \div$, sqrt, ln, x^2
MEP parameters	Number of Generations	500 to 2000	1500
	Subpopulation size	10 to 200	100
	Runs	2 to 15	10
	Crossover probability	0 to 1	0.8
	No. of subpopulations	100 to 2000	1000
	Functions	$+, -, \times, \div$, sqrt, x^2	$+, -, \times, \div$, sqrt, x^2
	Code length	20 to 200	100

Table 4. Hyperparameter setting of MEP and GEP algorithms.

Model development & authentication

The first step in the ML robust model's development is finding a set of variables which will result in the highest accuracy of the algorithms. This process is named as hyperparameter tunning. This step is crucial because every algorithm needs specific parameters which must be finalized through rigorous testing. For this purpose, a thoroughly crafted methodology was developed in this study. The hyperparameter selection process of each algorithm is given under.

MEP model development

MEP algorithm implementation for developing an equation to predict C-S of MK-based cement mortars was done with the help of MEPX 2021.05.18.0. software which enabled to efficiently execute MEP methodology by changing different hyperparameters. The MEP parameters tunned in MEPX 2021.05.18.0. software which affected the performance and accuracy of the algorithm included number of generations, probability of crossover, length of code etc. Same as for GEP model, the preliminary values for the parameters were selected after considering previous studies[84] and then extensive trials were performed varying the value of each hyperparameter across a wide range until the optimal values were found for each parameter. These optimal values which yielded the highest model accuracy are tabulated in (Table 4). Each of these parameters signify a key function of the MEP algorithm. The choice of the optimum value for each of these variables is of great importance to ensure the development of a robust model. For instance, crossover probability is used to show the proportion of parents that will go under the process of genetic crossover. It is selected as 0.8 in this study indicating that 80% of the total parents present in the population would go under crossover. Similarly, population size and number of populations deal with the total amount of random chromosomes generated initially and hence these two parameters control the computational power and time required by the algorithm in addition to controlling the accuracy of the resulting equation. As a general rule, increasing these values result in an increase in the accuracy, but selecting abnormally larger values will lead to overfitting of the algorithm[85]. Furthermore, code length needs

to be carefully determined since it affects the length and complexity of the resulting mathematical statement. Table 4 shows the MEP parameters which result in highest accuracy in current study. The functions used for development of MEP equation again include only simplest and easily calculated arithmetic functions to keep the final expression simple as in GEP algorithm.

GEP model development

To develop the MK-based predictive model using GEP algorithm, a software package called GeneXpro Tools (version 5.0) was utilized. Prior research suggestions were used to choose the initial set of these parameters[86] and these were changed across a wide range until a set of parameters resulting in highest accuracy was reached. In this context, the most important GEP parameters involved in this study include gene and chromosomes, head size, and linking function etc. The chromosomes numbers and size of dead influences the complexity of the algorithm. Raising the values of these variables might result in a rise in accuracy of the algorithm but it also calls for more computational power and beyond a certain point, increasing these values can cause overfitting[87]. In the same way, gene number means the number of sub-expressions made by the algorithm which will be joined by linking function to make the equation. Thus, these parameters have their own importance and their optimizing the value of each parameter is significant for convergence of GEP for optimal solution. Therefore, the number of genes, head size, and chromosomes were wisely chosen such that overfitting in the algorithm is not observed and also maintaining the accuracy of the resulting equation. In short, the GEP development process involved varying different hyperparameters of GEP and monitoring their effect on the accuracy of the results. The parameters which resulted in highest accuracy is given in (Table 4). For current study, the number of genes were chosen as 5 which is an indicator that the resultant will have 5 sub-expressions which will be joined by the already selected linkage function to form the complete equation. Table 4 further shows that the logarithmic and square root functions, as well as comparatively basic functions are selected to be included in the final equation to maintain the equation's simplicity and compactness.

XGB, adaboost, and BR model development

Although the hyperparameter tunning of GEP and MEP was done manually which involved extensive testing using a range of values, the hyperparameter tunning of XGB, BR, and AdaBoost was done in a much efficient way. In particular, a widely used approach known as grid search was utilized for this purpose which is available in scikit-learn library of python in the form of GridSearchCV function[88]. It is noteworthy that not only the hyperparameter optimization, but the loading of data, splitting, model training and then testing it on unseen data were done using python programming language. The data was split into 80:20 ratio for training and testing of algorithms respectively by employing test_train_split function in Anaconda. The gird search approach was utilized for parameter tunning since it is most efficient and most used approach to tune the parameters of these ML models[89]. This approach basically involves checking every possible combination of all the parameters involved in model development. While keeping all other parameters constant, only one parameter is varied across a range to find its optimal value. This process is repeated for each parameter which results in finding the optimal value of each parameter. Thus, the same methodology was used to find the optimal values of hyperparameters of XGB, AdaBoost, and BR algorithms whose result is given in (Table 5). These are the parameters which resulted in the highest accuracy of the respective algorithms.

For XGB, AdaBoost, and BR algorithms, there is one common parameter called n-estimators which specifies the total DTs number (base learners) made by the model for prediction purposes. However, XGB and AdaBoost also has a parameter known as maximum depth which is used to control the depth of each DT formed in the search space of the algorithm. These parameters are important to tune carefully since the accuracy of algorithms like XGB and BR etc. depends greatly on these parameters. Moreover, XGB and AdaBoost has an additional parameter known as learning rate. In simple words, this learning rate is a shrinking factor applied to the predictions of each new tree added by the algorithm to reduce the rate of fitting of algorithms because by virtue of the sequential adding of trees for reduction of the residual, XGB and AdaBoost can fit quickly to the training set[90]. However, if this fitting is left unchecked, it can also lead to overfitting[91]. Thus, this shrinking factor makes sure that algorithm fitting is done slowly to reduce any potential overfitting. It is recommended in the literature to keep its value between 0.01 and 0.1 so that the new trees addition is not performed swiftly[92]. Hence, 0.01 learning rate was chosen for both XGB and AdaBoost algorithms in current study as shown in (Table 5).

	Hyperparameters	Search space	Optimal value
XGB parameters	n-estimators	10 to 100	40
	Max. depth	2 to 20	5
	Learning rate	0.01 to 0.1	0.01
AdaBoost parameters	n-estimators	10 to 200	100
	Max. depth	2 to 20	8
	Learning rate	0.01 to 0.1	0.01
BR parameters	n-estimators	10 to 200	80

Table 5. Hyperparameter setting of XGB, BR, and adaboost algorithms.

Model authentication criteria

After the training of algorithms using the selected hyperparameters, the next step is to validate their findings and access their accuracy using some error evaluation metrices. To make sure that model predictions meet the requirements for being categorized as acceptable models, this error evaluation will be carried out throughout both the training and testing stages of the algorithms. Table 6 provides a summary of error evaluation criteria used to evaluate the ML models. Each of the employed metric is used to measure a specific quality of the ML algorithm as given in the table. Also, the mathematic formulas to calculate these error metrices along with the criteria to be classified as acceptable is given in the table. The formulae utilize "n" to denote the total number of data points, and "x" and "y" to represent the actual and model-generated values, respectively.

Results and discussion
GEP result

Owing to its grey-box nature, the GEP algorithm provided an empirical equation to predict the output unlike other black-box algorithms. However, the equation wasn't directly given as an output, rather several sub-expression trees were generated by the algorithm which were combined by addition function to get the final expression given by Eq. (7). In particular, five sub-expression trees indicated by $T_1, T_2, \ldots, T_5$ were generated by the algorithm since the number of genes were selected as five. The sub-expression trees were decoded in five

Metric	Significance	Formula	Criteria	Reference
Mean absolute error (MAE)	This reflects the average deviation of the actual values from the predicted ones. Most commonly used metric for accuracy assessment. Its value should be The value of as close to zero as possible.	$\frac{\Sigma\ \lvert x - y\rvert}{n}$	Close to zero	[93]
Root mean square error (RMSE)	Another widely used metric for accessing ML models. It gives greater weight to bigger mistakes by using the squares of the residuals before calculating the mean and used as an indicator of larger errors. Should be kept as minimum as possible for a good model.	$\sqrt{\dfrac{\sum (x - y)^2}{n}}$	Close to zero	[94]
Coefficient of determination (R^2)	It serves to measure the general accuracy of models based on regression. It cannot be used as a sole indicator of accuracy since it is unaffected by dividing or multiplying the result by a constant. Generally, R^2 value greater than 0.8 is considered acceptable.	$1 - \dfrac{\sum (x-y)^2}{\sum (y-y_{\text{mean}})^2}$	$R^2 > 0.8$	[95]
a20-index (a20)	A newly introduced metric employed to evaluate the deviations of the predictions. It quantifies the proportions of forecasting which deviate more than + 20% and – 20% from the actual values. Its value should be equal to 1 for an ideal model.	$\frac{n20}{n}$	Close to 1	[96]
Performance index (PI)	It simultaneously evaluates both R and the relative root mean squared error (RRMSE). PI value less than 0.2 is widely used as a threshold for overall model accuracy. However, it is advised to have PI as close to zero as possible.	$\frac{RRMSE}{1+R}$	$PI < 0.2$	[97]
Objective function (OF)	OF combines RRMSE, correlation, and data points in training and testing sets. It is used to check the performance of ML models as a whole. It should also be less than 0.2 for a good model.	$\left(\frac{n_{\text{Training}} - n_{\text{Testing}}}{n}\right) PI_{\text{Training}} + 2\left(\frac{n_{\text{Testing}}}{n}\right) PI_{\text{Testing}}$	$OF < 0.2$	[98]

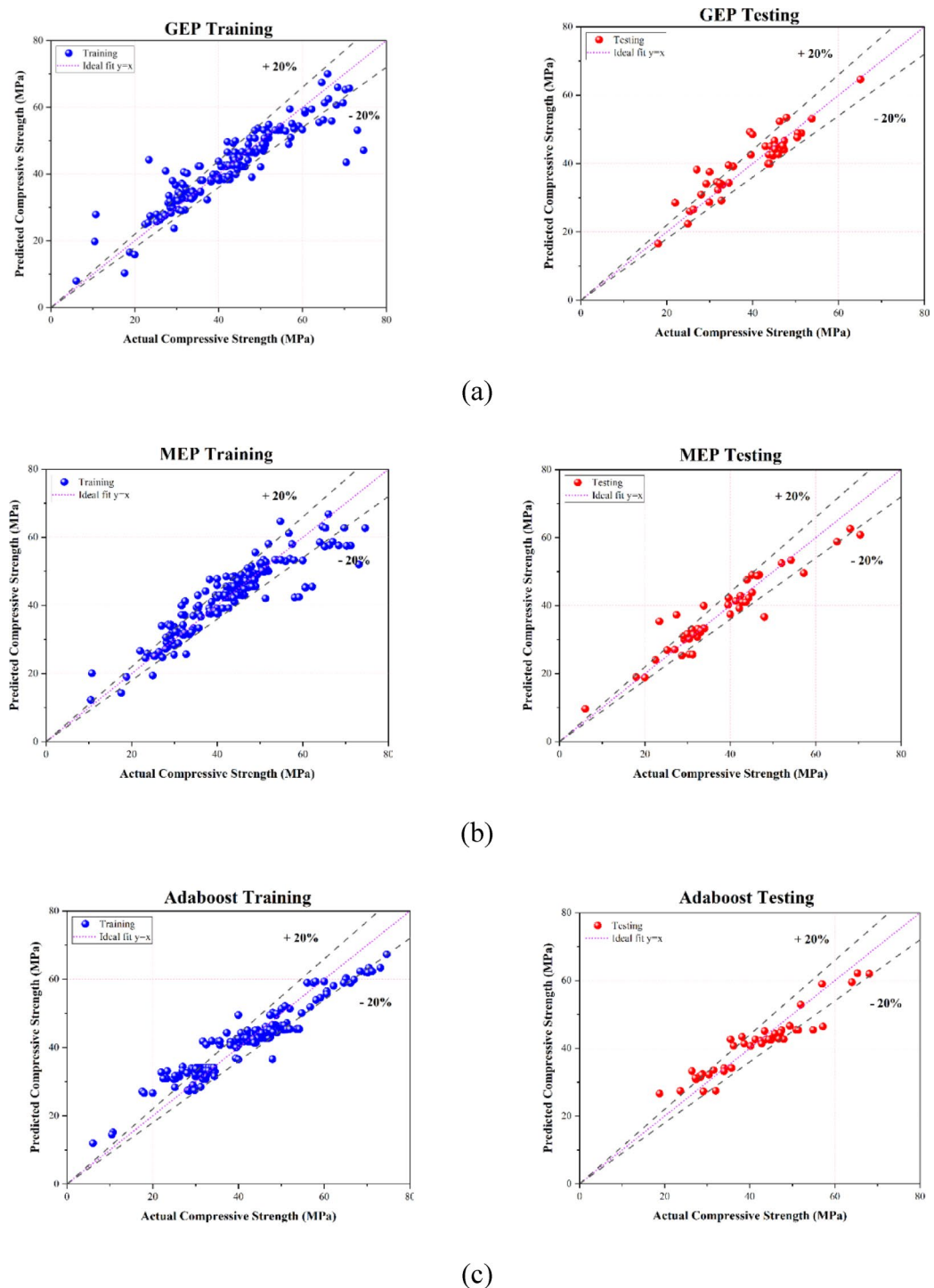**Table 6.** Summary of error evaluation criteria used in the study.

Fig. 10. Scatter plots of developed models for training and testing phases; (**a**) GEP, (**b**) MEP; (**c**) AdaBoost; (**d**) BR; (**e**) XGB.

parts, and the sub-expressions were summed to create the complete equation. The accuracy of GEP equation has been visualized by means of two separate scatter plots for training and testing datasets in (Fig. 10). It should be noted that in the training scatter plot, some points are lying beyond the error lines indicating that at some points, the error between predicted values and original values is more than 20%. However, the points going beyond error lines are comparatively lower in the testing data plot. Moreover, the result of performance evaluation of GEP is also displayed in Table 7 which shows the computed values for error metrices for both phases of training and testing sets. The R^2 value of the GEP is 0.83 for the training session and 0.84 for the test session, both of which are greater than 0.8. MAEs for training and test were 3.33 MPa and 2.87 MPa respectivelyS which shows that GEP equation generated predictions having an error of 3 MPa on average. The RMSE value is 3.8 for testing indicating that the algorithm didn't yield any abnormally larger errors in the testing phase. The PI value, which
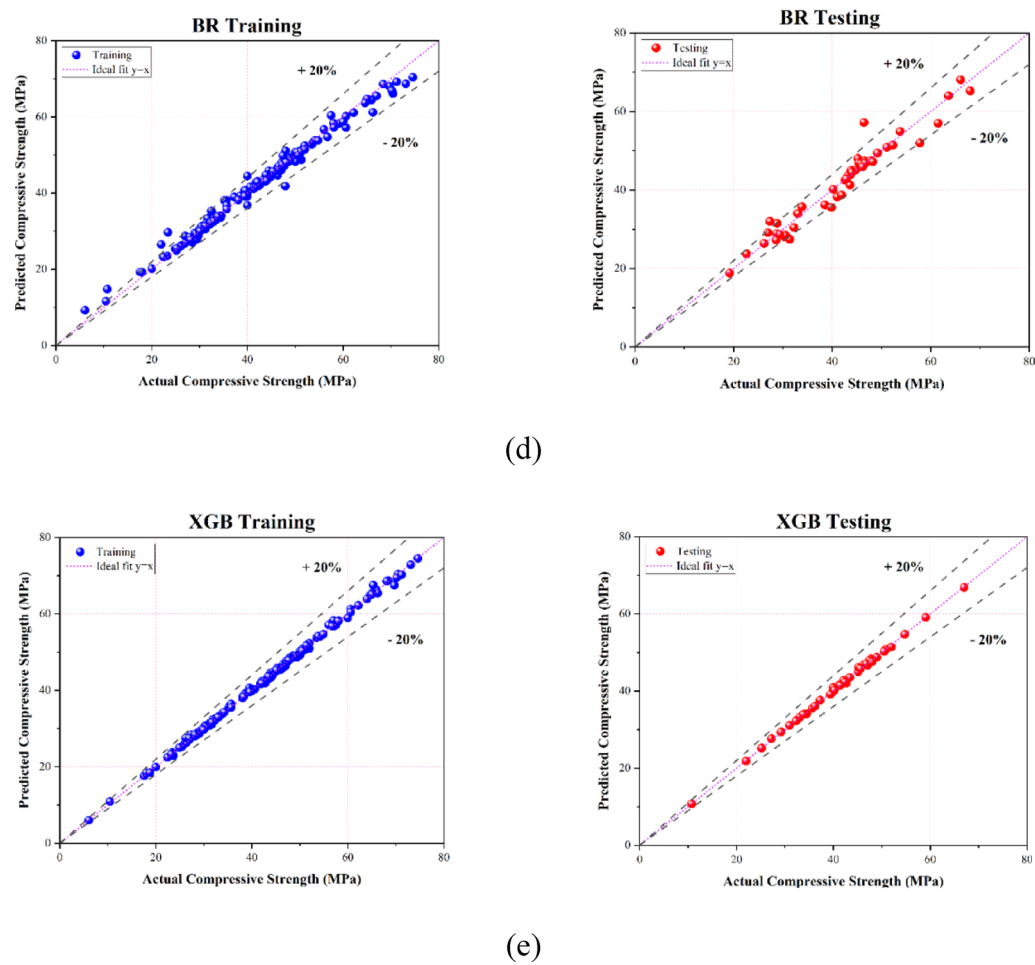
(d)

(e)

Fig. 10. (continued)

	GEP		MEP		AdaBoost		BR		XGB	
	Training	Testing	Training	Testing	Training	Testing	Training	Testing	Training	Testing
MAE	3.33	2.876	3.20	2.86	3.510	3.485	0.879	1.73	0.30	0.257
RMSE	5.24	3.832	4.850	4.14	4.37	4.18	1.45	2.59	0.450	0.347
R^2	0.838	0.846	0.836	0.8975	0.876	0.861	0.986	0.946	0.998	0.998
A20	0.92	0.93	0.9144	0.8936	0.91	1	1	0.98	1	1
PI	0.064	0.0500	0.058	0.056	0.053	0.051	0.017	0.031	0.005	0.0042
OF	0.0587		0.055		0.0521		0.022		0.00467	

Table 7. Summary of error evaluation of developed models.

combines RRMSE and R to indicate overall model performance is 0.064 for training set and 0.050 for testing which are very near to zero. Furthermore, the OF value of GEP is 0.0587 which again indicates that the GEP developed equation can be utilized practically for forecasting of C-S of MK-based cement mortars.

$$C - S \ (MPa) = T_1 + T_2 + T_3 + T_4 + T_5 \tag{7}$$

$$T_1 = x_1 \times \left[\ln \left\{ \frac{10.8342}{\{3.50 + \ln \left(\sqrt{\frac{(x_4 \times (\ln x_1))^4}{\sqrt{x_4} \times 5.239}} \right)\}^2} \right\} \right]$$

$$T_2 = \frac{\ln \left[\left\{ \frac{-1.488 x_1 + 2.9596}{x_3} \right\}^2 \right]}{(x_3)^2}$$

$$T_3 = \ln[x_2 + \{ \frac{\left\{\frac{x_4}{x_4 - x_1}\right\} / \left\{2.64x_3 + (\ln x_4)^2\right\}}{\left(\left\{\frac{x_4}{-0.539}\right\} / \left\{x_4 - x_1\right\}^2\right)} \}]$$

$$T_4 = 7.288 \times \left[\sqrt{\sqrt{x_0 \times \left[\left\{ \frac{\left\{\left(\frac{1.680/x_4}{x_0 - 1.680}\right) \times \left(\frac{x_4 - x_1}{x_3 - 6.95}\right)\right\} + x_3 - x_0}{x_0} \right\}^2 \right]}} \right]$$

$$T_5 = x_3 \times [\{(\frac{\sqrt{x_4 / \sqrt{\frac{0.585}{\ln\left(\frac{x_0 + 0.585}{x_0}\right)}}}}{x_4}) + \frac{x_1}{-2.18}\} \times -12.10]^2$$

MEP result

Equation (8) provides the output of the MEP algorithm as an empirical equation. It is worth mentioning that this equation wasn't directly yielded by MEP algorithm. Rather, the algorithm gave a python code which was used to get Eq. (8). Also, the predictive abilities of this equation to forecast C-S for both datasets of training and testing has been given in scatter plots shown in Fig. 10b from which it is evident that the MEP predictions lie closer to the best fit line. Also, fewer points lie beyond the ±20% error lines indicating that there are very few MEP forecasted values deviate more than 20% from the C-S actual values. Also, the validation of MEP model has been done by error metrices in (Table 7). Notice the R^2 value of MEP in training phase (0.836) and in testing phase (0.8975) from the table. Both of these values are greater than 0.8 which is the lowest recommended value for a good model. After R^2, it can be seen that MAE value of MEP is 3.20 MPa in training and 2.86 MPa in testing set which shows the minimal difference between predicted and the actual values. The trend is same for RMSE values too. Regarding a20-index, it is 0.9144 for training and 0.8936 for testing indicating the robustness of MEP. Furthermore, the overall performance indicators; PI and OF are also substantially lesser than 0.2. The OF value of MEP is 0.055 which is significantly lesser closer to zero indicating its robustness.

$$C - S \ (MPa) = \left\{ \frac{\sqrt{x_0 - x_4}}{C} + D - 2B + 2 \right\} + \left\{ \left[\sqrt{x_0 - x_4} \right] - \left[\frac{x_1 + (x_4 x_1)}{x_2 + \frac{x_4}{x_3}} \right] \right\} - \left\{ \frac{A}{x_0} \right\} \qquad (8)$$

$$A = \frac{1 + (x_4 x_3)}{x_4 x_1 - x_3}$$

$$B = (x_4 x_1) \times 2(x_3 + 1)$$

$$C = \frac{B - 1}{(x_4 x_1)^2} - (A - A^4)$$

$$D = \left[\left\{ \frac{2(x_3 + 1)}{3} \right\}^2 \right] + \left[\sqrt{x_0 - x_4} \times \frac{x_4}{x_3} \right] - \frac{\sqrt{x_0 - x_4}}{C}$$

AdaBoost result

The AdaBoost algorithm failed to provide an equation for output prediction. However, the scatter plots for AdaBoost training and testing data is shown in (Fig. 10c). From the plots, it is explanatory that compared to MEP and GEP, very few data points lie out of the 20% error lines. It shows that most of the AdaBoost predictions are accurate and have less than 20% deviation from the real-life C-S values. The result of AdaBoost error evaluation is shown in Table 7 from which it is evident that the R^2 values of AdaBoost for both data partitions is greater than 0.8. The R^2 value is 0.876 for training while it is 0.861 for testing phase. Also, the MAE of AdaBoost is 3.510 MPa in training and 3.45 MPa in testing which is slightly larger than the those of MEP and GEP. Additionally, the a20-index value of AdaBoost is 0.91 in both training and testing phases indicating its robustness. Moreover, the training set PI value of AdaBoost is 0.053 and the testing set PI value is 0.051 both of which are less than those of MEP indicating its accuracy over the evolutionary algorithm models. Furthermore, the indicator of model's overall efficiency; the OF value of AdaBoost is 0.0521 which is substantially less than 0.2 and very close to zero. Also, it is less than GEP and MEP which shows its superior accuracy in comparison to GEP and MEP. Therefore, the AdaBoost model can be used practically for C-S prediction of MK-based cement mortars with more than 86% accuracy on unseen data. The model exhibited a good accuracy overall (greater than 0.8) and acceptable performance indicator values. However, it is still lower accuracy than other black box models including XGB and BR which may be due to the model being sensitive to noisy data or being unable to accurately pick the complex relationship between inputs and output.

BR result

The scatter plots and in Fig. 10 depict the actual versus predicted values by BR algorithm for both data distributions from which it is evident that almost all of the points lie above the linear line. Some values deviate from the ideal fit line but are still lying within the ±20% error range which shows the excellent accuracy of BR algorithm. Also, a significant improvement in prediction accuracy offered by BR algorithm is evident from its training a20-index

value which is perfectly equal to 1 while it is 0.98 for testing data. It has been shown in Table 7 along with error metrices of other algorithms. These error metrices given in Table 5 verify that the predicted values of BR do not vary greatly from the actual experimentally determined values. The R^2 of BR training is 0.98 while it is 0.94 for testing. Another important thing to mention is the minimal MAE value of BR in training stage (0.8 MPa) and in testing phase (1.73 MPa). Both values are considerable lesser than the MAE values of other algorithms like MEP and GEP. The same is true for RMSE values of BR. Regarding the performance indicators of BR algorithm, the PI values are remarkably lesser than 0.2 for both BR training and testing. The PI value is 0.017 for training and 0.031 for testing. Similarly, the OF value is 0.022 which is the second lowest after XGB algorithm. It shows that BR proved to be very accurate for peak strength prediction of C-S of MK-based cement mortars.

XGB result

The XGB algorithm too like BR and AdaBoost didn't yield an empirical equation as an output due to its black-box nature. Thus, its results are interpreted and shown in the form of scatter plots in (Fig. 10e) and the results of error evaluation of XGB are also given in (Table 7). The superior accuracy of the XGB model is shown from the scatter plots where all predictions are lying seamlessly on the line. No point deviate from the ideal fit line and all points lie within the ± 20% error range. Due to this reason, the a20-index of XGB model is perfectly 1 for both data splits. The R^2 value of XGB is equal to 0.998 in both phases of training and testing showing that XGB has more than 99% accurate to predict the output. The RMSE and average deviation of XGB are also significantly lesser than the values of all other algorithms. The training MAE of XGB is at the minimal value of 0.30 MPa while
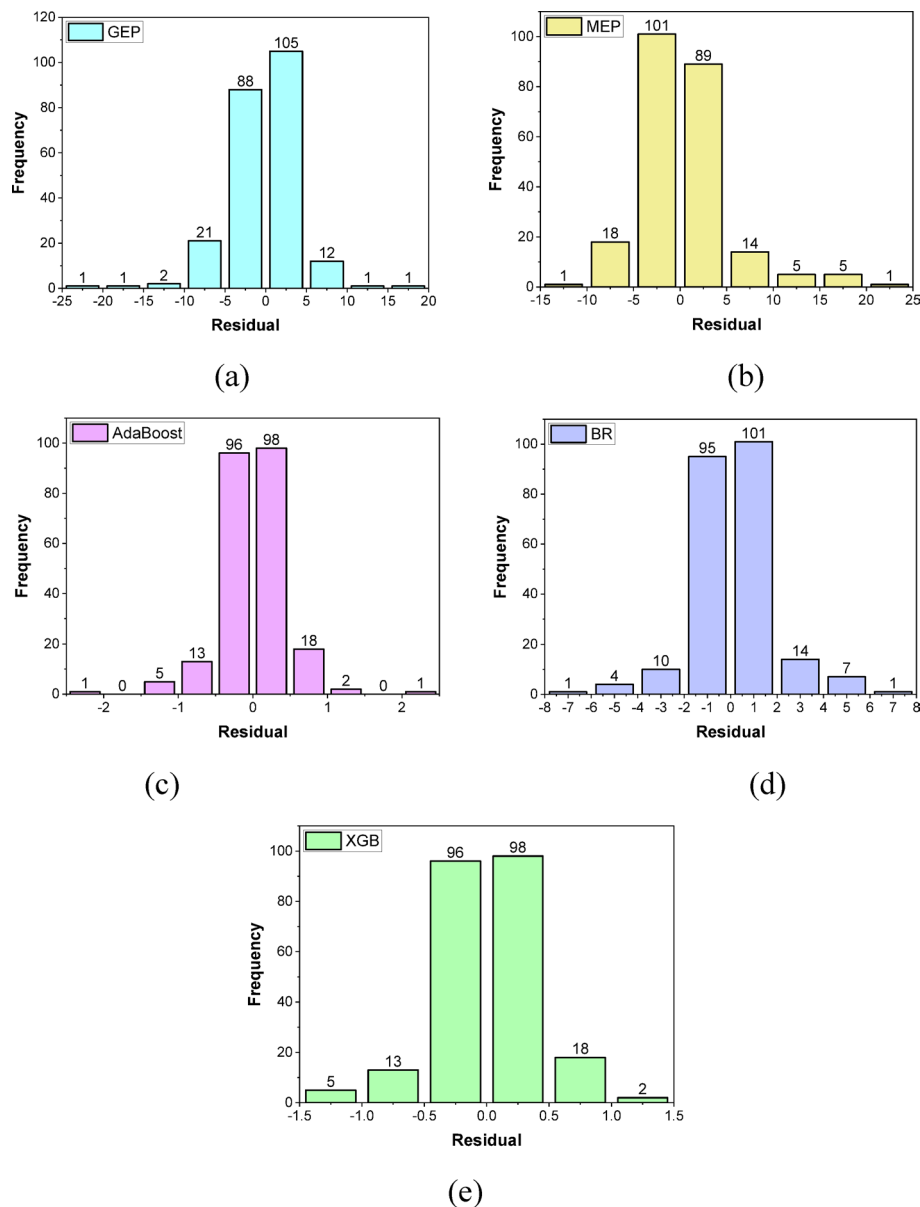

Fig. 11. Residual distribution of developed models; (**a**) GEP, (**b**) MEP; (**c**) AdaBoost; (**d**) BR; (**e**) XGB.

it is 0.257 MPa for testing set. These values are significantly lesser than all other algorithms. Moreover, the PI value of XGB training is 0.005 while it is around 0.042 for testing sets which shows the overall efficiency of XGB. These values are again significantly lesser than PI values of its counterparts. Finally, XGB OF is 0.0467 which in addition to being considerably less than 0.2 is less than the OF values of all other algorithms. Therefore, it can be concluded that the performance of XGB surpasses that of all other algorithms used in the current study.

Residual comparison of models

Based on the validation criteria defined in Sect. 6.4, Table 7 proves that the developed are acceptable and the values of all error metrics lie within the literature's range. However, to draw a comparison between the algorithms based on their errors made for both training and testing data (the combined dataset), Fig. 11 shows the residual distribution plots for the employed algorithms. From Fig. 11a, notice the residual distribution of GEP algorithm which shows that most of the GEP residuals (226 out of 234) lie within the range of -10 to 10. It means that while using the GEP-based equation, the maximum difference between actual and anticipated outcome will be 10 MPa most of the times. While, from Fig. 11b which gives residual distribution for MEP, it can be seen that 222 out of 234 residuals lie in the range −10 to 10. This indicates slightly more residuals beyond the range −10 to 10 for MEP algorithm. However, this range reduces significantly for the AdaBoost algorithm in which more than 95% of the residuals (232) residuals lie in the narrowest range of −2 to 2. For BR algorithm, in the range −4 to 4, 220 out of 234 residuals lie and there is no residual lying beyond the range −8 to 8. Finally, for XGB algorithm, all of the residuals lie in the range −1.5 to 15. It means that at no point the XGB algorithm will give an error more than 1.5 MPa from the actual values. In particular, 225 residuals are lying in the smallest range of −1.0 to 1.0. It again indicates that XGB has the smallest errors and has the superior accuracy compared to its rivals.

External validation of models

To further validate the model's accuracy, this study utilizes several external validation metrices in addition to the model validation by error metrics, scatter plots, and residual plots etc. The summary of the results of external validation is given in (Table 8). Moreover, the acceptable ranges of the external utilized checks for models to be classified as valid are also given in (Table 8). The external validation values for all algorithms were calculated using the whole dataset without any partition to confirm that the models are overall valid according to the external checks. The parameters P' and P depict the slope of regression lines passing through the origin and both of these values must lie within 0.85 to 1.15 for a model to be valid. Also, R_o^2 shows the correlation between real life and anticipated values. It is recommended that its value must be close to 1 for a good model. Similarly, R_m considers the combined effect of R_o^2 and R^2. From Table 8, it can be seen that overall, all models qualify the checks and the values of all metrices lie within the suggested range. The XGB algorithm, however, exhibited the highest values of all external validation metrices. The correlation (R value) is the highest for XGB (0.999) followed by BR, AdaBoost, MEP, and lastly GEP. Similarly, R_m which measures the difference between R^2 and R_o^2 is also highest for the XGB model. The R_m value of XGB is 0.899 followed by 0.895 for BR, 0.787 for AdaBoost, 0.574 for MEP, and it is 0.533 for GEP which is marginally larger than the lowest recommended value (0.5). The trend is same for almost all other external checks too. Therefore, based on the values from Table 8, it can be inferred that all of the ML algorithms presented in this study are efficient and robust since they passed all external validation checks in addition to their validation through error evaluation, scatter plots, and residual analysis.

Best model selection

The current study's findings demonstrate that the algorithms exhibit acceptable performance based on residual comparison and error evaluation, with XGB algorithm emerging as the most effective among those employed. However, a more robust method for model comparison is the Taylor plot, which evaluates models simultaneously based on their standard deviations and correlation coefficients (R values)[101]. In a Taylor plot, proximity to the

No.	Reference	Expression	Criteria	GEP	MEP	AdaBoost	BR	XGB
1	[99]	$P = \dfrac{\sum_{i=1}^{n}(x_i \times y_i)}{\sum_{i=1}^{n}\left(x_i^2\right)}$	$0.85 < P < 1.15$	0.980	0.977	0.972	0.994	0.999
2	[99]	$P' = \dfrac{\sum_{i=1}^{n}(x_i \times y)}{\sum_{i=1}^{n}\left(y_i^2\right)}$	$0.85 < P' < 1.15$	1.0066	1.010	1.0188	1.004	1.000
3	[100]	$R_m = R^2 \times (1 - \sqrt{|R^2 - R_o^2|})$	$R_m > 0.5$	0.533	0.574	0.787	0.895	0.899
4	[80]	$R_o^2 = 1 - \dfrac{\sum_{i=1}^{n}\left(y_i - x_i^o\right)^2}{\sum_{i=1}^{n}\left(y_i - y_i^o\right)^2}$, $x_i^o = P \times y_i$	$R_o^2 \approx 1$	0.970	0.958	0.928	0.978	0.989
5	[80]	$R_o'^2 = 1 - \dfrac{\sum_{i=1}^{n}(x_i - y_i^o)^2}{\sum_{i=1}^{n}(y_i - y_i^o)^2}$, $y_i^o = P' \times x_i$	$R_o'^2 \approx 1$	0.996	0.991	0.990	0.990	1
6	[100]	$R = \dfrac{(n\sum y - (\sum x)(\sum y))}{\sqrt{(n\sum x^2 - (\sum x)^2)(n\sum y^2 - (\sum y)^2)}}$	$R > 0.8$	0.915	0.924	0.9539	0.990	0.999

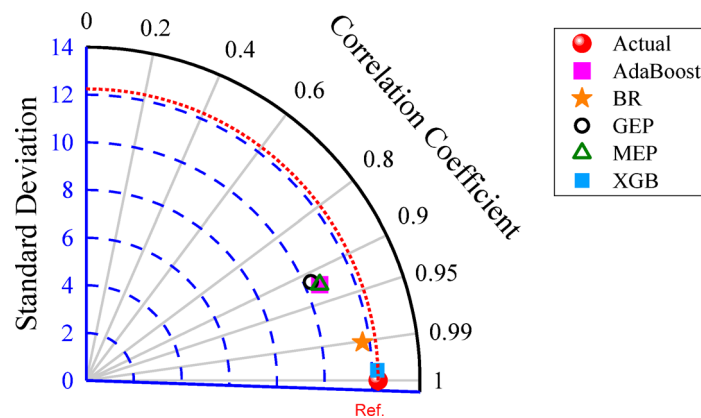**Table 8**. Summary of external validation criteria employed in the study.

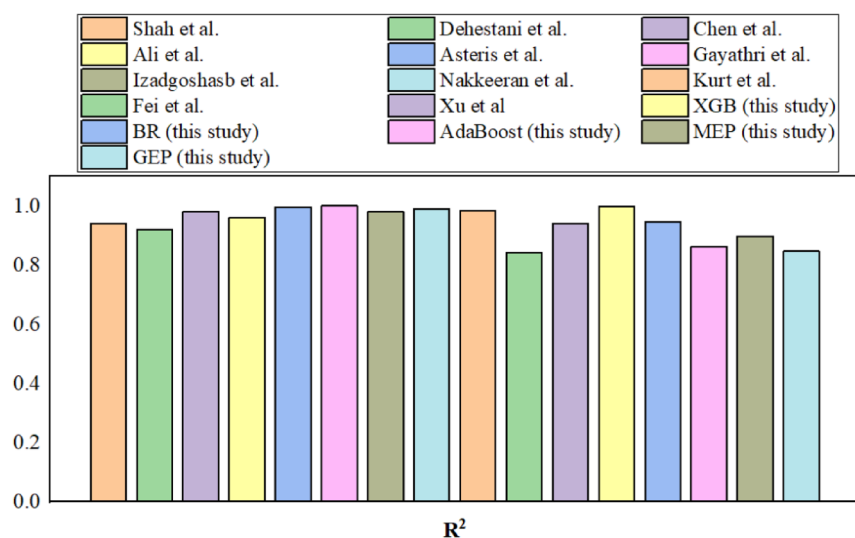**Fig. 12**. Taylor plot-based comparison of models.

Fig. 13. Accuracy comparison of ML models with existing studies in literature.

point of actual data indicates greater reliability of the algorithm's predictions and vice versa. The standard deviation of predictions for each algorithm has been calculated, and the resulting data points are plotted on the taylor plot as shown in (Fig. 12). Notably, the XGB algorithm, depicted by a blue square, is positioned closest to the actual data, underscoring its superior performance and indicating that the outcomes generated by XGB are nearly equivalent to the experimental C-S values. Adjacent to the XGB point, the BR algorithm is positioned closest to the actual data, indicating it ranks as the second-best performer following XGB. In contrast, the other algorithms like MEP, GEP, and AdaBoost exhibit performance levels that are quite similar, as evidenced by the aggregation of their points in close proximity to one another. Specifically, the MEP point is positioned directly above that of AdaBoost, highlighting their nearly identical performance, while the GEP point is slightly lower, suggesting a marginally lesser standard deviation when compared to MEP and AdaBoost. Consequently, based on the analysis derived from the taylor plot, the order of algorithm accuracy is established as follows: XGB > BR > AdaBoost > MEP > GEP. Given that the XGB model consistently demonstrates the highest level of accuracy, it is concluded that this algorithm will be utilized for subsequent explanatory analyses.

Comparison with existing models

In addition to drawing a comparison between algorithms employed in current study, it is also important to compare their accuracy with the accuracy of already existing models to predict the C-S of cement mortars having diverse types of secondary cementitious materials. Thus, a comparison between models in this study and already existing models described in Table 1 has been made on the basis of R^2 value and the result are shown in the form of a bar graph in (Fig. 13). From Fig. 13, it can be seen that the perfect R^2 value of 1 has been achieved by the study of Gayathri et al.[48] followed closely by the XGB model presented in this study. XGB model depicted accuracy higher than all previous studies conducted previously except that of Gayathri et al.[48] and all ML algorithms investigated in current study with R^2 value of 0.998. The XGB model was closely followed by Asteris et al.[47] having R^2 equal to 0.996 and 0.990 depicted by Nakkeeran et al.[51]. After that, the study conducted

by Chen[41] which predicted flexural strength of cement mortar having RHA had 0.980 R^2 value. Next to it is the accuracy of Kurt et al.[52] with 0.98 and Ali et al.[45]. However, the least R^2 value (0.846) was exhibited by the GEP model developed in current study while the accuracy of all other algorithms lies in between the XGB and GEP models. In addition to the superior accuracy of XGB model as compared to previous studies, the model presented in this study also offers the significant advantage that it has been made available in the form of a GUI which can be used to practically implement the findings of the XGB model. This feature was not available in the previously conducted studies. Another advantage is that this study presents explanatory techniques on the developed XGB model which further helps to validate its findings and reinforce its contribution.

Interpretable machine learning

In order to provide transparency to the prediction-making process of the algorithms, post-hoc interpretable techniques are widely used which impart interpretability and transparency characteristics to the black-box ML models. The application of various interpretable techniques on ML models is also crucial since they enhance the understanding of black-box models and enable non-technical users and different stakeholders to understand the predictions made by the algorithms. Thus, this study employs three types of techniques known as SHAP (Shapely Additive Explanatory Analysis), ICE (Individual Conditional Expectation), and PDP (Partial Dependence Plots) analyses on the XGB model since it is consistently highlighted in Sect. 7.6 to 7.9 that XGB is the most effective technique employed in current study. SHAP was selected for its ability to deliver global and local interpretability for sophisticated machine learning models. Based on cooperative game theory, SHAP provides a robust foundation for determining the effect of each input feature on the model's predictions. Moreover, SHAP values are consistent and readily visualized, making them accessible to non-technical stakeholders. PDPs further elucidate the relationship between a feature and the predicted outcome, while ICE plots illustrate how the prediction changes for an individual instance as a feature varies, providing granular insights into the model's behaviour. All of these models were used to get a clear understanding of effect of different variables on prediction of output.

Shapely analysis

Taking its inspiration from the coalition game theory, the shapely analysis was introduced by Lundberg and Lee[20] as a method to interpret black-box ML models. At its core, it assumes the input features as players and the predictions of the model are assumed to be payouts and thus calculates the contribution (significance) of each player (input) in payouts (predictions). The TreeSHAP version is widely used for interpretation of ML models built on trees (like XGB and AdaBoost etc.). Thus, the TreeSHAP version was employed in Anaconda using python to interpret the trained XGB-based predictive model for C-S of MK-based cement mortars. There are two sub-types of SHAP analysis known as global interpretation and local interpretation. SHAP global interpretation talks about the essence of the inputs as a whole while SHAP local interpretation dissects each individual prediction to give information about how the algorithm reached this particular prediction. The SHAP global interpretation is widely represented by means of a SHAP summary plot (given in Fig. 14). This plot serves two important purposes; it ranks the features based on their SHAP values and also tells which features are positively or negatively correlated with the outcome. A row of dots is present in front of each input on the beeswarm plot with the dots indicating the SHAP values for the particular input variable. The dots range in colour from blue (representing lower input values) to red (representing higher input values). In the summary plot, from top to bottom, the features are arranged in order of decreasing significance in predicting the outcome. Thus, the most influential variables are the ones present at the top and vice versa. Moreover, if there are more red dots on the lower side (negative side) of SHAP plot, it indicates an increase in variable values lead to a decrease in its corresponding SHAP value while the blue plots present on lower side of the plots shows that decreasing variable values causes a decrease in the SHAP values too. From Fig. 14, it can be seen that the feature age is present on the top indicating that it is the most crucial factor to predict the C-S of MK-based cement mortars. After age, W/B is highlighted as the second most essential variable followed by B/S ratio. However, the contribution aggregate diameter and MK/B ratio is the least in anticipation of the outcome. For age, it can be seen that most of the blue dots lie on the negative side on the SHAP values showing that for lower age values, the corresponding SHAP
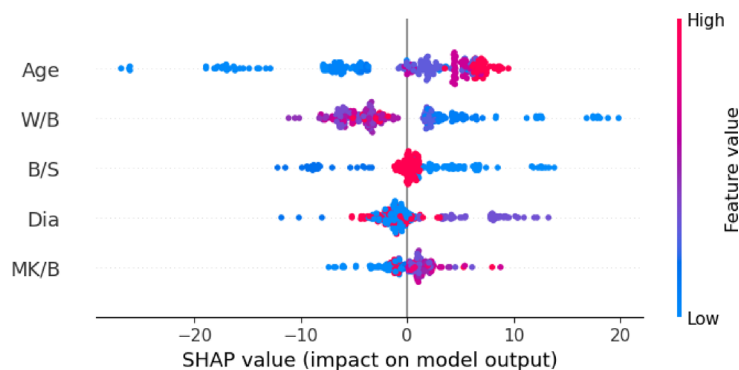

Fig. 14. SHAP global interpretation by means of beeswarm plot to highlight most crucial variables.
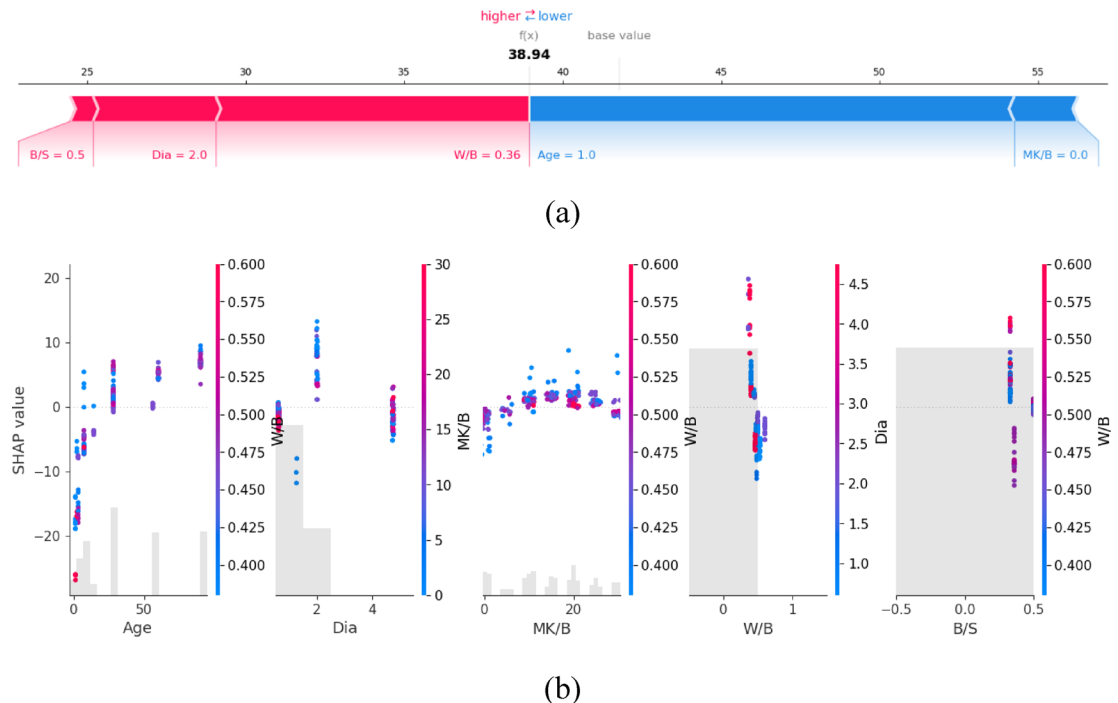
(a)

(b)

Fig. 15. SHAP local interpretation; (**a**) SHAP force plot for C-S value of 38.94 MPa; (**b**) SHAP partial dependence plots.

values are also lower and vice versa. It is also true for features MK/B and diameter in which the red dots lie on the positive side of the SHAP values. However, for W/B ratio, the blue dots lie on the positive side while red dots are lying on the negative side showing that increase in W/B ratio decreases its corresponding SHAP values. For B/S, the red dots are clustered near to the zero line while some blue dots lie on both sides of the SHAP values.

In addition to the SHAP global interpretation to know which variables are more significant in prediction of outcome, the SHAP local interpretation to get insight into how the algorithm generates individual predictions is also important. The most common ways to represent the SHAP local interpretation of a prediction is by using a graph known as force plot which helps to depict the local interpretation of a prediction in a concise and compact manner as shown in (Fig. 15a) for the predicted C-S value of 38.94 MPa. The centre point of the force plot indicates the base value, which is the average of all SHAP values from the predictions. From Fig. 15, it can be seen that the base value in our case is around 41.5 MPa. Each input feature which helps to predict the outcome is represented by an arrow in the force plot with the length and direction of the arrow indicating magnitude and the effect (positive or negative) on the output respectively. The colour coding of the arrows in force plot is done in the same way as in summary plot with blue indicating lower input values and red indicating higher input values. The contribution or significance of each variable is proportional to its size and the predicted value is located at the intersection of blue and red arrows as shown in (Fig. 15a). Notice that the blue and red arrows meet at 38.94 MPa which is the predicted C-S value of MK-based cement mortar. The arrow sizes for features age and W/B are the largest indicating their greater influence compared to other variables. The age value for this particular instance was 1 and MK/B ratio was 0 since they are represented by blue coloured arrows (indicating lower input values). However, dia, B/S, and W/B are represented by red colour since their values are on the higher side for this particular data instance. Thus, by means of this force plot, the user can get insight into the individual predictions knowing what input parameter values were used and what outcome did they yielded.

After carrying out SHAP local and global interpretations by means of beeswarm and force plots respectively, it is also beneficial to check the interdependence between the SHAP values of parameters with change in variable values of another input variable. This can be done by using SHAP partial dependence plots which are given in (Fig. 15b). The five plots in Fig. 15b are for the five input features considered in this study. The x-axis contains the input values of a feature, with the left y-axis displaying the corresponding SHAP values. To check the interdependency, the variable values of another parameter are given along the right-hand side of the plots along y-axis. First of all, it can be seen that as W/B ratio varies from 0.4 to 0.6, the SHAP values of age feature changes from −10 to 10. This shows that the W/B values can greatly impact the SHAP values of age feature. It is because these two features were classified as the most significant ones through the SHAP summary plot. Also, notice that increasing MK/B values from 0 to 30 causes SHAP values of diameter to 0.450 to almost 0.575 in a discontinuous manner. Moving on, the SHAP values of MK/B changes slightly from 12 to close to 20 when W/B is changed across its range of 0.4 to 0.6. Moreover, when diameter is varied from 1 to 4.5 mm, the W/B SHAP values change consistently from 0.450 to almost 0.60. Finally, the last sub-plot in (Fig. 15b) is between B/S and W/B from which it is evident that increasing W/B ratio from 0.45 to 0.55 causes the SHAP values of B/S to go from 2 to 4. This comparatively larger change in SHAP values of B/S with a minimal change in W/B

is due to the reason that W/B and B/S are some of the most contributing factors to C-S of MK-based cement mortars according to the SHAP global interpretation. In addition, a comparison of the SHAP results with the existing literature demonstrates that the feature contributions identified using the SHAP method are consistent with the known behaviour of metakaolin-based materials. This theoretical validation ensures that the XGB model accounts for the most crucial elements influencing the strength of MK-based mortars, enhancing the credibility of its predictions. The experimental findings indicate, for instance, that the influence of metakaolin on the strength at later ages (7–28 days) was more significant than at the early ages which is consistent with results from SHAP analysis[102]. Similarly, it has been identified by previous studies regarding strength prediction of different cementitious materials that age and water-to-cement ratio are some of the most important parameters for predicting the output[103,104]. It is again in accordance with the SHAP results.

ICE analysis

The ICE plots between C-S of cement mortars having MK and the five input variables used in current study are given in (Fig. 16). The plots for five input features are labelled from (i) to (v) as shown in the figure. These plots are developed by keeping all inputs constant and varying only one input across its range. Then, the change in output values with varying input features values is plotted in the form of a curve known as ICE plot. These plots provide an interactive visual way to investigate the relationship of each input feature with the predicted output. Figure 16i the ICE plot for age and it can be seen that the predicted outcome (C-S) generally increased with increase in age. The C-S value showed a consistent increase from almost 20 MPa to more than 50 MPa when age of the specimen was varied from 1 to 90 days. This increase of almost 30 MPa in C-S value over the range of age explains why it was identified as the most significant input feature by the SHAP summary plot. In Fig. 16ii, the plot of MK/B is shown from which it can be seen that C-S slightly increases up to a particular point with increase in MK/B ratio, after which it decreases minimally but there is not a significant change in output value with change in MK/B ratio. It is because MK/B is the least contributing variable to C-S of MK-based cement mortar according to SHAP analysis. In particular, when MK/B ratio was changed from 0 to 15, C-S increased slightly from 39.5 MPa to 43 MPa, after which it dropped again to almost 41 MPa when MK/B further reached 30. Similarly, the relationship between C-S and B/S ratio is depicted by the curve in Fig. 16iii. It is evident from the plot that when B/S ratio was changed from 0.325 to 0.350, the C-S dropped from 52.5 MPa to less than 37.5 MPa. After that, the C-S remain unaffected by increase in B/S ratio up to 0.425. However, when B/S was increased beyond 0.425, the C-S increased slightly to almost 40.5 MPa from less than 37.5 MPa. The relationship between aggregate diameter and predicted outcome is given by Fig. 16iv from which it can be seen that the C-S varies across a narrow range of 37 MPa to 45 MPa when diameter values were changed from less than 1 to more than 5 mm. After a slight initial decrease, the C-S increases from 37 MPa to 45 MPa when diameter value was increased from 1 to 2 mm. After that, the C-S decreased to 41 MPa with diameter reaching 6 mm. This comparatively lesser deviation of C-S with change in diameter values is owed to lesser contribution of diameter in predicting C-S as identified previously. Finally, the effect of change in W/B on the C-S can be seen from Fig. 16v which shows that there is a decreasing relationship between the two variables. The C-S decreased
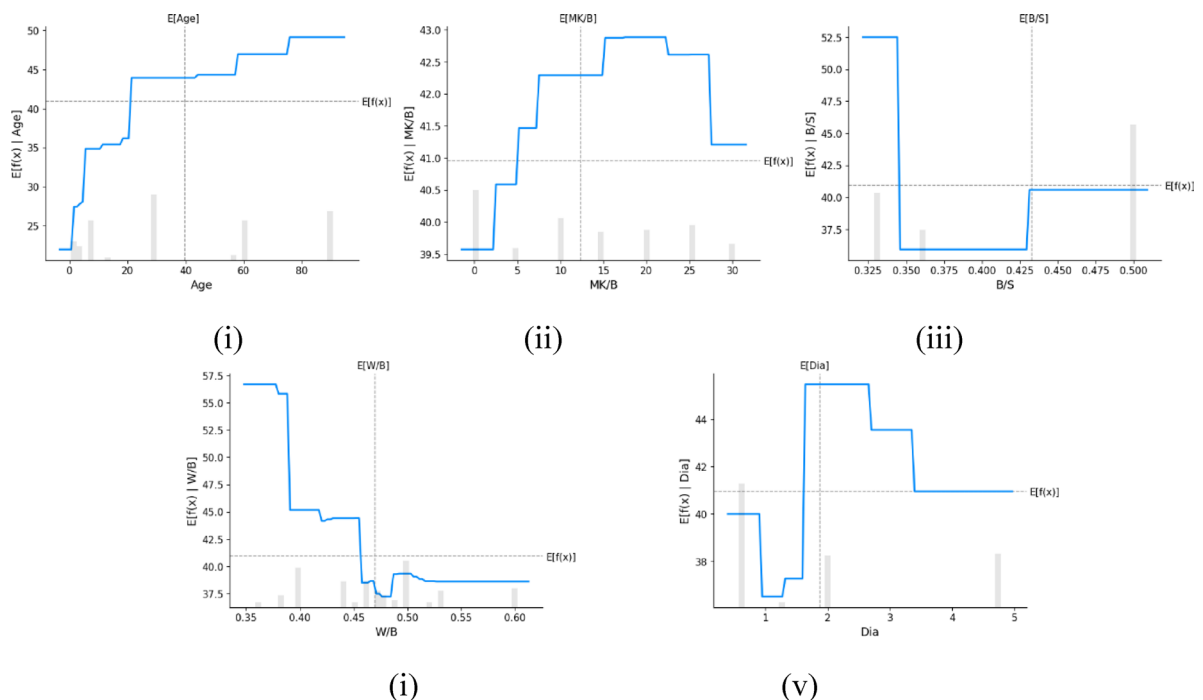

Fig. 16. ICE analysis to depict the effect of input features on output; (**i**) age; (**ii**) MK/B; (**iii**) B/S; (**iv**) W/B; (**v**) diameter.

from 57.5 MPa to a value of 37 MPa when W/B was changed from 0.35 to 0.60. This large change is C-S is due to the fact that the strength of cement mortars is dependent on the W/B ratio to a higher degree. It has also been previously identified in the SHAP summary plot where W/B ratio was the second most influential input to predict the outcome (C-S).

PDP analysis

PDP analysis is used to portray the relationship between a specified number of inputs and the output (prediction). This relationship can be linear, intricate, or monotonic etc. It was first introduced by Friedman and since then, has become a really popular method to check the global interpretation of ML models[105]. It can be used to check how changing a particular variable, or changing two variables simultaneously can affect the output. If the relationship between only one variable and output is investigated, the resulting plots will be known as 1-D PDP plots while if the simultaneous relationship between two inputs and outcome is being monitored, the PDP plots would be called 2-D plots. Since the individual relationship between inputs and outcome has already been covered by means of ICE analysis, this section will focus only on 2-D plots to depict the combined effect of two inputs on the predicted C-S of MK-based mortars.

The partial dependence plots with two variables (2D PDP plots) are shown in (Fig. 17). In these plots, one input feature's values (main variable) are given on the horizontal axis (x-axis) while the other input is represented on the y-axis. The legend given on the right side of the plots is used to show the probability of XGB model yielding high C-S values for different combinations of values of main input variable (on x-axis) and secondary interaction input feature (on y-axis). Also, the more the slope of a contour line on the graph, the higher is the influence of interaction of both variables on the output and vice versa. First of all, notice Fig. 17i which is a 2D PDP interaction plot developed between age and B/S ratio. From the slope of contours on the graph, it can be seen that maximum value is more than 47 which indicates that the combined effect of age and B/S affects the outcome to a higher degree. At higher B/S ratios (more than 0.48), the combined effect of the two variables is more profound. This is because age and B/S are some of the most contributing factors to C-S of MK-based cement mortars. In the same way, the plot between age and aggregate diameter has been given in Fig. 17ii with age as the primary variable and diameter as a secondary variable. The effect of this interaction is also high on the output a depicted by the maximum contour slope of 62.47 for diameter values around 2 mm and age values around 30 to 40 days. This is also attributed to the fact that age was the most contributing factor to C-S and its interaction with other factors affect the output to a higher degree. Next plot (Fig. 17iii) shows interaction between diameter and MK/B ratio. Although these two variables were identified as least contributing by SHAP analysis, their combined effect has a considerable effect on the output given by the slope values reaching 55. This slope value is again achieved at diameter values around 2 mm which shows that the influence of diameter values around 2 mm has significant effect on the other variables. However, it is still less than the maximum slope value between age and diameter because age was the most important variable. This slope between diameter and MK/B proves the significance of 2D PDP plots by which the influence of combined effect of two variables can be found on the output which were otherwise labelled as least contributing by SHAP. Moving forward, Fig. 17iv shows MK/B as primary variable and age as secondary variable. Due to the age factor being involved, the maximum
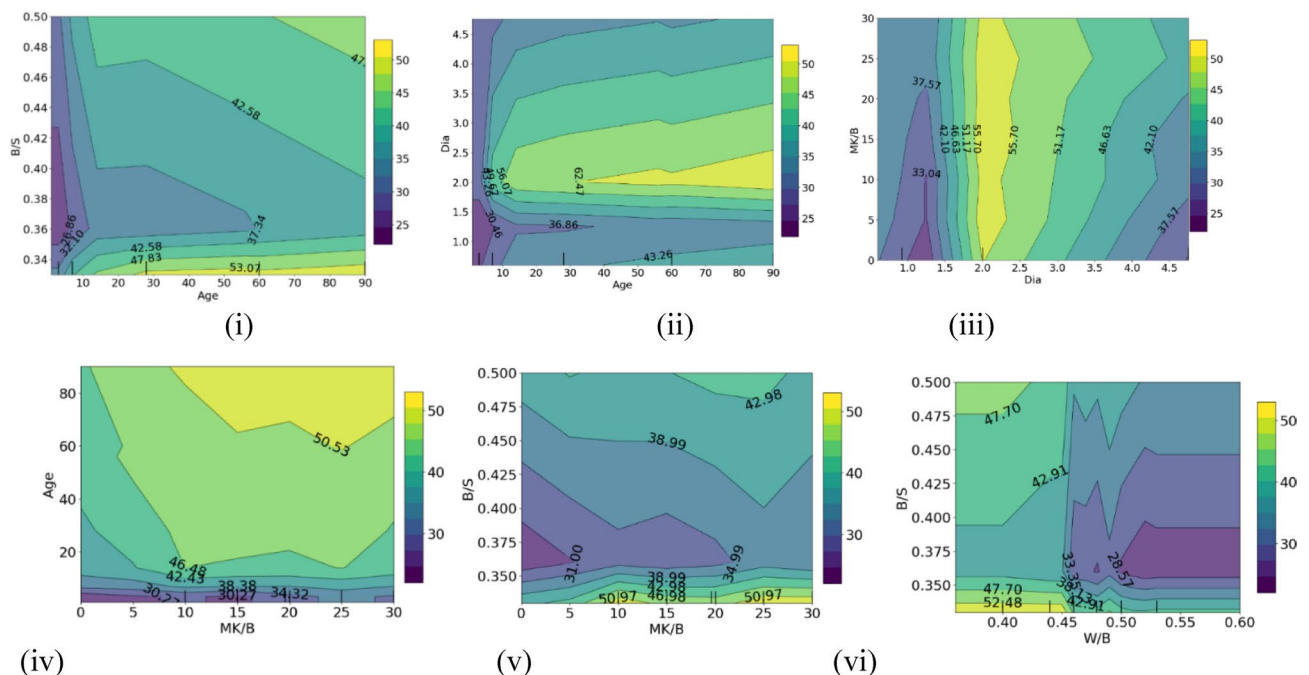

Fig. 17. 2-D PDP analysis to depict the combined effect of input features on output; (**i**) age and B/S ratio; (**ii**) age and diameter; (**iii**) diameter and MK/B; (**iv**) MK/B and age; (**v**) MK/B and B/S; (**vi**) W/B and B/S.

Fig. 18. User-friendly GUI developed for efficient prediction of compressive strength.

slope of contours is around 50 for age values more than 60 days and MK/B ratio around 25. Similarly, Fig. 17v depicts the interaction between MK/B and B/S with the former being primary variable and later being secondary variable. From Fig. 17v, it can be seen that at MK/B values 10 and 25 and at lower B/S ratios, the maximum slope is achieved hence the maximum effect on the outcome. Finally, the interaction between W/B ratio and B/S ratio is given in Fig. 17vi. Notice that at lower W/B values (close to 0.40) and B/S value less than 0.350, the influence of combined effect is more profound. The effect reduces as W/B ratio increases and reaches 0.60. This also aligns with experimental studies since the increase in W/B ratio causes a decrease in the C-S of mortars.

Practical implications and GUI

The accurate determination of C-S of cement mortars containing MK as a replacement of cement can help achieve a sustainable civil engineering industry by minimizing the utilization of emission intensive binder (cement) and fostering the use of secondary cementitious materials like MK etc. This will help conserve the natural environment by reducing the stress of depleting natural resources required for cement production i.e., limestone, gypsum etc. The incorporation of MK in the mortar mix imparts unexpected and non-linear behaviour to its C-S in relation to the mixture composition which is hard to determine by means of laboratory testing. Thus, this study developed empirical prediction models for C-S of MK-based cement mortar using MEP, GEP, XGB, AdaBoost, and BR algorithms. The predictive models presented in this study can be used practically by professionals in the industry to anticipate the C-S of cement mortar having MK without conducting extensive testing in the laboratory. Also, this study can serve as an important tool for mixture optimization of MK-based cement mortars based on the relationships given between different inputs and C-S based on ICE and PDP analyses. The explanatory analyses techniques conducted in this study including SHAP, ICE, and PDP provide information about which variables are most significant to predict C-S and which variables have a positive or negative correlation with the output. This information can be used to optimize mortar mixture to cater specific site requirements.

Although the empirical models given in the study can prove very effective for practical estimation of C-S of MK-based cement mortar, there are several limitations in their practical usage. It was identified in Sect. 7.9 that XGB not only surpassed its counterparts in this study in terms of accuracy but also proved to be more accurate than already existing studies. Thus, it can be used for output prediction with a greater accuracy. However, owing to its black-box nature, it is very challenging for the professionals working in the industry to effectively use the XGB model given in this study. Also, the equations yielded by MEP and GEP algorithms are complex and require significant computational resources to calculate the required C-S value. Thus, to overcome these challenges, and enable the professionals to benefit from the findings of current study, a GUI has been developed which can help predict the C-S based on mixture composition using XGB, MEP, and GEP algorithms and given in (Fig. 18). A significant advantage of this GUI is that it can be used by stakeholders and professionals having limited technical capabilities. The user can simply enter the input feature values in the designated boxes and upon pressing predict button, the user can get the desired output. Also, the python code to GUI has been deposited in a repository and the link to repository is given in the supplementary materials for the users to use the developed GUI for C-S prediction. To conclude, this GUI is a crucial tool which is an excellent example of practical implementation of the transformative power of ML to take data-driven informed decisions related to construction and building materials. This GUI will help to practically implement the findings of current study in practical settings.

Conclusions

This research aimed to develop reliable predictive models for compressive strength of MK-based cement mortars using multiple grey-box (MEP and GEP) and black-box (XGB, BR, and AdaBoost) ML algorithms to help effectively determine CS of MK-based mortar without extensive experimentation in order to reduce the

utilization of cement and consequently reduce the CO_2 emissions from construction industry. In addition, novel interpretable ML techniques were applied to foster data-driven decision making in the construction industry. The ML models were built using a dataset sourced from published literature, with the main findings of the study outlined below:

- GEP and MEP algorithms provided their outputs as empirical equations for the prediction of C-S owing to their grey-box nature while all other black-box algorithms failed to do so.
- The developed models were validated by error metrices, scatter plots, residual distribution, and external validation which proved that all the algorithms satisfied the recommended criteria to be deemed acceptable.
- XGB achieved the highest accuracy for determining C-S having the highest testing R^2 value equal to 0.998 followed by BR value of 0.946 while GEP has the lowest testing R^2 value of 0.846.
- SHAP analysis performed on the XGB model identified age and water-to-binder ratio as essential inputs for predicting the C-S of MK-based cement mortars.
- Further analysis including ICE, and PDP were used to investigate the effect of change in individual and combined input values on the output.
- An interactive GUI was developed to practically implement the findings of current study by efficiently predicting the C-S based on XGB, GEP, and MEP algorithms.

Future recommendations

While this study significantly advances the literature on predicting the compressive strength of metakaolin-based cement mortars using grey-box (GEP and MEP) and black-box (BR, AdaBoost, and XGB) algorithms, and includes innovative explanatory analyses, the limitations of the study must also be acknowledged. Future research should address these limitations to further refine the models and broaden their applicability.

- Future studies should prioritize the collection of more extensive datasets that incorporate a broader range of input parameters, enabling the models to accurately extrapolate beyond the limitations of this study.
- Future research should explore the use of additional machine learning algorithms, including LightGBM, CatBoost, GA, GBR, ANFIS, and ANNs, combined with various optimization techniques like Particle Swarm Optimization and Grey Wolf Optimization. These techniques can contribute to building more resilient models, which are less sensitive to variations in data and distributions, helping to reduce overfitting.
- This study focused only on the XGB model for explanatory purposes; however, future research should extend these analyses to a range of machine learning models and evaluate their results. This approach would provide a more thorough understanding of how each algorithm predicts values and help identify any discrepancies in feature importance or prediction mechanisms across different models.

Data availability

Data is available in the supplementary materials.

References

1. Rehan, R. & Nehdi, M. Carbon dioxide emissions and climate change: policy implications for the cement industry. *Environ. Sci. Policy*. **8** (2), 105–114. https://doi.org/10.1016/j.envsci.2004.12.006 (2005).
2. Benhelal, E., Zahedi, G., Shamsaei, E. & Bahadori, A. Global strategies and potentials to curb CO_2 emissions in cement industry. https://doi.org/10.1016/j.jclepro.2012.10.049 (Elsevier, 2013).
3. Shahmansouri, A. A., Akbarzadeh Bengar, H. & Ghanbari, S. Compressive strength prediction of eco-efficient GGBS-based geopolymer concrete using GEP method. *J. Building Eng.* https://doi.org/10.1016/j.jobe.2020.101326 (2020).
4. Sizirici, B., Fseha, Y., Cho, C. S., Yildiz, I. & Byon, Y. J. A review of carbon footprint reduction in construction industry, from design to operation. *MDPI* https://doi.org/10.3390/ma14206094 (2021).
5. Yu, J. et al. Stress relaxation behaviour of marble under Cyclic weak disturbance and confining pressures. *Meas. (Lond)*. https://doi.org/10.1016/J.MEASUREMENT.2021.109777 (2021).
6. Badogiannis, E., Kakali, G. & Tsivilis, S. Metakaolin as supplementary cementitious material: optimization of Kaolin to Metakaolin conversion. *J. Therm. Anal. Calorim.* **81** (2), 457–462. https://doi.org/10.1007/s10973-005-0806-3 (2005).
7. Kakali, G., Perraki, T., Tsivilis, S. & Badogiannis, E. Thermal treatment of Kaolin: the effect of mineralogy on the pozzolanic activity. *Appl. Clay Sci.* **20**, 1–2. https://doi.org/10.1016/S0169-1317(01)00040-0 (2001).
8. He, H. et al. A general and simple method to disperse 2D nanomaterials for promoting cement hydration. *Constr. Build. Mater.* https://doi.org/10.1016/J.CONBUILDMAT.2024.136217 (2024).
9. Li, Q. et al. Splitting tensile strength prediction of Metakaolin concrete using machine learning techniques. *Sci. Rep.* https://doi.org/10.1038/s41598-023-47196-4 (2023).
10. Chen, R. S. et al. Experimental study on ultimate bearing capacity of short thin-walled steel tubes reinforced with high-ductility concrete. *Structures* https://doi.org/10.1016/j.istruc.2024.107109 (2024).
11. Elfadaly, E., Othman, A. M., Aly, M. H., Elgarhy, W. A. & Abdellatief, M. Assessing performance and environmental benefits of high-performance geopolymer mortar incorporating pumice and rice straw Ash. *Sustain. Chem. Pharm.* **44**, 101918. https://doi.org/10.1016/J.SCP.2025.101918 (2025).
12. Liu, C. et al. The role of TBM asymmetric tail-grouting on surface settlement in coarse-grained soils of urban area: field tests and FEA modelling. *Tunn. Undergr. Space Technol.* https://doi.org/10.1016/J.TUST.2021.103857 (2021).
13. Saberi, F. & Hosseini-Barzi, M. Effect of thermal maturation and organic matter content on oil shale fracturing. *Int. J. Coal Sci. Technol.* **11** (1), 1–19. https://doi.org/10.1007/S40789-024-00666-0/FIGURES/16 (2024).
14. Wasim, M., Duc Ngo, T. & Law, D. Durability performance of reinforced waste-based geopolymer foam concrete under exposure to various corrosive environments. *Case Stud. Constr. Mater.* https://doi.org/10.1016/j.cscm.2021.e00703 (2021).
15. Diao, X. et al. Enhanced dynamic compressive performance of silicon carbide whiskers-modified mortars. *Constr. Build. Mater.* https://doi.org/10.1016/J.CONBUILDMAT.2024.136626 (2024).

16. Liu, X., Liu, X. & Zhang, Z. Application of red mud in carbon capture, utilization and storage (CCUS) technology. https://doi.org/10.1016/j.rser.2024.114683 (Elsevier, 2014).

17. Rahman, S. K. & Al-Ameri, R. Structural assessment of basalt FRP reinforced self-compacting geopolymer concrete using artificial neural network (ANN) modelling. *Constr. Build. Mater.* https://doi.org/10.1016/j.conbuildmat.2023.132464 (2023).

18. Wang, S., Guo, J., Yu, Y., Shi, P. & Zhang, H. Quality evaluation of land reclamation in mining area based on remote sensing. *Int. J. Coal Sci. Technol.* **10** (1), 1–10. https://doi.org/10.1007/S40789-023-00601-9/TABLES/6 (2023).

19. Cai, Y. et al. A review of monitoring, calculation, and simulation methods for ground subsidence induced by coal mining. *Int. J. Coal Sci. Technol.* **10** (1), 1–23. https://doi.org/10.1007/S40789-023-00595-4 (2023).

20. Soleimani, F., Si, G., Roshan, H. & Zhang, J. Numerical modelling of gas outburst from coal: a review from control parameters to the initiation process. *Int. J. Coal Sci. Technol.* https://doi.org/10.1007/S40789-023-00657-7 (2023).

21. Han, Z., Liu, K., Ma, J. & Li, D. Numerical simulation on the dynamic mechanical response and fracture mechanism of rocks containing a single hole. *Int. J. Coal Sci. Technol.* https://doi.org/10.1007/S40789-024-00718-5 (2024).

22. Qin, S., Lin, H., Yang, S. & Wei, Z. A mathematical model for parameter setting in discrete element numerical simulation. *Int. J. Coal Sci. Technol.* https://doi.org/10.1007/S40789-023-00644-Y (2023).

23. Sahu, A., Sinha, S. & Banka, H. Fuzzy inference system using genetic algorithm and pattern search for predicting roof fall rate in underground coal mines. *Int. J. Coal Sci. Technol.* https://doi.org/10.1007/S40789-023-00630-4 (2024).

24. Tran, V. L. & Kim, S. E. Efficiency of three advanced data-driven models for predicting axial compression capacity of CFDST columns. *Thin-Walled Struct.* https://doi.org/10.1016/j.tws.2020.106744 (2020).

25. Sarir, P., Chen, J., Asteris, P. G., Armaghani, D. J. & Tahir, M. M. Developing GEP tree-based, neuro-swarm, and Whale optimization models for evaluation of bearing capacity of concrete-filled steel tube columns. *Eng. Comput.* **37** (1), 1–19. https://doi.org/10.1007/s00366-019-00808-y (2021).

26. Memarzadeh, A., Shahmansouri, A. A. & Poologanathan, K. A novel prediction model for post-fire elastic modulus of circular recycled aggregate concrete-filled steel tubular stub columns. *Steel Compos. Struct.* **44** (3), 309–324. https://doi.org/10.12989/SCS.2022.44.3.309 (2022).

27. Waqas, H. A., Bahrami, A., Amin, F., Sahil, M. & Saud Khan, M. Numerical modeling and performance evaluation of carbon fiber-reinforced polymer-strengthened concrete culverts against water-induced corrosion. *Infrastructures (Basel)* **9** (5). https://doi.org/10.3390/infrastructures9050082 (2024).

28. Wang, M. et al. Sulfate diffusion in coal pillar: experimental data and prediction model. *Int. J. Coal Sci. Technol.* **10** (1), 1–12. https://doi.org/10.1007/S40789-023-00575-8/FIGURES/12 (2023).

29. Ismael Jaf, D. K. et al. Hybrid nonlinear regression model versus MARS, MEP, and ANN to evaluate the effect of the size and content of waste tire rubber on the compressive strength of concrete. *Heliyon* **10** (4). https://doi.org/10.1016/j.heliyon.2024.e25997 (2024).

30. Qi, C. & Tang, X. Slope stability prediction using integrated metaheuristic and machine learning approaches: A comparative study. *Comput. Ind. Eng.* **118**, 112–122. https://doi.org/10.1016/j.cie.2018.02.028 (2018).

31. Huang, F. et al. Slope stability prediction based on a long short-term memory neural network: comparisons with convolutional neural networks, support vector machines and random forest models. *Int. J. Coal Sci. Technol.* **10** (1), 1–14. https://doi.org/10.1007/S40789-023-00579-4/FIGURES/5 (2023).

32. Zheng, H., Jiang, B., Wang, H. & Zheng, Y. Experimental and numerical simulation study on forced ventilation and dust removal of coal mine heading surface. *Int. J. Coal Sci. Technol.* **11** (1), 1–17. https://doi.org/10.1007/S40789-024-00667-Z/FIGURES/20 (2024).

33. Qi, Q., Yue, X., Duo, X., Xu, Z. & Li, Z. Spatial prediction of soil organic carbon in coal mining subsidence areas based on RBF neural network. *Int. J. Coal Sci. Technol.* **10** (1), 1–13. https://doi.org/10.1007/S40789-023-00588-3/TABLES/4 (2023).

34. Lu, D., Wang, G., Du, X. & Wang, Y. A nonlinear dynamic uniaxial strength criterion that considers the ultimate dynamic strength of concrete. *Int. J. Impact Eng.* **103**, 124–137. https://doi.org/10.1016/J.IJIMPENG.2017.01.011 (2017).

35. Handousa, A. M., Abdellatief, M., Salem, F. A., Mahmoud, N. & Ghannam, M. Experimental and analytical study on axial behaviour of square corrugated concrete filled single and double skin tube stub columns. *Sci. Rep. Vol.* **15** (1), 1–26. https://doi.org/10.1038/s41598-025-85585-z (2025).

36. Abdellatief, M. et al. Sustainable foam glass property prediction using machine learning: A comprehensive comparison of predictive methods and techniques. *Results Eng.* **25**, 104089. https://doi.org/10.1016/J.RINENG.2025.104089 (2025).

37. Faraj, R. H., Ahmed, H. U., Fathullah, H. S., Abdulrahman, A. S. & Abed, F. Tensile strain capacity prediction of engineered cementitious composites (ECC) using soft computing techniques. *Comput. Model. Eng. Sci.* **138** (3), 2925–2954. https://doi.org/10.32604/cmes.2023.029392 (2024).

38. J. M.-J. of S. C. M. and and undefined Prediction of compressive strength of calcined clay based cement mortars using support vector machine and artificial neural network techniques. *dergipark.org.trJ MARANGUJournal Sustain. Construct. Mater. Technol.* **5** (1), 392–398. https://doi.org/10.29187/jscmt.2020.43 (2020).

39. Çalışkan, A., Demirhan, S. & Tekin, R. Comparison of different machine learning methods for estimating compressive strength of mortars. *Constr. Build. Mater.* https://doi.org/10.1016/j.conbuildmat.2022.127490 (2022).

40. Lu, D. et al. A dynamic elastoplastic model of concrete based on a modeling method with environmental factors as constitutive variables. *J. Eng. Mech.* https://doi.org/10.1061/JENMDT.EMENG-7206 (2023).

41. Chen, Z. Application of machine learning boosting and bagging methods to predict compressive and flexural strength of marble cement mortar. *Mater. Today Commun.* https://doi.org/10.1016/j.mtcomm.2024.108600 (2024).

42. Zou, B., Yin, J., Liu, Z. & Long, X. Transient rock breaking characteristics by successive impact of shield disc cutters under confining pressure conditions. *Tunn. Undergr. Space Technol.* https://doi.org/10.1016/J.TUST.2024.105861 (2024).

43. Shah, S. F. A., Chen, B., Zahid, M. & Ahmad, M. R. Compressive strength prediction of one-part alkali activated material enabled by interpretable machine learning. *Constr. Build. Mater.* https://doi.org/10.1016/j.conbuildmat.2022.129534 (2022).

44. Dehestani, A., Kazemi, F., Abdi, R. & Nitka, M. Prediction of fracture toughness in fibre-reinforced concrete, mortar, and rocks using various machine learning techniques. *Eng. Fract. Mech.* https://doi.org/10.1016/j.engfracmech.2022.108914 (2022).

45. Ali, Y., Hussain, F., Irfan, M. & Buller, A. S. An eXtreme gradient boosting model for predicting dynamic modulus of asphalt concrete mixtures. *Constr. Build. Mater.* https://doi.org/10.1016/j.conbuildmat.2021.123642 (2021).

46. Lu, H. et al. Study of synergistic behavior during bituminous coal-cow manure co-gasification: the role of intrinsic AAEM and organic matter. *Int. J. Coal Sci. Technol.* https://doi.org/10.1007/S40789-024-00694-W (2024).

47. Asteris, P. G., Koopialipoor, M., Armaghani, D. J., Kotsonis, E. A. & Lourenço, P. B. Prediction of cement-based mortars compressive strength using machine learning techniques. *Neural Comput. Appl.* **33**, 13089–13121. https://doi.org/10.1007/S00521-021-06004-8/TABLES/6 (2021).

48. Gayathri, R., Rani, S. U., Čepová, L., Rajesh, M. & Kalita, K. A comparative analysis of machine learning models in prediction of mortar compressive strength. *Processes* **10** (7), 1387. https://doi.org/10.3390/PR10071387 (2022). Page 1387.

49. Izadgoshasb, H., Kandiri, A., Shakor, P., Laghi, V. & Gasparini, G. Predicting compressive strength of 3d printed mortar in structural members using machine learning. *Appl. Sci. Vol.* **11** (22), 10826. https://doi.org/10.3390/APP112210826 (2021).

50. Fei, Z., Liang, S., Cai, Y. & Shen, Y. Ensemble machine-learning-based prediction models for the compressive strength of recycled powder mortar. *Materials* **16** (2), 583. https://doi.org/10.3390/MA16020583 (2023).

51. Nakkeeran, G. et al. Machine learning application to predict the mechanical properties of glass fiber mortar. *Adv. Eng. Softw.* **180**, 103454. https://doi.org/10.1016/J.ADVENGSOFT.2023.103454 (2023).

52. Kurt, Z., Yilmaz, Y., Cakmak, T. & Ustabaş, I. A novel framework for strength prediction of geopolymer mortar: renovative precursor effect. *J. Building Eng.* **76**, 107041. https://doi.org/10.1016/J.JOBE.2023.107041 (2023).
53. Shahin, M., Jaksa, M. B. & Maier, H. R. Physical modeling of rolling dynamic compaction view project artificial neural networks-pile capacity prediction view project. https://www.researchgate.net/publication/228364758 (2008).
54. He, L. et al. A quasi-exponential distribution of interfacial voids and its effect on the interlayer strength of 3D printed concrete. *Addit. Manuf.* https://doi.org/10.1016/J.ADDMA.2024.104296 (2024).
55. Ferreira, C. Gene expression programming: a new adaptive algorithm for solving problems. *Arxiv Org.* (2001).
56. C Ferreira, Mutation, transposition, and recombination: An analysis of the evolutionary dynamics., In H. J. Caulfield, S.-H. Chen, H.-D. Cheng, R. Duro, V. Honavar, E. E. Kerre, M. Lu,M. G. Romay, T. K. Shih, D. Ventura, P. P. Wang, Y. Yang, eds., *Proceedings of the 6th Joint Conference on Information Sciences, 4th International Workshop on Frontiers* 1–4. (2002).
57. Zhang, R. & Xue, X. A new model for prediction of soil thermal conductivity. *Int. Commun. Heat Mass Transfer* **129** 105661. https://doi.org/10.1016/j.icheatmasstransfer.2021.105661 (2021).
58. Oltean, M. & Dumitrescu, D. Multi expression programming. *J. Gen. Program. Evolv. Mach.* (2002).
59. Gandomi Amir, A. H. A., Yun, H. & Jin, G. Formulation of uplift capacity of Suction caissons using multi expression programming. *KSCE J. Civ. Eng.* **15** (2), 363–373. https://doi.org/10.1007/s12205-011-1117-9 (2011).
60. Chen Carlos, T. G. XGBoost: A scalable tree boosting system. https://doi.org/10.1145/2939672.2939785 (2016).
61. Inqiad, W. B. et al. Comparison of boosting and genetic programming techniques for prediction of tensile strain capacity of engineered cementitious composites (ECC). *Mater. Today Commun.* **39**, 109222. https://doi.org/10.1016/j.mtcomm.2024.109222 (2024).
62. Bushenkova, A., Soares, P. M. M., Johannsen, F. & Lima, D. C. A. Towards an improved representation of the urban heat Island effect: A multi-scale application of XGBoost for Madrid. *Urban Clim.* **55**, 101982. https://doi.org/10.1016/j.uclim.2024.101982 (2024).
63. Feng Liu, D. C. Z. et al. Machine learning-based compressive strength prediction for concrete: an adaptive boosting approach. *Constr. Build. Mater.* **230** 117000. https://doi.org/10.1016/j.conbuildmat.2019.117000 (2020).
64. Bakır, R., Orak, C. & Yüksel, A. Optimizing hydrogen evolution prediction: A unified approach using random forests, LightGBM, and bagging regressor ensemble model. *Int. J. Hydrogen Energy.* **67**, 101–110. https://doi.org/10.1016/j.ijhydene.2024.04.173 (2024).
65. Aslam, F., Alyousef, R., Hassan Awan, H. & Javed, M. F. Forecasting the self-healing capacity of engineered cementitious composites using bagging regressor and stacking regressor. *Structures* **54**, 1717–1728. https://doi.org/10.1016/j.istruc.2023.05.140 (2023).
66. Wang, N., Samavatian, M., Samavatian, V. & Sun, H. Bayesian machine learning-aided approach bridges between dynamic elasticity and compressive strength in the cement-based mortars. *Mater. Today Commun.* https://doi.org/10.1016/j.mtcomm.2023.106283 (2023).
67. Gayathri, R., Rani, S. U., Čepová, L., Rajesh, M. & Kalita, K. A comparative analysis of machine learning models in prediction of mortar compressive strength. *Processes* https://doi.org/10.3390/pr10071387 (2022).
68. Ozcan, G., Kocak, Y. & Gulbandilar, E. Estimation of compressive strength of BFS and WTRP blended cement mortars with machine learning models. *Computers Concrete.* **19** (3), 275–282. https://doi.org/10.12989/cac.2017.19.3.275 (2017).
69. D'Ayala, D., Significance, T. V. & Set E. F.-S. and and undefined Blended lime-cement mortars for conservation purposes: Microstructure and strength development, taylorfrancis.comD D'Ayala, E FoddeStructural analysis of historic construction: preserving safety and, 2008 taylorfrancis.com. https://www.taylorfrancis.com/chapters/edit/10.1201/9781439828229-126/blended-lime-cement-mortars-conservation-purposes-microstructure-strength-development (Accessed 19 May 2024). (2008).
70. Sarkar, A. et al. Spectral signatures of solvent-extracted macromolecules in Indian coals of different rank: insights from fluorescence excitation-emission matrix. *Int. J. Coal Sci. Technol.* https://doi.org/10.1007/S40789-024-00721-W (2024).
71. Courard, L. et al. Durability of mortars modified with Metakaolin. *Cem. Concr Res.* **33** (9), 1473–1479. https://doi.org/10.1016/S0008-8846(03)00090-5 (2003).
72. Potgieter-Vermaak, S. S. & Potgieter, J. H. Metakaolin as an extender in South African cement. *J. Mater. Civil Eng.* **18** (4), 619–623. (2006).
73. Abdalla, A. A. & Mohammed, A. S. The efficiency of hybrid intelligent models to evaluate the effect of the size of sand and clay Metakaolin content on various compressive strength ranges of cement mortar. *Neural Comput. Appl.* **36** (11), 6209–6229. https://doi.org/10.1007/S00521-023-09384-1 (2024).
74. Kadri, E. H., Kenai, S., Ezziane, K., Siddique, R. & De Schutter, G. Influence of Metakaolin and silica fume on the heat of hydration and compressive strength development of mortar. *Appl. Clay Sci.* **53** (4), 704–708. https://doi.org/10.1016/j.clay.2011.06.008 (2011).
75. Al-Chaar, G. K., Alkadi, M. & Asteris, P. G. Send orders for reprints to reprints@benthamscience.net natural Pozzolan as a partial substitute for cement in concrete. (2013).
76. Dhawan, H. & Sharma, D. K. Process engineering of demineralisation of moderate to high Ash Indian coals through NaOH-HCl leaching and HF leaching. *Int. J. Coal Sci. Technol.* https://doi.org/10.1007/S40789-024-00679-9 (2024).
77. Mahanta, A., Attry, B. & Saikia, B. K. Aqueous oxidation of coal-associated pyrite and standard pyrite mineral towards Understanding the depyritization kinetics and acid formations. *Int. J. Coal Sci. Technol.* https://doi.org/10.1007/S40789-024-00735-4 (2024).
78. Zimek, A., Schubert, E. & Kriegel, H. P. A survey on unsupervised outlier detection in high-dimensional numerical data. *Stat. Anal. Data Min. ASA Data Sci. J.* **5** (5), 363–387. https://doi.org/10.1002/SAM.11161 (2012).
79. Khan, M. I. & Abbas, Y. M. Robust extreme gradient boosting regression model for compressive strength prediction of blast furnace slag and fly Ash concrete. *Mater. Today Commun.* **35**, 105793. https://doi.org/10.1016/J.MTCOMM.2023.105793 (2023).
80. Chu, H. H. 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.* **12** (4), 3603–3617. https://doi.org/10.1016/j.asej.2021.03.018 (2021).
81. Iftikhar, B. et al. Predictive modeling of compressive strength of sustainable rice husk Ash concrete: ensemble learner optimization and comparison. *J. Clean. Prod.* https://doi.org/10.1016/j.jclepro.2022.131285 (2022).
82. Sarveghadi, M., Gandomi, A. H., Bolandi, H. & Alavi, A. H. *Development of prediction models for shear strength of SFRCB using a machine learning approach.* https://doi.org/10.1007/s00521-015-1997-6 (Springer, 2019).
83. Modiga, A., Eterigho-Ikelegbe, O. & Bada, S. Extractability and mineralogical evaluation of rare Earth elements from Waterberg coalfield run-of-mine and discard coal. *Int. J. Coal Sci. Technol.* **11** (1), 65. https://doi.org/10.1007/S40789-024-00702-Z (2024).
84. Inqiad, W. B. et al. Soft computing models for prediction of bentonite plastic concrete strength. *Nat. Com.* **14** 18145, 123AD. https://doi.org/10.1038/s41598-024-69271-0
85. Inqiad, W. & Javed, M. M. S.-J. of B., and undefined Predicting natural vibration period of concrete frame structures having masonry infill using machine learning techniques. https://www.sciencedirect.com/science/article/pii/S2352710224019855 (Accessed 07 September 2024) (Elsevier, 2024).
86. Khan, M. S. et al. Predicting residual strength of hybrid fibre-reinforced Self-compacting concrete (HFR-SCC) exposed to elevated temperatures using machine learning. *Case Stud. Constr. Mater.* https://doi.org/10.1016/j.cscm.2024.e04112 (2025).
87. Inqiad, W. B. et al. Utilizing contemporary machine learning techniques for determining soilcrete properties. *Earth Sci. Inf.* **18** (1), 1–30. https://doi.org/10.1007/S12145-024-01520-2/FIGURES/2 (2025).

88. Alhakeem, Z. M. et al. Prediction of ecofriendly concrete compressive strength using gradient boosting regression tree combined with GridSearchCV hyperparameter-optimization techniques. *Materials* https://doi.org/10.3390/ma15217432 (2022).
89. Zöller, M. A. & Huber, M. F. Benchmark and survey of automated machine learning frameworks. (2021).
90. Hoang, N. D. A novel ant colony-optimized extreme gradient boosting machine for estimating compressive strength of recycled aggregate concrete. *Multiscale Multidisciplinary Model. Experiments Des.* https://doi.org/10.1007/s41939-023-00220-6 (2023).
91. Al-Taai, S. R. et al. XGBoost prediction model optimized with bayesian for the compressive strength of eco-friendly concrete containing ground granulated blast furnace slag and recycled coarse aggregate. *Appl. Sci. (Switzerland).* https://doi.org/10.3390/app13158889 (2023).
92. Lu, S. et al. A review of coal permeability models including the internal swelling coefficient of matrix. *Int. J. Coal Sci. Technol.* https://doi.org/10.1007/S40789-024-00701-0 (2024).
93. Rostami, A. et al. Rigorous framework determining residual gas saturations during spontaneous and forced imbibition using gene expression programming. *J. Nat. Gas Sci. Eng.* https://doi.org/10.1016/J.JNGSE.2020.103644 (2020).
94. Despotovic, M., Nedic, V., Despotovic, D. & Cvetanovic, S. Evaluation of empirical models for predicting monthly mean horizontal diffuse solar radiation. *Renew. Sustain. Energy Rev.* **56**, 246–260. https://doi.org/10.1016/J.RSER.2015.11.058 (2016).
95. Li, H., Yang, Y., Wang, X. & Tang, H. Effects of the position and chloride-induced corrosion of strand on bonding behavior between the steel strand and concrete. *Structures* **58** https://doi.org/10.1016/J.ISTRUC.2023.105500 (2023).
96. Asteris, P. G., Koopialipoor, M., Armaghani, D. J., Kotsonis, E. A. & Lourenço, P. B. Prediction of cement-based mortars compressive strength using machine learning techniques. *Neural Comput. Appl.* **33**, 13089–13121. https://doi.org/10.1007/s00521-021-06004-8 (2021).
97. Asiwaju, L., Mustapha, K. A., Abdullah, W. H., Sia, S. G. & Hakimi, M. H. Geochemistry of cenozoic coals from Sarawak basin, Malaysia: implications for paleoclimate, depositional conditions, and controls on petroleum potential. *Int. J. Coal Sci. Technol.* **11** (1). https://doi.org/10.1007/S40789-024-00690-0 (2024).
98. Hower, J. C., Eble, C. F., Hopps, S. D., Morgan, T. D. & Edwards, V. A. Petrology and geochemistry of the middle Pennsylvanian (Langsettian) Clintwood coalbed, Pike County, Kentucky. *Int. J. Coal Sci. Technol.* **12** (1), 22. https://doi.org/10.1007/S40789-025-00764-7 (2025).
99. Zhang, Y. et al. Research on coal-rock identification method and data augmentation algorithm of comprehensive working face based on FL-Segformer. *Int. J. Coal Sci. Technol.* https://doi.org/10.1007/S40789-024-00704-X (2024).
100. Khan, M. A., Zafar, A., Akbar, A., Javed, M. F. & Mosavi, A. Application of gene expression programming (GEP) for the prediction of compressive strength of geopolymer concrete. *Materials* **14** (5), 1–23. https://doi.org/10.3390/ma14051106 (2021).
101. Sun, M. et al. The release and migration mechanism of arsenic during pyrolysis process of Chinese coals. *Int. J. Coal Sci. Technol.* https://doi.org/10.1007/S40789-024-00715-8 (2024).
102. Amoih, F. et al. Comparative study on different coals from the Lorraine basin (France) by sorption isotherms, thermogravimetric analysis and breakthrough curves for CO2-ECBM recovery. *Int. J. Coal Sci. Technol.* https://doi.org/10.1007/S40789-024-00696-8 (2024).
103. Asteris, P. et al. Predicting concrete compressive strength using hybrid ensembling of surrogate machine learning models. *ElsevierPG Asteris, P. S.-C. and C., and undefined, K PilakoutasCement and Concrete Research, 2021* https://www.sciencedirect.com/science/article/pii/S0008884621000983 (Accessed 13 September 2024) (Elsevier, 2021).
104. He, L. et al. Development of novel concave and convex trowels for higher interlayer strength of 3D printed cement paste. *Case Stud. Constr. Mater.* **21**, e03745. https://doi.org/10.1016/J.CSCM.2024.E03745 (2024).
105. Quan Tran, V., Quoc Dang, V. & Si Ho, L. Evaluating compressive strength of concrete made with recycled concrete aggregates using machine learning approach. *Constr. Build. Mater.* https://doi.org/10.1016/j.conbuildmat.2022.126578 (2022).

Acknowledgements

Author contributions

Naseer Muhammad Khan: Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation. Saad S. Alarifi: Funding acquisition, Formal analysis, Data curation, Conceptualization. Muhammad Saud Khan: Writing – original draft, Visualization, Validation, Supervision, Software. Liqiang Ma: Writing – review & editing, Visualization, Project administration, Investigation, Funding acquisition, Formal analysis, Conceptualization. Waleed Bin Inqiad: Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Formal analysis, Data curation, Conceptualization. Imtiaz Iqbal: Project administration, Methodology, Data curation, Conceptualization. Muhammad Saud Khan: Writing – original draft, Visualization, Validation, Supervision, Software. All authors reviewed the manuscript.

Declarations

Competing interests
The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-01327-1.

Correspondence and requests for materials should be addressed to N.M.K. or L.M.

Reprints and permissions information is available at www.nature.com/reprints.

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