Journal of Building Engineering 95 (2024) 110330

Contents lists available at ScienceDirect

Journal of Building Engineering

journal homepage: www.elsevier.com/locate/jobe

Development of a learner model tool for predicting strength and embodied carbon for lightweight concrete production

Promise D. Nukah [a,*], Samuel J. Abbey [a,b], Colin A. Booth [a,b]

[a] School of Engineering, College of Arts, Technology and Environment, University of the West of England, Bristol, BS16 1QY, UK
[b] Centre for Architecture and Built Environment Research (CABER), College of Arts, Technology and Environment, University of the West of England, Bristol, BS16 1QY, UK

ARTICLE INFO

Keywords:
Machine learning
Embodied carbon
Foundation structure
Artificial intelligence

ABSTRACT

The demand for sustainable concrete in meeting the net zero carbon target places a burden in optimizing concrete response to structural strength that satisfy acceptable embodied carbon. In most cases, a low carbon concrete is deficient in structural requirement and vice versa. This dilemma informs the need for a tool that can predict compressive strength as well as embodied carbon using the same input data. Since the use of alternative materials as cement replacement to enhance sustainability is emerging in the quest for a sustainable concrete, an optimal material that satisfy both conditions of structural integrity and sustainability is still lacking. Paucity of data in the emerging lightweight low carbon concrete using alternative materials, portends an upheave in the bias for prediction of the behaviour of lightweight low carbon concrete. This study therefore uses concrete data of lightweight low carbon concrete from laboratory experiment for the prediction of compressive strength and embodied carbon with their performance evaluated using eight machine leaning regression models. The results obtained indicates that the XG boost regression model exhibited excellent performance with a low Mean Squared Error (MSE) of 50.15, Mean absolute error(MAE) = 5.26, Mean absolute percentage error(MAPE) = 11.76 %, Explained variance score = 0.97, Root mean square error(RMSE) = 7.08 and a high R squared value of 0.96. The tool predicted compressive strength and embodied carbon for lightweight carbon concrete using multiple output regression such that the output can be limited to the yearly structural embodied carbon threshold to achieving 2050 net zero target. The developed tool when compared with concrete of similar mix ingredients performed more than 95 % in predicting concrete compressive strength and associated embodied carbon. In line with the inclusion of embodied carbon in carbon regulations of buildings in the UK as suggested by the professionals in the construction industry, the developed learner model and prediction tool has integrated concrete strength and embodied carbon to initiate a holistic approach to design and construction, balancing performance, cost, and environmental impact.

1. Introduction

Machine learning techniques can optimize the prediction process by continuously refining models based on new data inputs. Through iterative learning and feedback loops, machine learning algorithms can improve the accuracy of embodied carbon predictions

* Corresponding author.
E-mail addresses: promise.nukah@uwe.ac.uk (P.D. Nukah), samuel.abbey@uwe.ac.uk (S.J. Abbey), colin.booth@uwe.ac.uk (C.A. Booth).

https://doi.org/10.1016/j.jobe.2024.110330
Received 28 May 2024; Received in revised form 10 July 2024; Accepted 28 July 2024
Available online 2 August 2024

over time. This iterative approach allows for the adaptation of models to changing construction practices and environmental conditions. Machine learning enables the automation of carbon footprint estimation processes, reducing the need for manual calculations and streamlining the prediction workflow. In the review of carbon emission in building [1], It was reported that Global carbon emissions increased from 24.69 billion tons in 2000 to 36.14 billion tons in 2014 with the building sector consuming 40 % of the total energy use and accounts for one-third of energy-related carbon emissions globally. On the regional level, it was noted that buildings account for 39 % of carbon emissions in the United States and In the European Union (EU), buildings contribute to about 59 % of the total electricity consumption. In China, commercial buildings account for about 20%–30 % of national energy consumption while in Hong Kong, buildings contribute to about 90 % of the total electricity consumption and 60 % of carbon emissions.

As the Production of raw material in the construction industry contributes to 93 % of the total emissions in the industry, 78 % of which are from concrete [2]. The demand for concrete is expected to rise with the world's global built-up area projected to double by 2060 and amount to an extra 230 billion square area of building needed, which is equivalent to adding the size of New York City every month [3]. As noted that an average of 900 kg of Carbon is emitted for every 1000 kg of Portland cement produced, emphasizes the urgency for alternative cement replacement using supplementary cementitious materials (SCM) in concrete production [4]. Several materials have been used as SCM in concrete with source from industrial and agricultural waste, however it has been suggested that the viability of SCM from industrial waste present effective potency with the possibility of greater than 40 % cement replacement at a low embodied carbon [5]. Notwithstanding the low embodied carbon with these materials when used as SCM, there are also issues related to workability, durability and poor strength development which can impact project delivery schedule and performance. Prominent among the SCMs from industrial waste is ground granular blast furnace slag (GGBS) owing to its cementitious and pozzolanic properties. The reduction of embodied carbon in concrete structures has also attracted several other approaches from researchers. These approaches as presented in Fig. 1 include structural section optimization [6] and self-weight reduction. Resizing of a member section alters the mechanical properties of the structures especially the moment of inertia as well as the stiffness. This alteration culminated in the reduction of the volume of concrete and embodied carbon. With the replacement of normal weight aggregate with a lightweight aggregate, the reinforcement demand in the structures reduces as well as the embodied carbon. While these approaches are driven mainly by human intuition, the margin of embodied carbon reduction is only affected by concrete volume reduction while the material composition remains the same. Considering that embodied carbon is determined by the carbon factor of the concrete ingredient, optimizing the mix composition to reduce embodied carbon that is devoid of substantial errors and bias will enhance significant reduction in embodied carbon with the use machine learning and artificial intelligence [7]. Since the effectiveness of prediction using machine learning is predicated on the viability of dataset [8–10], the use of experimental dataset as against empirical portends better reliability [11]. Findings from literatures as shown in Fig. 2 indicate that there are inconsistencies in dataset for prediction of the performance of low carbon concrete [12–15]. It was also shown that the use of alkaline solution improves the mechanical properties of GGBS based concrete when used in combination with silica fume [15].

By analysing historical data on construction projects, machine learning models can learn from past patterns and make more accurate predictions about the carbon footprint of ground beams. Studies have shown that foundation construction materials account for 67.9 % of the greenhouse gas (GHG) emissions, equipment usage contributed to 14.75 %, and transportation added 17.35 % for cast-in-situ piles construction. For precast piles construction, materials contributed to 74.62 % of the emissions, equipment usage was 6.5 %,

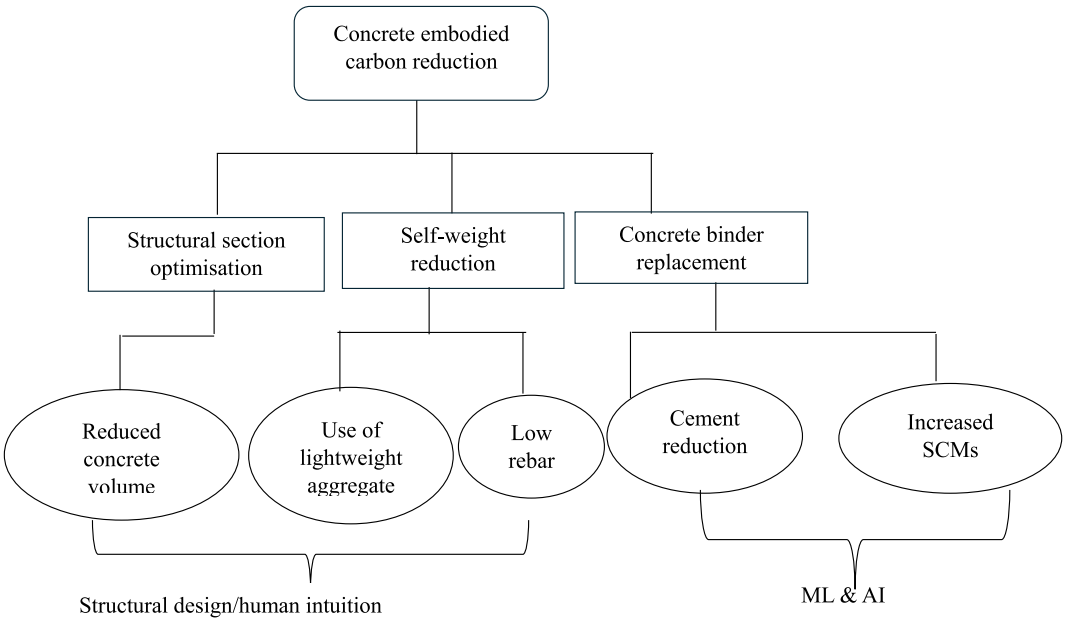


Fig. 1. Techniques used in embodied carbon reduction in concrete structures.

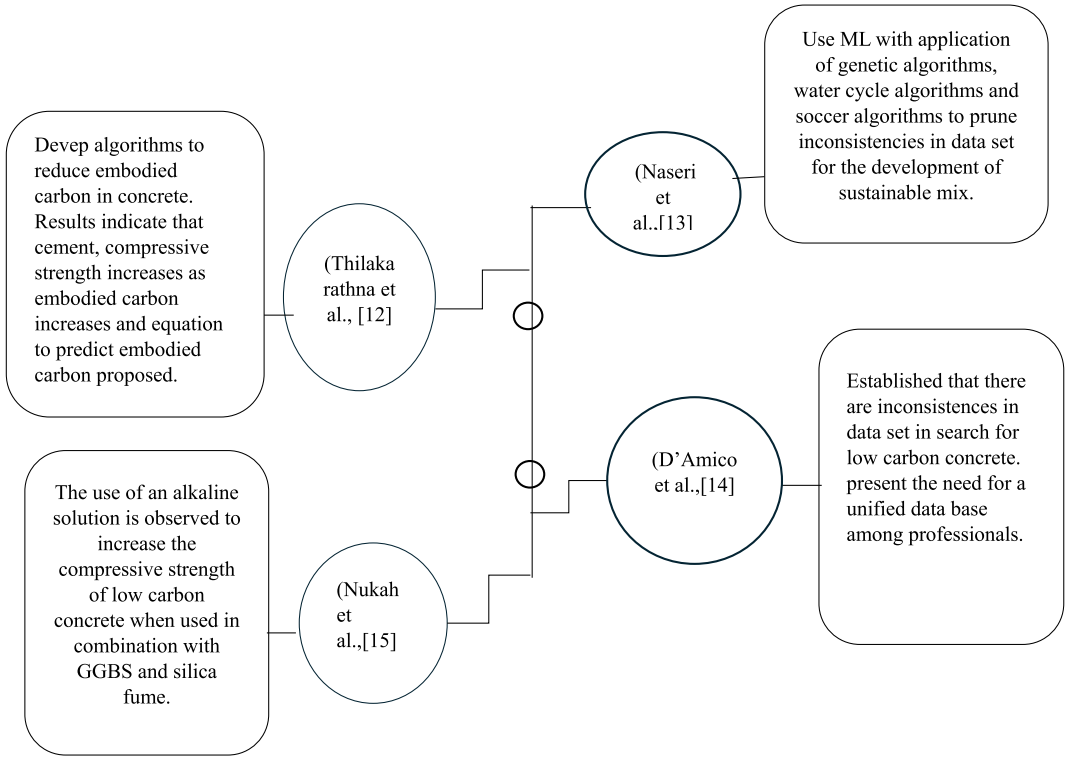


Fig. 2. Evidence of dataset inconsistence for low carbon concrete performance prediction and viability of GGBS as cement replacement.

and transportation accounted for 18.87 % of the emissions [16]. Machine learning contributes to predicting embodied carbon in ground beams by leveraging algorithms and statistical models to analyse complex data sets and identify patterns that traditional methods may overlook [17,18]. The results of machine learning techniques, specifically Artificial Neural Network (ANN), Support Vector Regression (SVR), and Response Surface Methodology (RSM), on the prediction of embodied carbon in ground beams indicates that ANN model successfully captures the variability in the data related to the prediction of embodied carbon in ground beams [19]. From the study, the SVR model demonstrates consistent superiority over both the ANN and RSM models in terms of diverse performance metrics such as residuals and correlation coefficients while the RSM regression model present a robust correlation between variables and observed outcomes, with optimal p-values, low R-squared values, and f-values. It was concluded that SVR, can be effectively utilized to predict the embodied carbon in ground beams and other construction elements.

In predicting construction-stage carbon emissions, the use of random forest (RF) method has been observed to play a significant role. On comparing traditional multilinear regression method, RF model demonstrated a higher coefficient of determination (R^2 = 0.64) and lower mean square error (MSE = 0.76) [20]. These algorithms can process large volumes of data related to construction materials, processes, and environmental factors to identify correlations and trends that influence embodied carbon. Machine learning algorithms excel at recognizing patterns in data that may not be apparent to human analysts [21]. By training models on diverse data sets, machine learning can identify subtle relationships between different variables that impact embodied carbon in ground beams. This ability to detect complex patterns enables more precise predictions of carbon emissions It was reported that in 2015, the total embodied carbon in the building sector represented approximately 17.62 % of the total reported carbon emmision (CO_2) emissions in China [22].

By automating repetitive tasks such as data preprocessing, feature selection, and model training, machine learning accelerates the prediction of embodied carbon in ground beams and enhances efficiency in sustainability assessments. Machine learning models can scale to handle large and diverse data sets, making them suitable for analysing complex factors that influence embodied carbon in construction projects. Whether dealing with multiple variables, varying project scales, or diverse material types, machine learning algorithms can adapt to different scenarios and provide scalable solutions for predicting carbon emissions in ground beams [23]. Artificial intelligence (AI) also offers several benefits in carbon footprint estimation for construction projects. AI algorithms is able to analyse vast amounts of data and complex relationships to provide more accurate estimates of carbon footprints in construction. By considering multiple variables and factors simultaneously, AI models can offer precise calculations that account for various parameters influencing carbon emissions. AI streamlines the carbon footprint estimation process by automating data analysis, modelling, and prediction tasks [24]. This automation reduces the time and effort required for manual calculations, allowing for faster and more efficient sustainability assessments in construction projects. AI can forecast future carbon emissions based on historical data and trends, enabling proactive decision-making to minimize environmental impact. By predicting potential carbon footprints at different stages of a project, AI helps stakeholders implement mitigation strategies and optimize resource usage for sustainability. AI algorithms

can optimize construction practices and material selection to reduce embodied carbon and overall environmental impact. By analysing alternative scenarios and recommending eco-friendly solutions, AI contributes to more sustainable building designs and construction processes [25]. AI generates data-driven insights into carbon footprint sources, trends, and mitigation opportunities in construction. By analysing patterns and correlations in data, AI helps identify areas for improvement and guides decision-making towards greener and more sustainable practices. AI technologies can also scale to handle large and diverse data sets, making them suitable for analysing carbon footprints across various construction projects and portfolios. Whether assessing individual buildings or entire developments, AI offers scalable solutions for estimating and managing carbon emissions in construction. AI systems can learn from new data inputs and feedback, continuously improving their accuracy and predictive capabilities over time. By adapting to changing environmental regulations, technological advancements, and industry trends, AI ensures that carbon footprint estimations remain up-to-date and relevant in the evolving construction landscape [26]. With the use of machine learning, it was observed that by optimizing the water content and cementitious material in the concrete paste, it is possible to achieve the desired workability, compressive strength, and mechanical properties while minimizing the waste of raw materials [27]. The integration of artificial intelligence in carbon footprint estimation for construction brings enhanced accuracy, efficiency, predictive capabilities, optimization opportunities, data-driven insights, scalability, and continuous learning to support sustainable practices and environmental stewardship in the built environment. In this study, results of lightweight aggregate low carbon concrete were collected and trained using some machine learning algorithms, the accuracy of the predictions were evaluated. By adopting this tool to carefully optimize mix designs of light weight concrete, following the rigorous testing procedures of the data set used, the waste of raw materials in light concrete production from trial mixes can be minimized, contributing to more sustainable and environmentally friendly construction practices that satisfy the desired structural requirement for concrete production.

2. Review of some existing machine learning models

Utilizing recycled aggregates and SCMs in concrete production can significantly reduce the environmental impact of the construction industry by decreasing the demand for natural resources and lowering embodied carbon (eCO_2). Some of the available machine learning models in literatures include regression models, decision tree, neural networks, support vector machines, random forest, ensemble learning and deep reinforcement learning. Regression analysis is commonly used in predicting carbon emissions based on various input variables such as material types, construction methods, project parameters, and environmental factors. Machine learning prediction of carbonation depth in recycled aggregate concrete incorporating SCMs using regression models, specifically the Gradient Boosting Regression Tree (GBRT) model was utilized to enhance the prediction of carbonation depth in recycled aggregate concrete. The dataset used in the study comprised of 713 experimental data records on carbonation of recycled aggregate concrete. This dataset was utilized to develop and test the gradient boosting regression tree (GBRT) model for predicting carbonation depth in recycled aggregate concrete with different mineral additions. The GBRT model achieved a Root Mean Squared Error (RMSE) of 1.51, Mean Absolute Error (MAE) of 0.95, and a coefficient of determination (R^2) of 0.97. The model's robustness was confirmed by consistent performance across different random seeds [28]. Despite the gain with this model, the GBRT algorithm is considered a black-box model due to the lack of comprehensibility in its prediction process. While regression trees can be interpreted by analysing their structure, GBRT models consist of thousands of regression trees, making it challenging to visualize and interpret all the trees. This lack of transparency in the prediction process can be a setback in understanding the underlying mechanisms involved in the carbonation process of recycled aggregate concrete (RAC). A machine-learning regression model with Logical Analysis of Data (LAD) has been developed with the advantage of transforming independent variables of the original data into patterns, while the dependent response is regressed on these patterns instead of the observations. The Logical Analysis of Data Regression (LADR) model when compared with other models shows good performance. The results indicate that LADR-KM (Threshold $\tau = 4$): MSE = 26.70, R^2 = 90.50, MAE = 3.77, Linear Regression (LR): MSE = 109.37, R^2 = 61.19, MAE = 8.30, Support Vector Regression (SVR): MSE = 117.21, R^2 = 60.14, MAE = 8.20, Decision Tree Regression (DTR): MSE = 56.57, R^2 = 80.38, MAE = 5.54, and random Forest (RF): MSE = 22.35, R^2 = 92.19, MAE = 3.31 [29].

Decision tree algorithms are employed to analyse complex datasets and make decisions regarding material selection, construction processes, and design choices that minimize embodied carbon. Decision trees are easy to interpret and understand, making them valuable for explaining the reasoning behind predictions. They are often used to capture nonlinear relationships between input variables and the target output, which is common in concrete prediction tasks. Evaluation of the performance of different machine learning models, including decision tree (DT), extreme gradient boosting (XGBoost), and random forest (RF), in predicting the compressive strength of geo-polymer concretes have been studied [30]. The XGBoost model demonstrated the highest degree of accuracy among the models, with mean absolute error (MAE) of 2.07, mean absolute percentage error (MAPE) of 5.55, Nash–Sutcliffe (NS) of 0.98, correlation coefficient (R) of 0.99, R^2 of 0.98, root mean square error (RMSE) of 2.46, Willmott's index (WI) of 0.79, weighted mean absolute percentage error (WMAPE) of 0.046, Bias of 2.073 square index (SI) of 0.054, p-value of 0.027, and mean relative error (MRE) of −0.014. It is suggested that XGBoost introduces a second-order Taylor expansion, which increases accuracy and allows for customization of loss functions through gradient descent. This helps in preventing overfitting and results in better generalization. According to the National Council for Cement and Building Materials (NCBM), several key factors like the compressive strength of concrete, Water-Cement Ratio, Mineral Admixtures such as fly ash, ground granulated blast furnace, silica fume, and fine aggregates and Water play a crucial role in determining the strength, durability, and workability of concrete mixes.

Modelling non-linear mechanical behaviour of concrete at elevated temperatures poses several challenges due to the complex nature of the material response under extreme conditions. The mechanical properties of concrete, such as modulus of elasticity, compressive strength, and thermal conductivity, are highly temperature dependent. Modelling these properties accurately as they

change with temperature is crucial but challenging. Concrete exhibits non-linear behaviour under high temperatures, including thermal expansion, changes in stiffness (Young's modulus), and potential damage mechanisms. Capturing these non-linear effects in a model accurately is a significant challenge. Studies have shown that Artificial neural networks (ANNs) excel at mapping complex relationships between input and output parameters without the need for explicit mathematical equations. This capability is beneficial for capturing the non-linear behaviour of concrete under varying temperature conditions and shows that ANNs can produce superior results compared to traditional semi-analytical expressions when predicting stress-strain relationships in concrete at high temperatures [31]. ANN consists of interconnected artificial neurons organized in layers (input, hidden, and output layers) to process information and make predictions. The development of a deep neural network model to predict the compressive strength of rubber concrete was carried out to using the Deep neural network (DNN) to predict the compressive strength of rubber concrete. The DNN model for predicting the compressive strength of rubber concrete was developed through a systematic process using database of 223 experimental results with 12 input variables, including specimen geometry, binder, cement, cement replacement materials, superplasticizer, water, aggregates, and rubber components [32]. The developed DNN model exhibited high reliability with a coefficient of determination (R^2) of 0.98. The correlation results indicated a high accuracy level with $R^2 = 0.966$ for training and $R^2 = 0.98$ for testing datasets, demonstrating the feasibility of using the DNN model for predicting compressive strength with low error.

The main idea behind Support Vector Machine (SVM) is to find the optimal hyperplane that best separates different classes in a dataset. SVM is a supervised machine learning algorithm that is effective for both classification and regression tasks and can handle high-dimensional data efficiently through the use of kernel functions, allowing it to capture complex relationships in the data that may not be linearly separable in the original feature space. The combination of SVM and K-Fold cross validation in predicting the compressive strength of concrete in the marine environment was carried out. The dataset was divided into k parts, with one part used as the testing dataset and the remaining k-1 parts as the training dataset with the aim to find the optimal penalty parameter (c) and radial basis kernel parameter (g) for the SVM model. The range for searching c and g was set as 2^{-5} to 2^5. After cross-validation training, the optimal penalty parameter c and radial basis kernel parameter g were found to be $c = 84.448$ and $g = 0.19$ [33]. The key findings regarding the prediction performance of Support Vector Machine (SVM), Artificial Neural Network (ANN), and Decision Tree (DT) models shows that the optimized SVM model demonstrated the best prediction performance compared to the ANN and DT models. Based on the comparison results, the SVM model was selected as the optimal prediction model for concrete strength degradation in marine environments. Similar studies present the use of SVM to predict the flexural and compressive strength of cement mortar with nano-silica and micro-silica. Four different kernels were investigated for predicting the mechanical properties of cement mortar which include Radial Basis Function (RBF), Polynomial, Linear and Sigmoid. The input parameters used in the dataset for the SVM approach in predicting the flexural and compressive strength of cement mortar containing nano-silica and micro-silica. The RBF kernel uses a Gaussian function to map the input data into a higher-dimensional space, where the data points are transformed based on their distance from a centre and calculates the similarity between data points in the input space using a radial basis function, which allows for non-linear transformations and better fitting of the data. It is observed that the Radial Basis Function (RBF) kernel outperformed the polynomial, linear, and sigmoid kernels in predicting the mechanical properties of cement mortar [34]. SVMs adopt a structural risk minimization induction principle, aiming to minimize a bound-on generalization error of a model rather than just minimizing the error on the training data. This approach helps in avoiding overtraining and enhances generalization capability and have been successfully applied in various civil engineering problems, showcasing their versatility and effectiveness in different domains [35]. Prediction of compressive strength of high-performance concrete (HPC) using the Multivariate Adaptive Regression Splines (MARS) model in combination with Gradient Tree Boosting Machine (GBM) has been studied. The MARS model is employed as a feature selection method to reduce the number of input variables and capture non-linearity in the data. It prunes the features to an optimal number of knots using a generalized cross-validation procedure. The MARS model establishes an empirical equation based on the most important inputs, such as cement, blast furnace slag, water, superplasticizer, fine aggregate, and concrete age. The empirical equation developed by the MARS model in the study is crucial for predicting the compressive strength of high-performance concrete [36]. Equation (1) is based on the most important input variables identified by the MARS model and is used to estimate the compressive strength of concrete (CCS).

$$\begin{aligned}CCS = {} & 1.718 + 4.404h(Age-0.151)-5.967h(0.151-Age)-0.582h(0.954-Cement)+0.351h(0.360-Superplasticizer)+0.276h \\ & (BlastFurnaceSlag-0.209)-0.498h(0.210-BlastFurnaceSlag)+1.244h(Water-0.784)-0.082h(0.867-Fineaggregate)- \\ & 4.360h(Age-0.036)-0.448h(Water-0.128)-0.329h(Superplasticizer-0.106) \ldots. \end{aligned} \quad (1)$$

Where h is the hinge function MARS model is a piecewise linear function that is used to model the relationship between input variables and the target output. The MARS-GBM model demonstrates improved accuracy in predicting compressive strength compared to previous studies. It effectively utilizes six key variables to estimate the impact of predictors on concrete strength, highlighting the effectiveness of the MARS model in feature selection and prediction. The dataset used in the study consisted of a total of 1030 data sets with eight input variables. These input variables included parameters such as cement, blast furnace slag, water, superplasticizer, fine aggregate, and concrete age. The dataset was utilized to estimate the compressive strength of high-performance concrete (HPC). The study compared the performance of different models, including Kernel Ridge Regression (KRR) and Gaussian Process Regression (GPR), against the MARS-GBM model. The MARS-GBM model achieved a high correlation coefficient (R) of 0.965 during the testing phase, indicating a strong relationship between the predicted and actual compressive strength values. The MARS-GBM model outperformed other models, including Kernel Ridge Regression (KRR) and Gaussian Process Regression (GPR), in terms of statistical indices such as correlation coefficient (R), Mean Absolute Error (MAE), Normalized Root Mean Squared Error (NRMSE), Residual Standard Deviation (RSR), and coefficient of determination (cp) during both training and testing phases. This performance was

comparable to the KRR model, demonstrating the effectiveness of the MARS-GBM approach. With regards to prediction of self-compacting concrete using various machine learning model, the performance of Linear Regression, Ridge Regression, Lasso Regression, Decision Tree and Random Forest have been demonstrated. Linear Regression is used to establish the relationship between variables by fitting a linear equation to observed data points while Ridge Regression which is a regularization technique adds a penalty term to the linear regression equation to prevent overfitting. Lasso regression which is another regularization technique is used to adds a penalty term to the linear regression equation to promote sparsity in the model. The role of the Decision Tree as a non-linear supervised learning method is to partitions the data into subsets based on feature values to make predictions while that Random Forest builds multiple decision trees and combines their predictions to improve accuracy [37]. The dataset used in the study consisted of 99 samples of self-compacting concrete (SCC) mixtures. These samples were obtained through various procedures such as V-funnel, slump flow, and L-box during the experiment. Based on the performance metrics, the Random Forest Regressor model outperformed the other models with the lowest RMSE, MSE, and MAE values, as well as the highest R^2 value. In a similar study, two models such as Kernel Ridge Regression (KRR) and Neural Networks (NN) were utilized to predict the compressive strength of concrete based on various input parameters. The kernel type for KRR was Linear and Nonlinear (4th order polynomial), while the architecture of the neural network is Up to four layers with three hidden layers and one output layer. The performance of both models was evaluated using gradient descent with momentum weight bias learning and mean square error (MSE) on 182 concrete samples of dataset [38]. The optimal model performance was achieved by the Kernel Ridge Regression model with a nonlinear 4th order polynomial kernel due to its low MSE, high regression value, and superior performance compared to the Neural Network model. The nonlinear kernel allows the model to capture nonlinear relationships between the input parameters and the compressive strength of concrete and exhibited effectiveness in handling complex and nonlinear patterns in the data, leading to improved prediction accuracy.

3. Data set and concrete preparation

The concrete mixes used in the data set was obtained from the study by Nukah et al. [39], and was produced using CEMII Ordinary Portland cement (OPC) 42.5 R with density of approximately 2.4–3 g/cm^3 and Ground Granulated Blast Furnace Slag (GGBS). Densified silica fume Microsilica, Pozzolan, GFRC, 90 % was used as an additive and superplasticizer (Flowaid SCC, UK) as a high-water reducing agent. Alkaline solution was prepared from sodium hydroxide, NaOH DRYI29–1000 and sodium Silicate, Na_2SiO_3. The sodium silicate solution was made of density between 1.30 and 1.60 kg/L with pH value of 11–13, % w/w = 20–60 and molar ratio >1.6. The aggregate used was Lytag 4/14 mm with loose bulk density and particle density of 8.2 × 10− 7 kg/m^3 and 1.42 × 10 − 6 kg/m^3 and ordinary sand. Lytag was pre-soaked for 24 hours before use while the sodium hydroxide solution was prepared by adding 522 g of sodium hydroxide, NaOH pellet to 1 litre of distilled water. The alkaline solution was obtained at Na_2SiO_3/NaOH ratio of 2.5. Cement replacement in the concrete mix was carried using GGBS at 60–80 % for the estimation of compressive strength and embodied carbon (CO_2) footprint.

The workability of concrete specimen was measured using slump test and it varies between 50 and 80 mm while compressive strength test was carried out in Matest compressive testing machine at a loading rate of 3 kN/s and a start load of 10 kN and conforms to BS EN 12390 [40].

3.1. Calculation of concrete embodied carbon

Embodied carbon was measured using the embodied carbon coefficients, kgCO_{2e}/kgm^3 of concrete ingredients for Cradle to gate (A1-A3) and calculated with global warming potentials (GWP) presented in equation (2) as recommended in BS EN15978 [41].

$$\mathrm{EwP} = \sum_{i=1}^{n} \mathrm{cm(i)\ x\ Ec(i)} \tag{2}$$

Where i = material element in the concrete, cm = concrete materials (kg/m^3), EwP = global warming potentials (kgCO_{2e}/m^3), Ec = embodied carbon coefficient (kgCO_{2e}/kg).

Embodied carbon in concrete is calculated by determining the amount of embodied carbon (CO_2) emissions associated with the production of the materials used in the concrete mix. The calculation considers the carbon footprint of each constituent material, such as cement, aggregates, water, and any supplementary cementitious materials. The embodied carbon calculation is typically expressed in terms of kilograms of CO_2 emitted per kilogram of concrete produced. Calculation of embodied carbon can vary based on factors such as the source of materials, production processes, transportation distances, and energy sources used. Therefore, accurate and up-to-

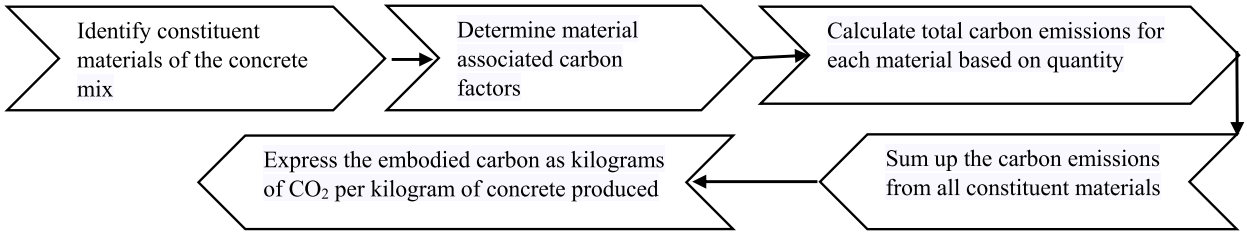


Fig. 3. Concrete embodied carbon calculation process flow.

date data on carbon emissions for each material are essential for a precise calculation of embodied carbon in concrete. The process flow in determining embodied carbon in concrete is shown in Fig. 3.

The Embodied carbon factors (E_{cf}) of the data set used in this study are shown in Table 1.

The dataset consisting of 64 concrete data points was utilized to model the behaviour of lightweight geopolymer concrete with the evaluation of its mechanical performance in terms of compressive strength and CO_2 footprint. The input features parameters used were concrete age, OPC, Lytag, sand, 60 % GGBS, 70 % GGBS, 80 % GGBS, superplasticizer, silica fume, alkaline binder ratio, and water to cement ratio. Multi-output output regression algorithms from scikit-learn python library were used to predict the output of compressive strength and embodied carbon.

3.2. Machine learning models

The training of machine learning models was carried out using Decision tree regression, XGBoost Regression, Neural network Regression, Ridge regression, Lasso regression. Support vector machine. Ridge regression, Random Forest, linear regression and implemented in Python 3.11. The training processes are conducted in the windows 10 platform with the processor of 11th Gen Intel(R) Core(TM) i5-1145G7 @ 2.60 GHz 2.61 GHz. The main machine learning and Python libraries used are scikit-learn 1.4.2 [42], seaborn 0.13.2 [43] scipy 1.13.0 [44], xgboost 2.0.3 [45], numpy 1.26.4 [44], pandas 2.2.2 [46], pickleshare 0.7.5 [47], joblib 1.4.2 [48], flask 3.0.3 [49], jinja2 3.1.4 [50], matplotlib 3.8.0 [51]. For each machine learning model, a random search on hyperparameters is performed to find the best performing model. It is noted that with the same number of combinations of hyperparameters, random search performs better than grid search and manual search for hyperparameter optimization [52] as larger hyperparameter search space leads to the better performance. The model was trained and evaluated with every hyperparameter combination to find the best performing one. In this study, the dataset was obtained from 64 concrete mix sample with target design strength of 20, 30, 40 and 50 MPa cured for 3, 7, 21 and 28 days. The dataset used for training in this study was scaled so that the magnitude of the variables does not produce erroneous or misleading statistics and to prepare data for machine learning. The methods used in scaling were normalization and standardization. For standardization, rescaling data was carried out, so it has a zero mean and unit variance while normalization was implemented by putting each observation on a relative scale between the values of 0 and 1 using scikit-learn preprocessing library.

Due to the non-linearity of concrete ingredients with compressive strength [53], the performance of the prediction was evaluated with machine learning regressions models with the fundamental of some of the models highlighted.

3.2.1. Adaboost regression

AdaBoost starts by fitting a weak regression model to the data. A weak learner is a model that performs slightly better than random chance on the given task it assigns initial weights to each observation in the training data. Observations that are difficult to predict are given higher weights, while observations that are easy to predict are given lower weights. AdaBoost adjusts the weights of the observations based on the errors made by the weak regression model. Observations with higher errors are given increased weights to focus the next model on the previously misclassified data points. The final prediction is made by aggregating the predictions of all the weak regression models in the ensemble, weighted by their individual performance.

Studies have demonstrated that this technique are popular and accurate in prediction [54,55]. The fundamental concept of AdaBoost regression as shown in Freund and Schapire [55] was demonstrated through equations (2)–(4) that demonstrate the process of boosting weak regression models to create a strong ensemble mode. Equations (3)–(5) represents the combination of weak regression models gt (x) weighted by αt to form the final ensemble model gi(x)

$$g_i(x) = \sum_{t=1}^{T} \propto_i g_t(x) \tag{3}$$

$$g\varepsilon_i(x) = \sum_{j=1}^{N} \omega_{ij} I\left(y_i = g_{i(X_i)}\right. \tag{4}$$

Where ϵi represents the error of the ensemble model $g_i(x)$ on the training data, where w_{ij} are the weights assigned to the observations based on their misclassification.

The weight αi assigned to the ensemble model $g_i(x)$ based on its error rate ϵi. is shown in equation (3).

Table 1
Embodied carbon factors E_{cf} of concrete materials.

Concrete material	E_{cf}	Reference
Lytag	0.249	[39]
GGBS	0.07	[39]
Cement	0.912	[39]
Superplasticizer	0.01	[39]
Silica fume	0.014	[39]
River sand	0.005	[39]
Sodium hydroxide	0.86	[39]
Sodium silicate	0.43	[39]
Granite	0.7	[39]

$$\propto_i = 1 + \frac{1}{2}\frac{(1-\varepsilon_i)}{\varepsilon_i} \tag{5}$$

By counting the number of correct and incorrect predictions, AdaBoost.RT calculates the error rate. When comparing AdaBoost.RT to boosting with AdaBoost.R2 and bagging using the M5 model tree as a weak learner, AdaBoost.RT showed a significant improvement in performance at 3.4 % to as much as 18.8 % [56].

3.3. Decision tree regression

Regression trees are a type of decision tree specifically designed to approximate real-valued functions rather than assigning class labels like in classification trees. In regression trees, each leaf typically contains a constant value, often the average value of the target attribute. Additionally, it has been shown that model trees is an extension of regression trees where the constant value in each leaf is replaced by a linear or nonlinear regression function. The concepts of Gini impurity and entropy are commonly used in decision tree algorithms, such as CART (Classification and Regression Trees), to make decisions about splitting nodes during the tree-building process by Czajkowski and Kretowski [57] as shown in equations (6) and (7).

The entropy for a node is calculated in equations (6) and (7) as

$$Entropy = \sum_{i=1}^{C} P_i \, log_2(P_i) \tag{6}$$

and.

The Gini impurity for a node is calculated as

$$Gini = 1 - \sum_{i=1}^{C} \left(P_i^2\right) \tag{7}$$

Where.

- c is the number of classes.
- pi is the probability of randomly selecting a sample of class i in the node.

Five evolutionary tree inducers with various tree representations were validated and highlighted their advantages and disadvantages in the work of Czajkowski and Kretowski [57]. A unified framework was proposed for the global induction of mixed trees that could incorporate both univariate and oblique tests in internal nodes and regression planes and models in leaves. By modifying the representation in different variants of the mutation operators during the evolutionary tree induction, it was possible to to switch the node representation.

Combination of algorithms in application to regression have shown significant improvement. Decision tree (DT) algorithm was combined with linear regression to improve data classification to mitigate the impact of these small examples on classification accuracy. It was observed that DT algorithm classifies data at the upper level, while linear regression is applied to build models at the leaf level of the tree for more granular classification [58].

3.3.1. XGBoost regression

XGBoost is based on the concept of gradient boosting, where multiple weak learners (decision trees) are combined to create a strong learner. Each new tree corrects errors made by the previous ones, leading to improved overall performance. It uses gradient descent optimization to minimize the loss function iteratively. It calculates the gradient of the loss function with respect to the model parameters and updates the parameters in the direction that minimizes the loss. The concept of XGBoost is related to the regularization term used in the algorithm to control the complexity of the model and prevent overfitting and was presented in Li et al. [59], as shown in equation (8):

$$\Omega(\mathrm{ft}) = \gamma \mathrm{T} + \tfrac{1}{2}\,\lambda \sum_{j=1}^{T} \omega_j^2 \tag{8}$$

Where:

- $\Omega(ft)$ represents the regularization term.
- γ controls the complexity of the tree by penalizing the number of leaf nodes (T).
- λ is the regularization parameter that penalizes the L2 norm of the weights (w) of the leaf nodes.
- T is the number of leaf nodes in the tree.
- wj is the weight function of the j-th leaf node.

When compared to other algorithmic models such as SVR, SGDR, LR, DTR, and RF, the XGBoost algorithm showed the best optimization performance with the highest correlation coefficient and the smallest error, highlighting its superiority [60].

3.4. Neural network regression

Neural networks in regression algorithms, such as the Generalized Regression Neural Network (GRNN), utilize a network of

interconnected nodes to model complex relationships between input and output variables. In a regression neural network, there are typically three main layers: input layer, hidden layer(s), and output layer. The input layer receives the input data, which is then processed through the hidden layers where the network learns the underlying patterns in the data, and finally, the output layer provides the predicted values. Each node in the network performs a weighted sum of its inputs, applies an activation function to the sum, and passes the result to the next layer. This process allows the network to capture non-linear relationships between variables. During the training phase, the network adjusts the weights of connections between nodes to minimize the difference between predicted and actual output values. This is typically done using optimization algorithms like gradient descent or Levenberg–Marquardt algorithm. In regression tasks, the neural network aims to predict a continuous output value based on input variables. The network learns the mapping between input features and the target output, allowing it to make predictions on new and unseen data.

From equations (9)–(12), the input to a node in a neural network is calculated by Specht [61,62]
as the weighted sum of the inputs such that:

$$Z_j(x) = \sum_{i=1}^{n} \omega_{ij} * \ X_i + b_j \tag{9}$$

where:

- *zj* is the total input to node *j*,
- *wij* is the weight of the connection between input node *i* and hidden node *j*,
- *xi* is the input value at node *i*,
- *bj* is the bias term for node *j*

The total input is then passed through an activation function to introduce non-linearity such that.

$$a_j = f(z_J) \tag{10}$$

where:

- *aj* is the output of node *j*,
- $f(\cdot)$ is the activation function (e.g., sigmoid, ReLU, tanh).

For a simple feedforward neural network, the output is calculated as the weighted sum of the hidden layer outputs.

$$y_k = \sum_{j=1}^{m} v_{jk} * \ a_j \ + c_k \tag{11}$$

where:
yk is the final output of the network, v_{jk} is the weight of the connection between hidden node *j* and output node *k*, *aj* is the output of hidden node *j*, and c_k is the bias term for output node *k*.

Equation (10) shows that the network is trained by minimizing a loss function L, such as Mean Squared Error (MSE), which measures the difference between predicted and actual outputs such that.

$$L = \frac{1}{N} \sum_{i=1}^{N} (y_i \ - \ \hat{y}_i)^2 \tag{12}$$

where:
N is the number of training samples, y_i is the actual output, and $\hat{y}_i$ is the predicted output.

In utilizing the Generalized Regression Neural Network (GRNN) for river suspended sediment estimation as carried out in Cigizoglu and Alp [63]. The performance of the GRNN model was assessed using metrics based on the best performance criteria, such as lowest Mean Squared Error (MSE) and highest determination coefficient (R^2). It has been suggested that GRNN is advantageous in real-time environments with sparse data, as it can define regression surfaces instantly even with minimal sample points, making it suitable for applications requiring quick adaptation [61].

3.4.1. Ridge regression

The main concept of ridge regression is to address the issue of multicollinearity in linear regression models. Multicollinearity occurs when independent variables in a regression model are highly correlated with each other, leading to unstable estimates of the regression coefficients.

In ridge regression, a penalty term is added to the ordinary least squares (OLS) estimation to shrink the coefficients towards zero, reducing the impact of multicollinearity. This penalty term is controlled by a tuning parameter (often denoted as *k*), which determines the amount of shrinkage applied to the coefficients. By introducing this penalty term, ridge regression helps to stabilize the estimates, particularly when the number of predictors is large compared to the number of observations. This aims to improve the predictive performance and stability of the regression model by balancing the bias-variance trade-off through the introduction of a regularization term in the estimation process.

Given a linear regression model in equation (13) and (14).

$$y = X\beta + \epsilon \tag{13}$$

Table 2
Data training and prediction.

Compressive strength	Concrete age	CEMII	Lytag	Sand	60 % GGBS	70 % GGBS	80 % GGBS	SP	SF	A/B ratio	w/c ratio	Embodied carbon
3.53	3	260	812	437	0	0	0	13	2.6	0	0.37	442
4.38	3	330	780	420	0	0	0	16	3.3	0	0.18	499
6.38	3	475	670	361	0	0	0	24	4.75	0	0.13	602
7.70	3	570	608	327	0	0	0	28	5.7	0	0.1	673
3.86	3	104	812	437	156	0	0	13	2.6	0.3	0.04	310
3.49	3	78	812	437	0	182	0	13	2.6	0.3	0.05	289
5.04	3	52	812	437	0	0	208	13	2.6	0.3	0.08	267
6.85	3	132	780	420	198	0	0	16	3.3	0.6	0	333
5.36	3	99	780	420	0	231	0	16	3.3	0.3	0.06	305
6.45	3	66	780	420	0	0	264	16	3.3	0.3	0.06	277
5.94	3	190	670	360	285	0	0	24	4.75	0.3	0.03	362
5.76	3	142	670	360	0	332	0	24	4.75	0.5	0	322
4.05	3	95	670	360	0	0	380	24	4.75	0.6	0	282
5.38	3	228	608	327	342	0	0	28	5.7	0.6	0.6	385
4.38	3	171	608	327	0	399	0	28	5.7	0.5	0.6	337
5.80	3	114	608	327	0	0	456	28	5.7	0.5	0.6	289
2.80	3	260	812	437	0	0	0	13	2.6	0	0.37	442
20.79	7	260	812	437	0	0	0	13	2.6	0	0.37	442
25.81	7	330	780	420	0	0	0	16	3.3	0	0.18	499
37.54	7	475	670	361	0	0	0	24	4.75	0	0.13	602
45.29	7	570	608	327	0	0	0	28	5.7	0	0.1	673
22.73	7	104	812	437	156	0	0	13	2.6	0.3	0.04	310
20.55	7	78	812	437	0	182	0	13	2.6	0.3	0.05	289
29.63	7	52	812	437	0	0	208	13	2.6	0.3	0.08	267
40.27	7	132	780	420	198	0	0	16	3.3	0.6	0	333
31.50	7	99	780	420	0	231	0	16	3.3	0.3	0.06	305
37.92	7	66	780	420	0	0	264	16	3.3	0.3	0.06	277
34.97	7	190	670	360	285	0	0	24	4.75	0.3	0.03	362
33.89	7	142	670	360	0	332	0	24	4.75	0.5	0	322
23.83	7	95	670	360	0	0	380	24	4.75	0.6	0	282
31.63	7	228	608	327	342	0	0	28	5.7	0.6	0.6	385
25.75	7	171	608	327	0	399	0	28	5.7	0.5	0.6	337
34.14	7	114	608	327	0	0	456	28	5.7	0.5	0.6	289
22.19	21	330	780	420	0	0	0	16	3.3	0	0.18	499
31.68	21	475	670	361	0	0	0	24	4.75	0	0.13	602
33.72	21	570	608	327	0	0	0	28	5.7	0	0.1	673
20.37	21	104	812	437	156	0	0	13	2.6	0.3	0.04	310
18.06	21	78	812	437	0	182	0	13	2.6	0.3	0.05	289
23.45	21	52	812	437	0	0	208	13	2.6	0.3	0.08	267
29.202	21	132	780	420	198	0	0	16	3.3	0.6	0	333
23.92	21	99	780	420	0	231	0	16	3.3	0.3	0.06	305
27.20	21	66	780	420	0	0	264	16	3.3	0.3	0.06	277
24.73	21	190	670	360	285	0	0	24	4.75	0.3	0.03	362
28.48	21	142	670	360	0	332	0	24	4.75	0.5	0	322
17.91	21	95	670	360	0	0	380	24	4.75	0.6	0	282
33.70	21	228	608	327	342	0	0	28	5.7	0.6	0.6	385
27.50	21	171	608	327	0	399	0	28	5.7	0.5	0.6	337
32.38	21	114	608	327	0	0	456	28	5.7	0.5	0.6	289
24.66	28	260	812	437	0	0	0	13	2.6	0	0.37	442
33.11	28	330	780	420	0	0	0	16	3.3	0	0.18	499
47.29	28	475	670	361	0	0	0	24	4.75	0	0.13	602
50.33	28	570	608	327	0	0	0	28	5.7	0	0.1	673
30.40	28	104	812	437	156	0	0	13	2.6	0.3	0.04	310
26.95	28	78	812	437	0	182	0	13	2.6	0.3	0.05	289
35.00	28	52	812	437	0	0	208	13	2.6	0.3	0.08	267
43.58	28	132	780	420	198	0	0	16	3.3	0.6	0	333
35.70	28	99	780	420	0	231	0	16	3.3	0.3	0.06	305
40.60	28	66	780	420	0	0	264	16	3.3	0.3	0.06	277
36.92	28	190	670	360	285	0	0	24	4.75	0.3	0.03	362
42.51	28	142	670	360	0	332	0	24	4.75	0.5	0	322
26.73	28	95	670	360	0	0	380	24	4.75	0.6	0	282
50.30	28	228	608	327	342	0	0	28	5.7	0.6	0.6	385
41.04	28	171	608	327	0	399	0	28	5.7	0.5	0.6	337
48.32	28	114	608	327	0	0	456	28	5.7	0.5	0.6	289

w/c = water to cement ratio, A/B = alkaline binder ratio, SP = superplasticizer, SF = silica fume.

Where:

y is the vector of observed responses, X is the design matrix of predictors, β is the vector of regression coefficients to be estimated, and ϵ is the vector of errors. The ridge regression estimator for the coefficient vector β

$$\beta_{ridge} = \left(X^T X + kI\right)^{-1} X^T y \tag{14}$$

Where:

k is the ridge parameter that controls the amount of shrinkage, I represent the identity matrix, X^T denotes the transpose of matrix X and β_{ridge} is the ridge regression coefficient estimate.

By adding the kI term to the matrix X^T X in the normal equations, ridge regression introduces regularization to the estimation process. This regularization helps to stabilize the estimates, particularly in cases of multicollinearity where the matrix X^T X may be ill-conditioned [64]. The algorithm combats the "curse of dimensionality" by utilizing kernel functions. Kernel functions allow the construction of a linear regression function in a high dimensional feature space, which corresponds to a non-linear regression in the

Table 3
Dataset from other literatures with varying light weight aggregate.

Reference	Compressive strength	LWA	OPC	PFA	Sand	SF	SP	w/c	LWA type
Ergül Yaşar et al. [67]	28.3	500	300	0	150	0	0	0.55	Pumice
	29	400	275	100	150	0	0	0.55	Pumice
Domagała [68]	13.8	338	308	0	406	0	0	0.55	Lytag
	16.9	446	287	0	458	0	9	0.37	Lytag
	37.1	338	749	0	406	0	0	0.55	Lytag
	47.6	446	699	0	458	0	9	0.37	Lytag
Shannag [69],	29.3	400	550	0	200	0	1	0.625	Volcanic tuff
	28.8	380	548	0	199.3	20	3	0.66	Volcanic tuff
	38	360	546	0	198.6	40	3	0.69	Volcanic tuff
	43.2	340	550	0	200	60	4	0.73	Volcanic tuff
	27.7	380	548.2	20	199.2	0	2	0.65	Volcanic tuff
	22.5	360	546.6	40	198.7	0	1.5	0.69	Volcanic tuff
	32.5	360	546.3	20	198.6	20	3	0.70	Volcanic tuff
	32.4	340	544.6	40	198	20	2	0.73	Volcanic tuff
	39	340	544.3	20	197.9	40	4	0.73	Volcanic tuff
	33.7	320	542.6	40	197.3	40	4	0.77	Volcanic tuff
	36.7	320	548.2	20	199.2	60	3	0.78	Volcanic tuff
Sikora et al. [70],	15	480	56.4	0	34	0	7.2	0.4	Liaver
	16.1	432	56.4	0	34	48	8.4	0.44	Liaver
	16	474	56.4	0	34	0	7.5	0.39	Liaver
	17.5	468	56.4	0	34	0	9	0.38	Liaver
	13.7	456	56.4	0	34	0	9.5	0.37	Liaver
	12	432	56.4	0	34	0	10.8	0.33	Liaver
	2.85	200	98	0	0	0	5	0.7	Liaver
	4.3	180	98	0	0	20	6	0.78	Liaver
	1.35	197.5	98	0	0	0	6	0.70	Liaver
	1.2	195	98	0	0	0	6.5	0.69	Liaver
	0.97	190	98	0	0	0	7.2	0.68	Liaver
	0.75	180	98	0	0	0	7.5	0.67	Liaver
	12.28	500	77	0	108	0	2.5	0.38	Liaver
	13.08	500	75	0	53	0	2.5	0.39	Liaver
	13.93	500	74	0	0	0	2.5	0.42	Liaver
	17.65	500	0	0	104	0	2.5	0.42	Liaver
	17.85	500	37	0	52	0	2.5	0.42	Liaver
	19.22	500	0	0	52	0	2.5	0.43	Liaver
Sim et al. [71],	46.2	667	401	0	334	0	0	0.30	Lytag
	43.7	667	401	0	311	0	0	0.30	Lytag
	40.4	667	401	0	622	0	0	0.30	Lytag
	40.5	571	401	0	373	0	0	0.35	Lytag
	37.3	571	401	0	348	0	0	0.35	Lytag
	34.2	571	401	0	695	0	0	0.35	Lytag
	35.6	500	401	0	403	0	0	0.4	Lytag
	33.5	500	401	0	375	0	0	0.4	Lytag
	30.4	500	401	0	750	0	0	0.4	Lytag
	39.4	465	401	0	417	0	0	0.43	Lytag
	37.8	426	401	0	866	0	0	0.469484	Lytag
	35.5	364	401	0	862	0	0	0.549451	Lytag
Shafigh et al. [72],	41.38	550	333	0	891	0	4.4	0.35	Crushed OPS
	53.05	550	333	0	713	0	6	0.305	Crushed OPS
	43.25	500	435	0	726	0	6.1	0.354	Crushed OPS
	34.29	360	381	0	826	0	6.5	0.448	Crushed OPS

OPS – Oil palm shell, SP – Superplasticizer, SF – Silica fume, LWA – Lightweight aggregate.

input space. By using kernel functions, the algorithm avoids the need to carry out computations in the high dimensional space, thus overcoming the computational difficulties associated with many parameters [65].

3.4.2. Lasso regression

The Lasso (Least Absolute Shrinkage and Selection Operator) regression is a popular method used for variable selection and regularization in linear regression models. The Lasso regression aims to minimize the residual sum of squares subject to the constraint that the sum of the absolute values of the coefficients is less than a constant.

It is expressed mathematically by Alhamzawi and Ali [66] as:

$$\text{Minimise} \sum_{i=1}^{n} \left(y_i - \beta 0 - \sum_{j=1}^{p} \left(x_{ij} \beta_j \right. \right. \tag{15}$$

subject to

$$\sum_{j=1}^{p} |\beta_j| \leq t \tag{16}$$

Where n: Number of observations in the dataset, p: Number of predictor variables (features) in the model. yi: Observed value of the dependent variable for the ith observation. $\beta 0$: Intercept term in the linear regression model. xij: Value of the jth predictor variable for the ith observation. βj: Coefficient associated with the jth predictor variable and t: Constant representing the constraint on the sum of absolute values of coefficients.

3.4.3. Support vector machine

Support Vector Machine (SVM) aims to find the optimal hyperplane that best separates different classes in a dataset by maximizing the margin between the classes. SVM constructs a hyperplane using equation (17) in a high-dimensional space that acts as a decision boundary to classify data points.

It is formulated as:

$$\text{Minimise } \frac{1}{2}||w||^2 \text{ subject to } y\_i(w * x\_i + b) >= 1 \tag{17}$$

for all I where w represents the weights of the features, x_i represents the feature vectors, b is the bias term and y_i is the class label (+1 or −1).

The dataset for 7 and 28 days of concrete age used in this study has been reported by Nukah et al. [39], in addition to that of age 3 and 21 days in this study. The study employed advanced AI optimization techniques, including Decision tress regression, linear regression, Xgboost regression, random forest regression, adaboost regression, support vector machine (SVM), Ridge regression, and neural network models. These models were used to predict compressive strength and embodied carbon CO_2 emissions from the GGBS-LC geopolymer concrete materials. The performance of the models was evaluated using metrics such as Mean Absolute Error (MAE) and Root Mean Square Error (RMSE). The dataset is shown in Table 2. The performance of the model and the dataset in this study were compared with dataset from other literatures with varying light weight aggregate type as shown in Table 3.

3.5. Exploratory data analysis (EDA)

Exploratory Data Analysis (EDA) plays a crucial role in machine learning by helping data scientists understand the underlying patterns, relationships, and characteristics of the data before applying machine learning algorithms. EDA helps in understanding the structure of the data, identifying missing values, outliers, and anomalies, and gaining insights into the distribution of variables [73]. EDA was used for Data Understanding, Principal component analysis (PCA), Feature Selection (FS), feature engineering (FE), Data Preprocessing, Model Selection, Assumption Checking and Visualization**.** EDA was also used in understanding the structure of the data, identifying missing values, outliers, and anomalies, and gaining insights into the distribution of variables. It was able to show visualization techniques like histograms, box plots, scatter plots, and heat maps to visually represent the data, making it easier to interpret and communicate findings.

EDA in Python was carried out using libraries such as Pandas, NumPy, Matplotlib, Seaborn, and Plotly. The following were carried during the EDA on the dataset. Pandas was used to read data from MICROSOFT CSV and Excel.

3.5.1. Data Cleaning

The training data were read with the raw data into Python using Panda's and NumPy library. The structure of the data was adjusted by examining its shape, data types, summary statistics, and initial observations. Potential issues with missing values, outliers, or inconsistent data-were carried out. The issue of missing values was examined by more advanced techniques like interpolation and other machine learning models. Check for duplicate rows were carried out since duplicates can skew analysis and model training results. Transformations such as scaling, normalization carried out on the variables to prepare the data for modelling.

3.5.2. Data visualization

Data visualization in Python was carried out using Matplotlib, seaborn and Panda's libraries. Matplotlib is a widely used plotting library in Python that provides a variety of plots, including line plots, bar plots, scatter plots, histograms, and more. It allows for

detailed customization of plots, such as labels, colors, and styles.

Jupyter Notebooks was used for the model code development, interactive analysis, and visualization. Pycharm was also used for the user interface prediction application for compressive strength and embodied carbon tool development. The process of the Machine learning model development was carried out in the sequence as shown in Fig. 4.

4. Result and discussion

A quantile-quantile (Q-Q) plot was used as a graphical tool used to assess whether a set of data follows a specific distribution. In a Q-Q plot, the quantiles of the observed data are plotted against the quantiles of a theoretical distribution (such as a normal distribution). If the data points fall along a straight line, it indicates that the data closely follows the theoretical distribution [74]. It can be seen in Fig. 5 that the dataset is not significantly Deviated from the straight line in the Q-Q plot and hence do not suggest departures and skewness from the assumed distribution. However, the dataset was further put to standard scaling in python for more refining before data training, testing and evaluation. 30 % of the dataset was used for testing and 70 % for training and correlation matrix of the input features shown in Fig. 6. A strong correlation with concrete age with compressive strength is suggested. Strong influence of cement, 60 % GGBS, silica fume and superplasticizer on compressive strength is observed compared to alkaline binder ratio and water to cement ratio that has low influence. Sand and Lytag appears not to make any significant impact. The correlation between compressive strength and embodied is 1 and suggest a linear relationship which implies that an increase in compressive strength results to an increase in embodied carbon. Silica fume and superplasticizer has a high correlation with compressive strength at value of 0.25 greater than CEMII with value of 0.16. The duration of curing in days have the greatest impact on compressive strength with a correlation value of 0.67.

4.1. Model performance

Mean squared error (MSE) calculates the average of the squares of the errors. The errors are represented as the difference between the predicted and actual values.

$$MSE = \frac{1}{n}\left(\sum_{i=1}^{n}(yi - \widehat{y}_i)^2\right.$$

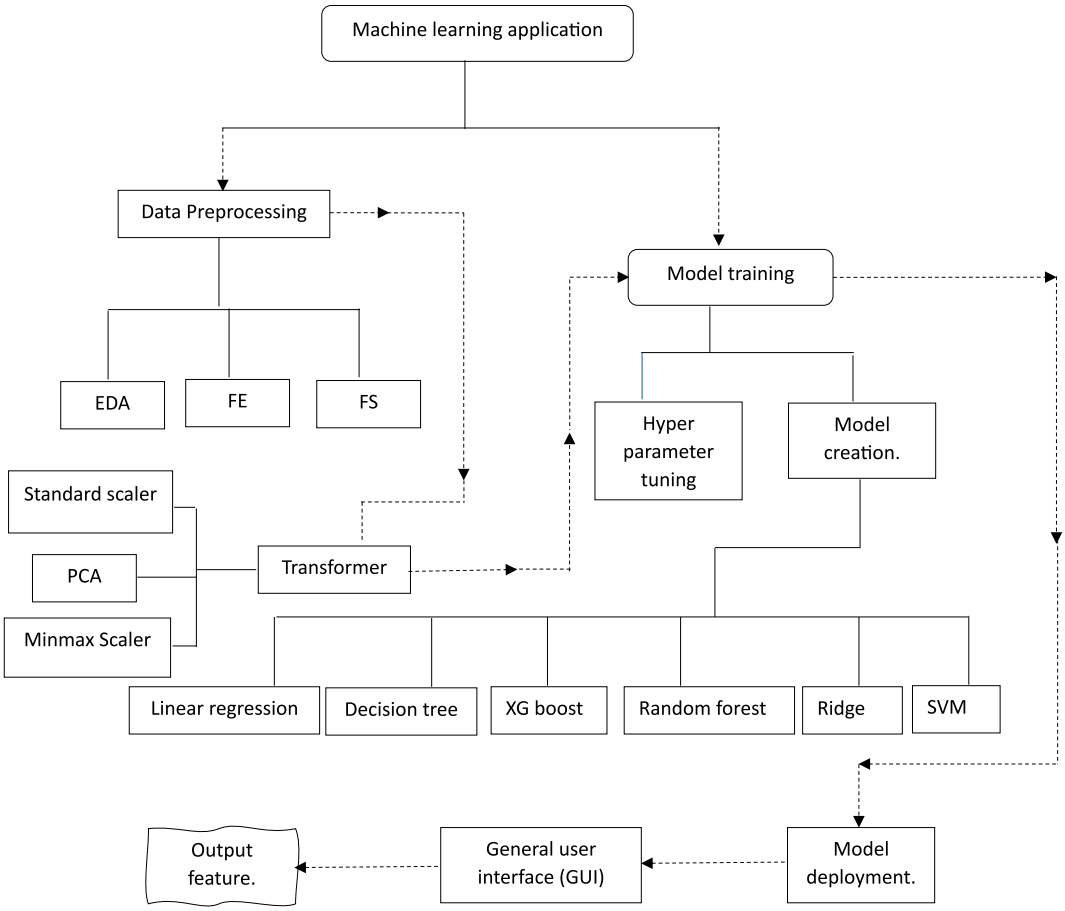
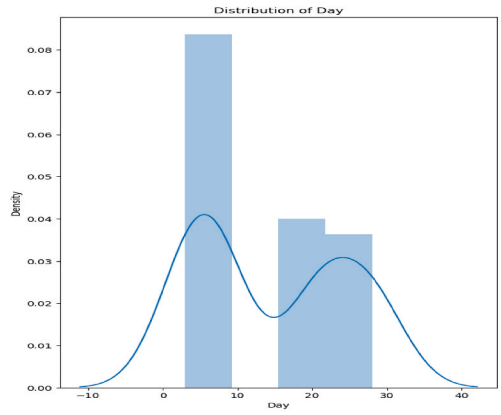
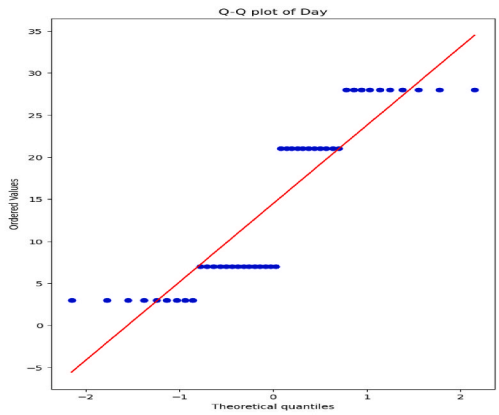


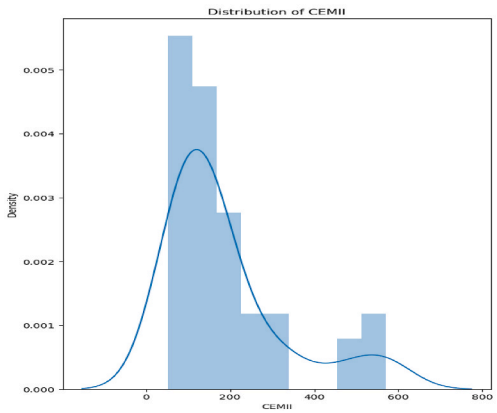
Fig. 4. Proposed learner model.



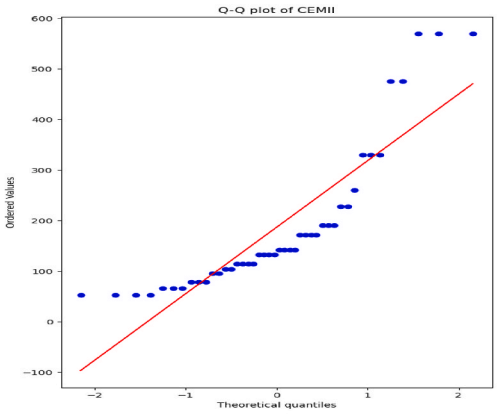
(a) Normal distribution of day



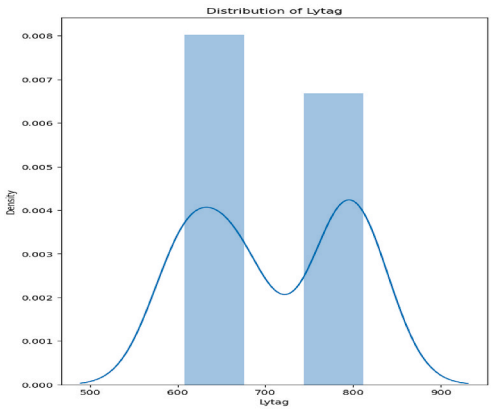
(b) Q-Q plot of day



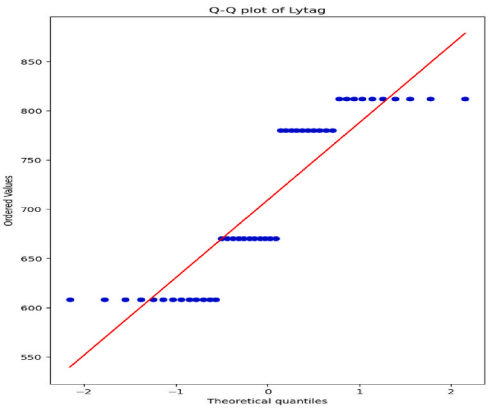
(c) Normal distribution of CEMII



(d) Q-Q plot of CEMII



(e) Normal distribution of Lytag

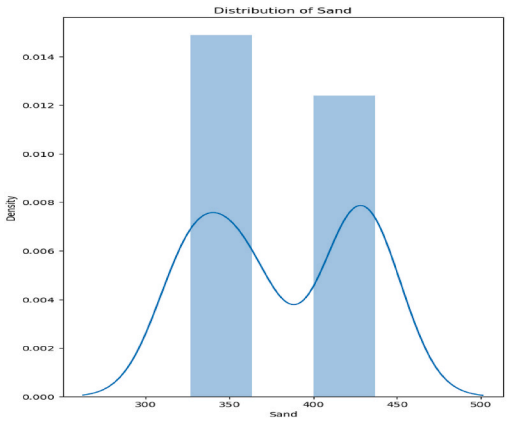


(f) Q-Q plot of Lytag

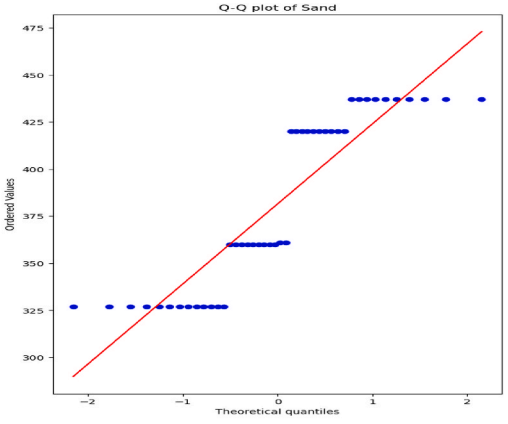
Fig. 5. (a–j): Distribution of input features of dataset for this study.

where *yi* is the actual values and $\hat{y}_i$ are the predicted values, and *n* is the number of samples. it penalizes larger errors more heavily than smaller errors due to squaring each error term. It gives a sense of the average magnitude of the errors the model makes.

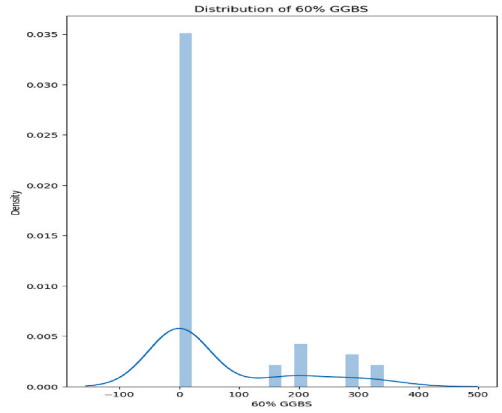
Root means square error (RMSE) is the square root of MSE, which gives the error metric in the same unit as the target variable. it is expressed as:



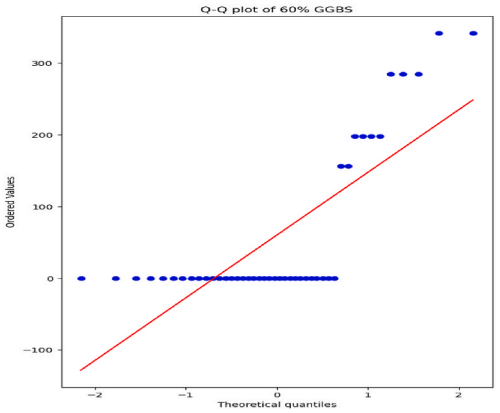
(g) Normal distribution of sand



(h) Q-Q plot of sand



(i) Normal distribution of 60% GGBS



(j) Q-Q plot of 60% GGBS

Fig. 5. (*continued*).

$$RMSE = \sqrt{MSE}$$

RMSE provides an idea of how much the model's predictions deviate, on average, from the actual values.

Mean absolute error (MAE) calculates the average of the absolute differences between predicted and actual values.

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |yi - \widehat{y}_i|$$

where yi is the actual values and $\widehat{y}_i$ are the predicted values, and n is the number of samples

MAE provides a measure of the average magnitude of errors without considering their direction. It's less sensitive to outliers compared to MSE because it doesn't square the differences.

Mean absolute percentage error (MAPE) calculates the average of the absolute percentage differences between predicted and actual values.

$$MAPE = \frac{100}{n}\sum_{i=1}^{n} \left|\frac{yi - \widehat{y}_i}{yi}\right|$$

MAPE gives a relative measure of the size of errors compared to the magnitude of the actual values. It's expressed as a percentage, making it easy to interpret and compare across different datasets.

The explained variance score (also known as R^2 *s* core or coefficient of determination) measures the proportion of the variance in the dependent variable that is predictable from the independent variables.

$$R^2 = 1 - \frac{var(yi - \widehat{y}_i)}{var(yi)}$$

R^2 score ranges from 0 to 1, where 1 indicates perfect predictions and 0 indicates that the model's predictions are no better than simply

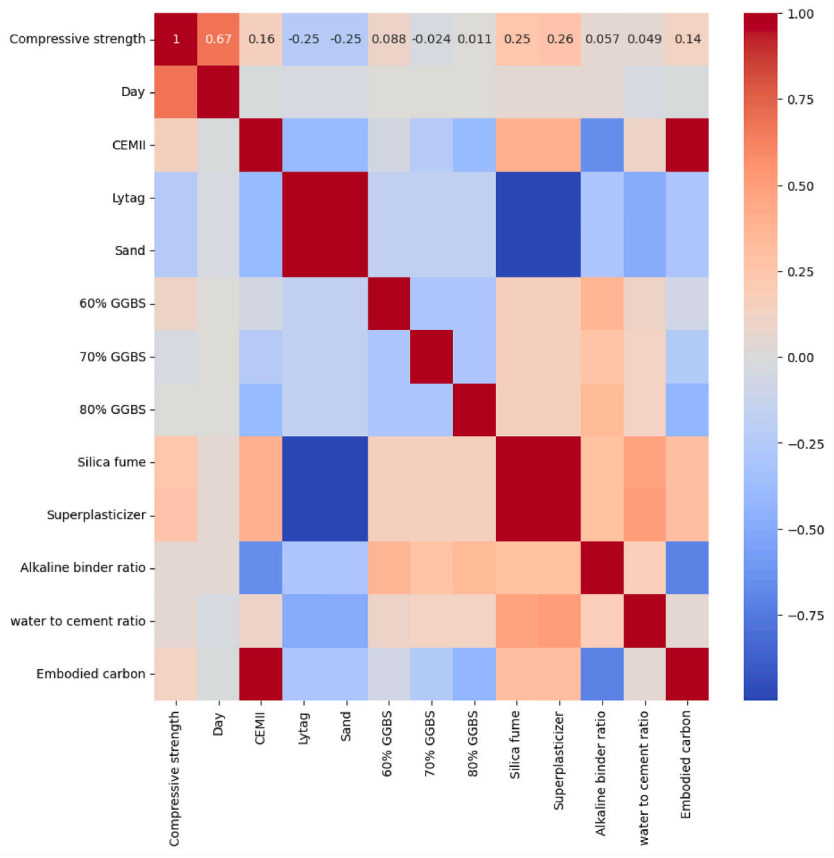


Fig. 6. Input features correlation matrix.

predicting the mean of the target variable. It's a measure of how well the model captures the variance in the data.

R squared R^2 score measures the proportion of the variance in the dependent variable (y) that is predictable from the independent variables (X). It is a statistical measure that indicates how well the regression predictions approximate the real data points.

$$R^2 = 1 - \frac{var(yi - \widehat{y}_i)^2}{var(yi - \overline{y})^2}$$

Where $yi =$ the actual values, $\widehat{y}_i =$ predicted values, $\overline{y} =$ mean of the actual values, and n is the number of samples. R^2 score allows for comparison between different models. A higher R^2 score suggests a better fit. R^2 score complements MSE, RMSE, MAE, and MAPE by providing a standardized measure of how well the regression model fits the data, which is crucial for assessing predictive accuracy and model quality

Table 4 compares the performance of different regression models based on various metrics such as Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and the percentage of explained variance. Additionally, it includes the R-squared (R^2) values for both the present study and the other literatures.

1. Linear Regression: The linear regression model shows relatively low values for MSE, RMSE, MAE, and MAPE compared to other models. It also has a decent R-squared value, indicating a reasonable fit to the data.
2. Lasso Regression: The Lasso regression model has higher error metrics compared to linear regression but slightly better performance in terms of R-squared. It penalizes the absolute size of the regression coefficients, potentially leading to a more interpretable model.
3. Ridge Regression: Like Lasso regression, Ridge regression also has higher error metrics compared to linear regression. It performs slightly worse than Lasso in terms of R-squared but may offer better stability by shrinking the coefficients.
4. Decision Tree Regressor: The decision tree model shows lower error metrics compared to linear, Lasso, and Ridge regression. It has a relatively high R-squared value, indicating a good fit to the data. However, decision trees are prone to overfitting.
5. XGBRegressor: The XGBoost regression model performs well with low error metrics and high R-squared values. XGBoost is an ensemble learning method known for its speed and performance.

Table 4
Model performance comparison.

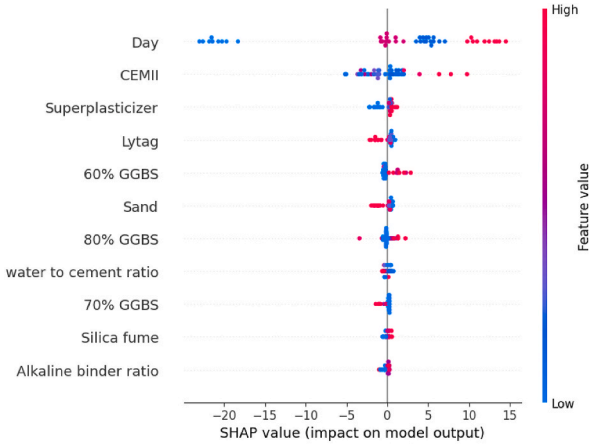
Model	MSE		RMSE		MAE		MAPE %		Explained variance		R-squared (R^2)	
	Present study	Literature	Present study	Literature	Present study	Literature	Present study	Literature	Present study	Literature	Present study	Literature
Linear Regression	59.44	980.28	7.71	31.31	5.98	20.79	46.2	17.52	0.81	0.77	0.79	0.75
Lasso Regression	462.41	921.49	21.50	30.36	16.17	20.29	58.9	17.13	0.82	0.79	0.78	0.76
Ridge Regression	483.07	950.28	21.98	30.83	15.91	20.69	48.0	17.48	0.79	0.78	0.77	0.75
Decision Tree Regressor	173.46	395.54	13.17	19.89	8.51	10.00	70.89	10.56	0.56	0.80	0.55	0.79
XGBRF Regressor	50.15	770.43	7.08	27.76	5.26	14.48	11.76	13.12	0.97	0.84	0.96	0.83
Neural Network:	6886.93	18142.88	82.99	134.70	56.65	83.43	64.49	28.98	0.56	0.24	0.26	−1.03
Random Forest Regressor	124.12	297.75	11.14	17.26	8.97	12.19	69.67	11.79	0.66	0.83	0.64	0.83

6. Neural Network: The neural network model exhibits very high error metrics compared to the other models. It has a negative R-squared value when applied on dataset from other literaures and a low value in this present study, indicating a poor fit to the data. This suggests that the neural network model is not suitable for this dataset. This can liken the model architecture defined by the number of neurons in the input and the output layers
7. Random Forest Regressor: The random forest model shows good performance with low error metrics and a high R-squared value. Random forests are known for handling complex relationships in data and reducing overfitting.

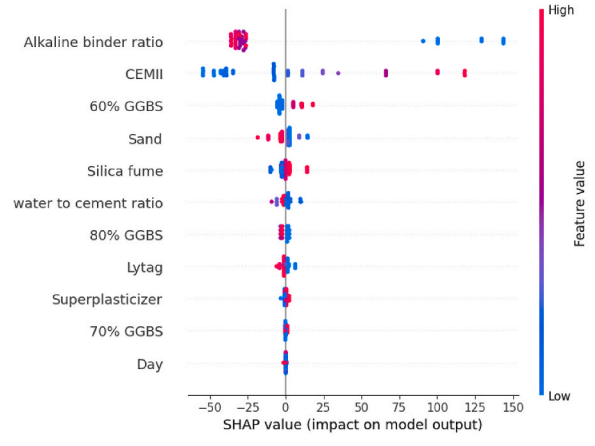
The XGBoost regressor is suggested to be the best-performing models among those listed and agrees with result from previous study [75]. These models have low error metrics, high R-squared values, and are known for their robust performance in various scenarios. The neural network, on the other hand, shows poor performance in this study, indicating that it may not be the most suitable model for this dataset.

4.2. SHapley additive explanation (SHAP) analysis of features

SHAP analysis is a method that explain the prediction of an instance by attributing each feature a certain importance value. The sum of these values equals the difference between the average prediction and the instance's prediction. SHAP was carried out by importing Shap library into python3.11 environment and applied on the trained model. Shap explainers was created, and the values generated for each target. The SHAP summary is shown in Fig. 7(a) with target output of compressive strength. The SHAP value represents the impact of a feature on the model's output. Positive values push the prediction higher, and negative values push it lower. The color gradient (from blue to red) indicates the feature value from low to high. This helps in understanding the relationship between the feature value and its impact. Insights from the Plot indicates that Day has a wide spread of SHAP values and is mostly red on the positive side, it suggests that higher values of Day (more curing time) lead to higher strength prediction. Superplasticizer shows a mix of red and blue dots spread around zero, it indicates that its impact varies depending on other conditions. Also, Water to Cement Ratio has mostly blue dots on the negative side. It suggests that a higher ratio (more water relative to cement) generally decreases the

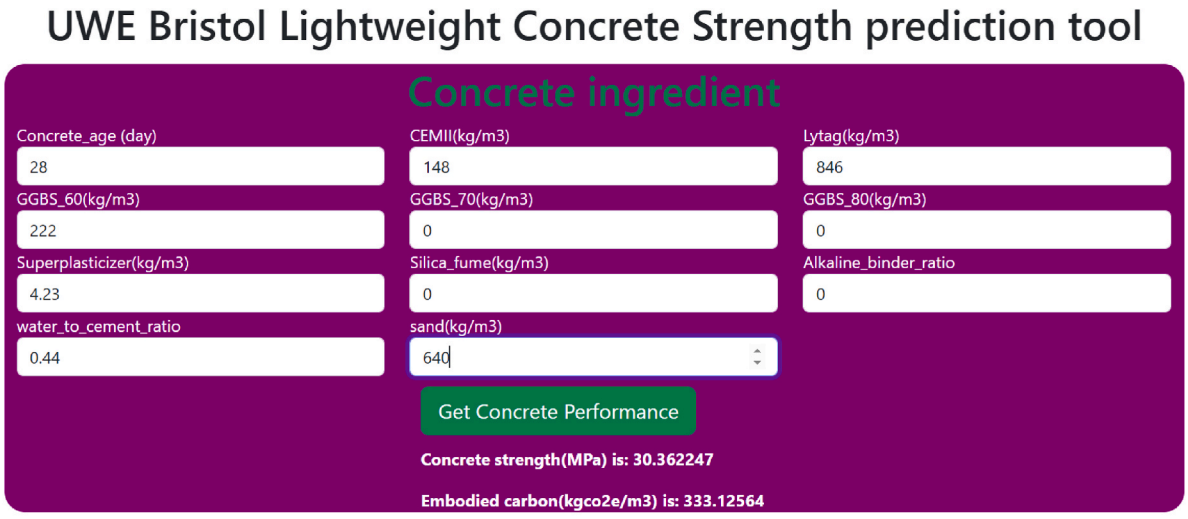


a: SHAP values for the first target (Compressive strength)



b: SHAP values for the second target (Embodied carbon)

Fig. 7. (a–b): Summary plot of SHAP values.



(a): Model input interface

Executive Summary

Lightweight concrete provides a viable solution in the quest for high strength low carbon cost effective concrete owing to reduce selfweight which culminates into a low reinforcement demand.

The addition of supplementary cementitious materials to replace cement activated with alkaline solution, silica fume and superplasticizer enhances workability, durability, low embodied carbon with high structural performance.

This tool provides an attempt to predict embodied carbon as well as compressive strength using recycled sustainable materials by the application of machine learning models.

Data Training

The dataset used in this tool was obtained from rigorous laboratory experiments using Lytag as coarse aggregate with cement replaced between 60-80%.

Modelling

About eight machine learning models were used in the data training after which XGBoost model was optimal having a low RMSE and high R-squared. It was adopted as the preferred model which is used in the development of the tool.

Performance

The tool has the ability to predict concrete embodied carbon and compressive strength of lightweight concrete for age 3 to 28 days.

(b): Readme file

Disclaimer

This tool is strictly for the prediction of embodied carbon and compressive strength of lightweight concrete using the same concrete materials used in the model development and will not be applicable for other concrete materials.

The authors would not be responsible for inaccuracies in results due to the lack of demonstration of knowledge in concrete mix design to acceptable standards and code of practice.

(c): Disclaimer

Fig. 8. (a–c): GUI tool of the proposed super learner model.

predicted strength. Sand has a narrow range of SHAP values, its effect is more predictable compared to Lytag which might have a broader range, indicating more variability. Since CEMII is near the top, it's a crucial predictor with Positive SHAP values for red dots (high values of CEMII) suggesting that higher amounts increase strength predictions. With a consistent color pattern (all red dots on one side) as indicated on CEMII and day, there is a high indication of a reliable effect. The behaviour of embodied carbon as shown Fig. 7(b) is like that of compressive strength except for duration of curing(day) where no remarkable change was noted.

4.3. Prediction of compressive strength and embodied carbon

Prediction of compressive strength and embodied were carried only with XG boost model, based on the performance using multi-output regression from sci-kit learn library in python. Output from the prediction were stored using pickle library for the development of lightweight concrete prediction tool (LCPT) with a general user interface as shown in Fig. 6.

The relationship between embodied carbon and compressive strength was derived using a linear regression model. After making predictions for compressive strength and embodied carbon using the MultiOutputRegressor with XGBRFRegressor, the predicted values were extracted and used to fit a linear regression model.

The linear regression model was trained using the predicted compressive strengths as the independent variable and the predicted embodied carbon values as the dependent variable. By fitting a linear regression model to these predicted values, the model calculates the slope and intercept of the linear relationship between compressive strength and embodied carbon. The linear relationship is shown in equation (17) as

$$Ec = 1.564Cs + 355.47 \quad (17)$$

where Ec is the predicted embodied carbon, $kgco_{2e}/m^3$ and Cs is the predicted compressive strength, MPa. The slope of the model is 1.501 and 357.343 is the intercept.

The equation obtained from the linear regression model represents the estimated linear relationship between compressive strength and embodied carbon. By analysing the slope and intercept of this equation, we can understand how changes in compressive strength are associated with changes in embodied carbon, providing insights into the relationship between these two variables.

The user interface input is shown in Fig. 8 takes input from user and was developed using the algorithm response as demonstrated in Fig. 4 and compared with SHAP analysis in Fig. 7 which shows the effectiveness of the model in predicting compressive strength and embodied carbon of lightweight geopolymer. Predicting compressive strength and embodied carbon during concrete production is a necessity to appraise the need for the expected outcome within the required threshold. Using the predictive developed application is essential for a friendly interaction with the user to produce lightweight concrete with ingredients consistent with that of GGBS, Lytag, CEMII, A/B ratio, w/c ratio, Superplasticiser, Silica fume and sand. The performance of LCPT was applied on similar work of Kanavaris et al. [76], of similar materials with input features comprising of cement, GGBS, Lytag, sand, superplasticizer, water to cement ratio and limestone. However, limestone was added to some concrete mix. The result obtained as presented in Table 4 indicate that LWC predicted is like that from the experiment. With the expectation of yearly design targets for structural embodied carbon as reported by Arnold et al. [77], meeting the 2050 net zero carbon emissions, requires concrete mix optimization. As noted in the report, design expectations of structural embodied carbon emissions, target the average as shown in Table 6 and indicates that the embodied carbon is expected to reduce by 10 % each year. This implies that concrete design with a target embodied carbon as noted in Table 6 and compliment concrete performance in Table 5 which would facilitate the accomplishment of the net zero carbon target. This implies that most typical structural design for concrete must achieve an A rating by 2030 which is equivalent to not more than 150 gt/co2e. When applied to concrete of similar ingredients, the developed tool (UWE LCPT) was able to perform at more than 95 % in predicting compressive strength and embodied carbon of lightweight low carbon concrete as shown in Figs. 9 and 10. The learner model and prediction tool have been developed to combine concrete strength and embodied carbon considerations in response to the evolving building regulations in the UK construction sector [78]. This integration aims to offer a comprehensive approach to design and construction, ensuring a balance between performance, cost-effectiveness, and environmental impact. By providing essential data, this tool empowers engineers, architects, and project managers to make well-informed decisions that meet performance standards while also supporting sustainability objectives related to embodied carbon. The developed tool can be accessed using the link Lightweight Concrete Strength prediction tool (pnukah2006.pythonanywhere.com).

The performance of the data set from Table 2 is as shown in Fig. 11 which suggests that the model has a good level of explanatory power of experimental data compared to the data from other literatures in Table 3. The data from Table 3 indicates a reasonable fit, however there is still substantial room for improvement. From Fig. 12, Linear Regression and Ridge Regression have similar performance with MSEs around 46 and R^2 of 0.56. These models are relatively simple and provide a moderate fit to the data. Lasso Regression

Table 5
Validation of compressive strength and embodied carbon of similar materials from Experiment work in Kanavaris et al. [76],

Concrete mix	Compressive strength, MPa	Embodied carbon, $kgco_{2e}/m^3$
50 % GGBS	38.7	348
50 % GGBS	35.1	328
50 % GGBS	41.4	328
60 % GGBS	41.4	309
50 % GGBS, limestone	41.4	324
50 % GGBS, limestone	45	324
60 % GGBS, limestone	42.3	305

Table 6
Structural embodied carbon, $gtCO_{2e}$.

Global structural carbon budget, $gtCO_{2e}$			
Year	Global average target	40 % worse than average	40 % better than average
2020	350	>400	250
2025	230	325	170
2030	140	100	195
2035	75	55	110
2040	45	40	60
2045	45	40	25
2050	0	0	0

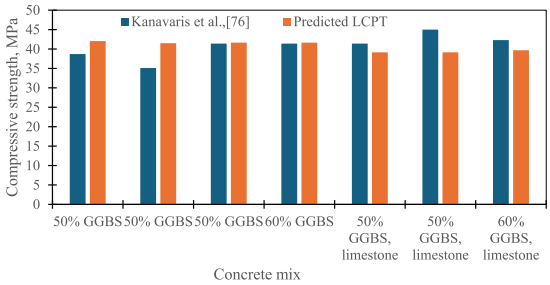


Fig. 9. Compressive strength predictions using developed applications.

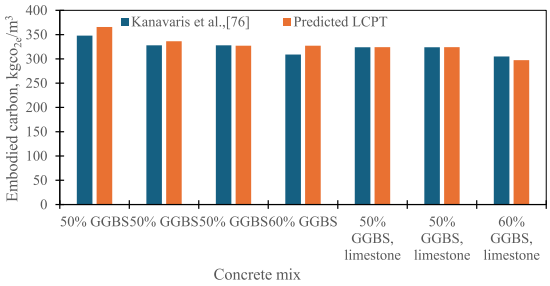
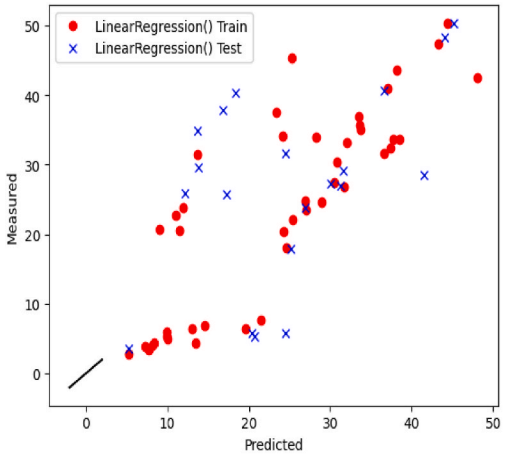


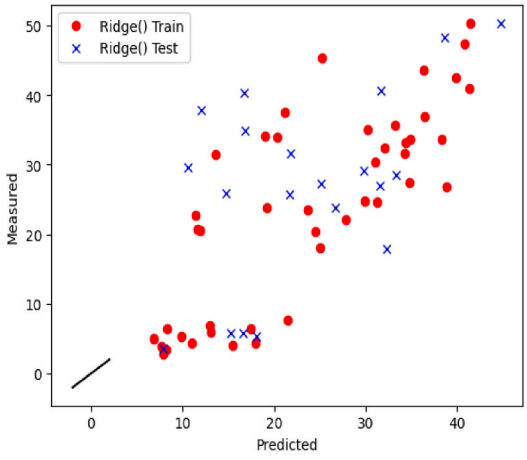
Fig. 10. Embodied carbon predictions using developed tool.

performs better with a lower MSE of 39.67 and a higher R^2 of 0.63, indicating that it can handle feature selection and possibly reduce overfitting better than Linear and Ridge Regression.

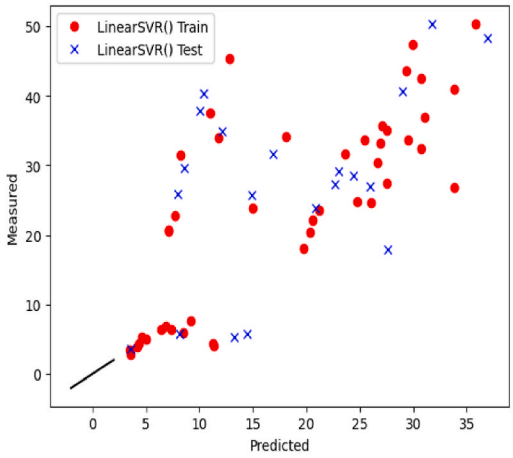
Studies has demonstrated that the embodied carbon (eCO_2) of concrete is not solely determined by its compressive strength but is influenced by various mix design parameters. Initially, it is observed that there is a general trend where eCO_2 increases with concrete strength. This is primarily due to the higher cement content required in high-strength mixes to maintain workability and compaction. The higher the compressive strength of concrete, the more cement is typically needed, leading to increased carbon emissions associated with cement production. Contrary to a linear relationship between eCO_2 and compressive strength, it was revealed that there is an optimum strength range for concrete to minimize eCO_2 per unit of structural performance, typically between 50 and 70 MPa [79]. The study demonstrates through empirical data of 512 theoretical concrete mixes and model analysis that there is a general trend where eCO_2 increases with concrete compressive strength. The use of supplementary cementitious materials like pulverised fly ash (PFA) can significantly reduce carbon emissions in concrete production. By replacing a portion of cement with PFA, substantial eCO_2 savings can be achieved without compromising performance. Factors such as water-to-binder ratio, aggregate type (crushed vs. uncrushed), use of superplasticizer, and cement strength class can all impact eCO_2 levels in concrete. Adjusting these variables intelligently can lead to notable reductions in carbon emissions. Different SCMs, such as fly ash, ground granulated blast furnace slag (GGBS), natural pozzolans, and limestone filler, have varying effects on green house gas (GHG) emissions reduction. The specified compressive strength of the concrete mixture influences its environmental sustainability, however higher compressive strengths may result in higher GHG



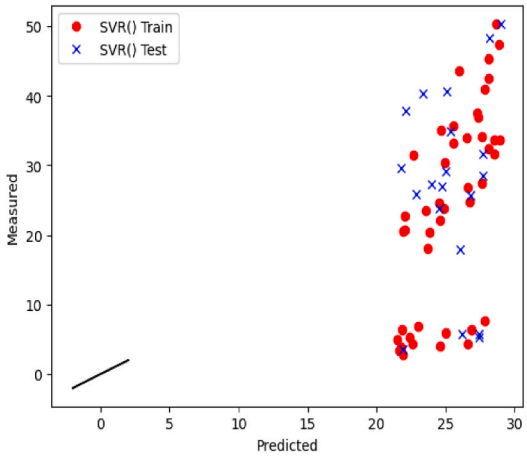
(a) Scatter plot of linear regression



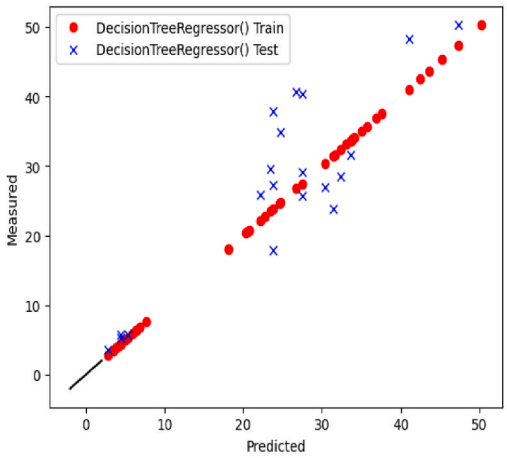
(b) Scatter plot of Ridge regression



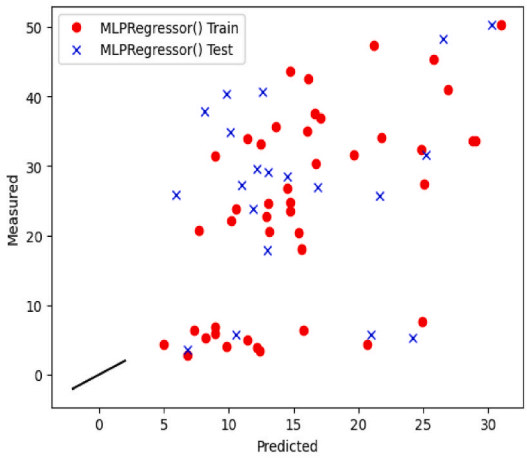
(c) Scatter plot of linear SVR regression



(d) Scatter plot of linear SVR regression

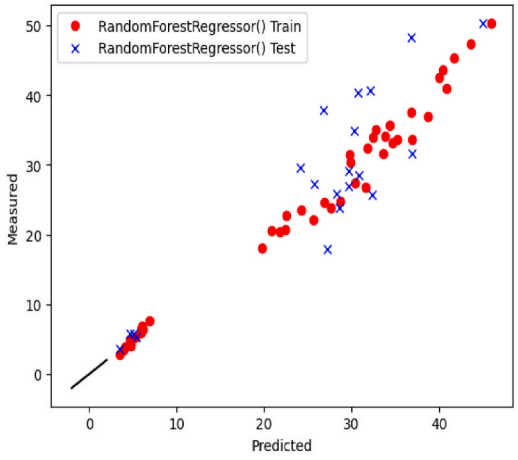


(e) Scatter plot of Decision regression

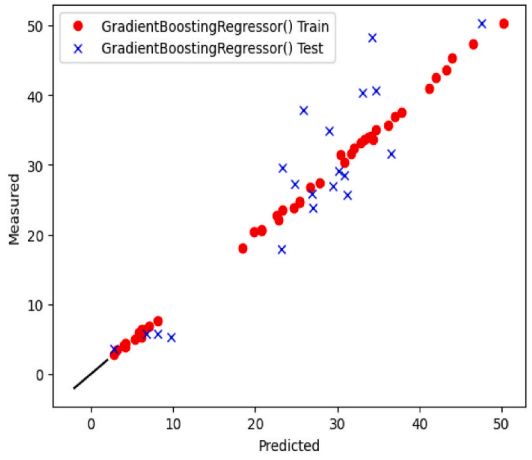


(f) Scatter plot of MLP regression

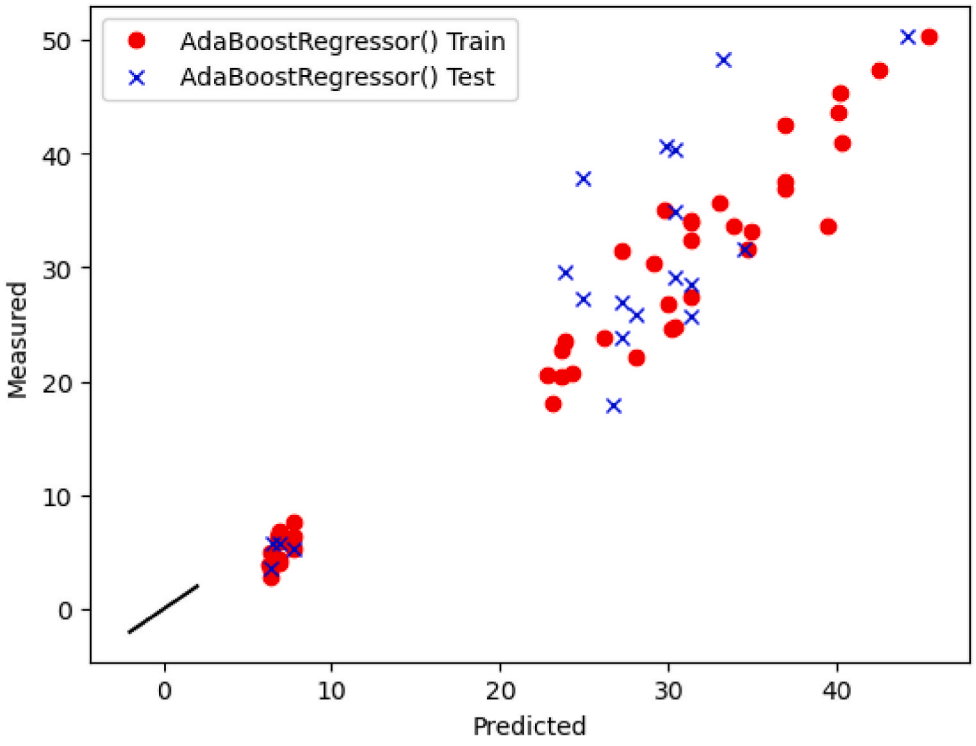
Fig. 11. (a–j): Scatter plot of model for training and testing from Table 2.



(g) Scatter plot of Random forest regression



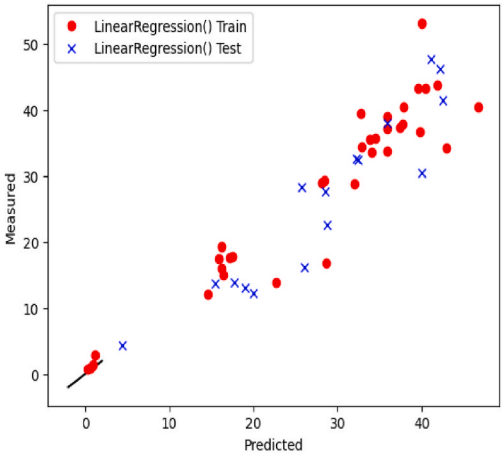
(h) Scatter plot of Gradient boost regression



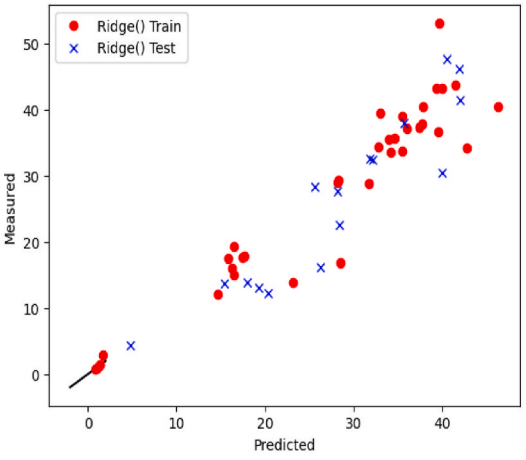
(j): Scatter plot of AdaBoost boost regression.

Fig. 11. (*continued*).

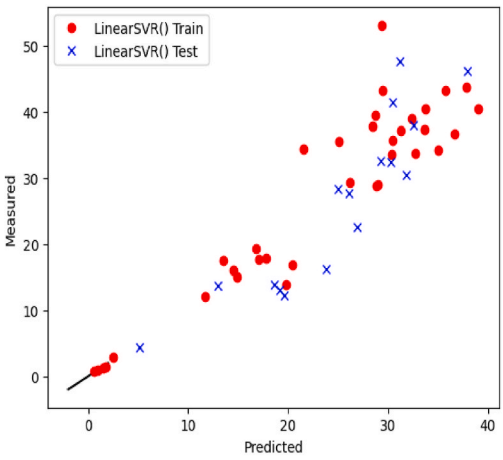
emissions, so it is essential to balance strength requirements with emission reduction goals [80]. From Table 7, it is suggested that by optimizing the mix design and incorporating supplementary cementitious materials like GGBS, it is possible to achieve the desired compressive strength levels while reducing the embodied carbon footprint of concrete. It is also shown that the use of a superplasticizer in concrete mixtures can reduce the overall embodied carbon by approximately 26 % while achieving the desired workability with reduced water content, leading to a corresponding reduction in binder content. Higher compressive strength requirements in concrete mixtures can lead to increased carbon emissions. It is crucial to consider the trade-off between achieving higher strength and minimizing embodied carbon to meet sustainability goals. The choice of cement replacement materials, such as GGBS and PFA, can



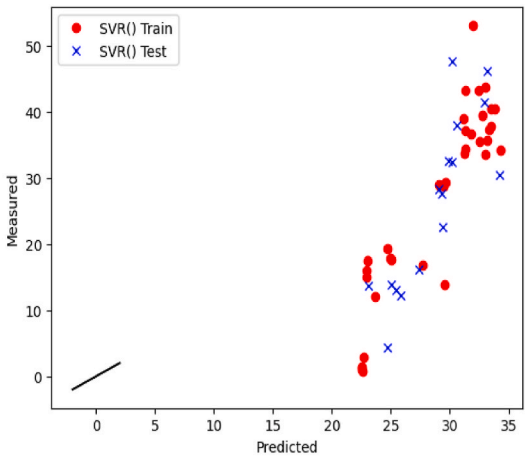
(a) Scatter plot of Linear regression



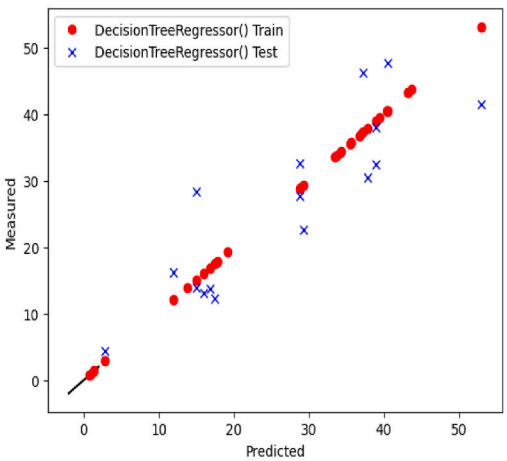
(b) Scatter plot of Linear regression



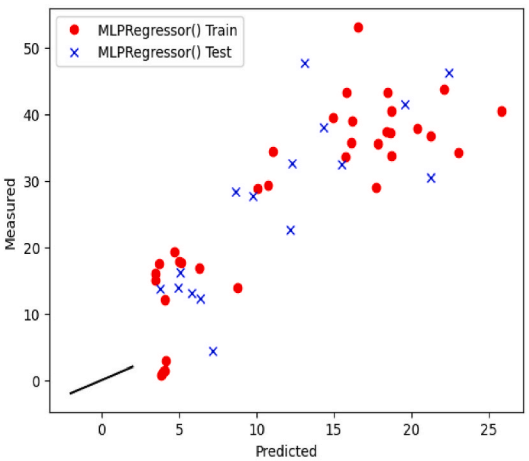
(c)Scatter plot of Linear SVR regression



(d) Scatter plot of Linear SVR regression

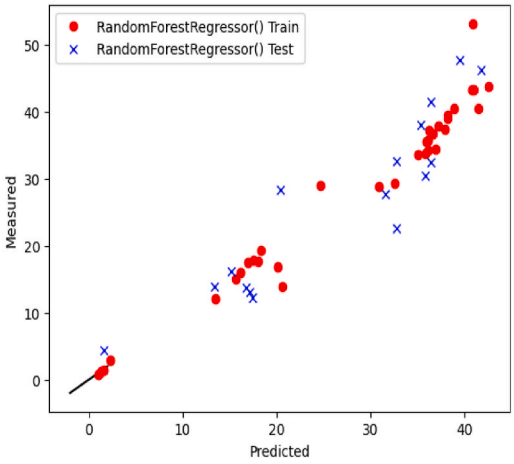


(e) Scatter plot of Decision tree regression

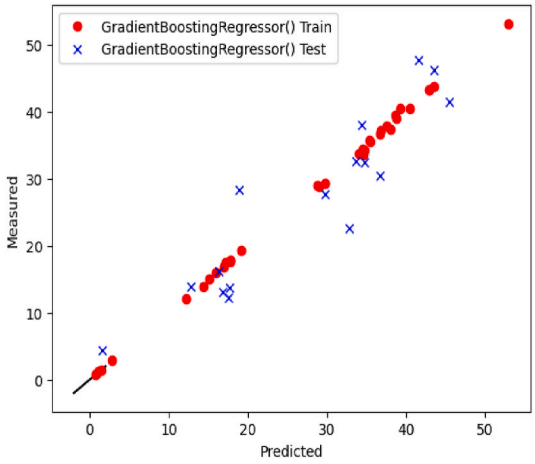


(f) Scatter plot of Decision of tree regression

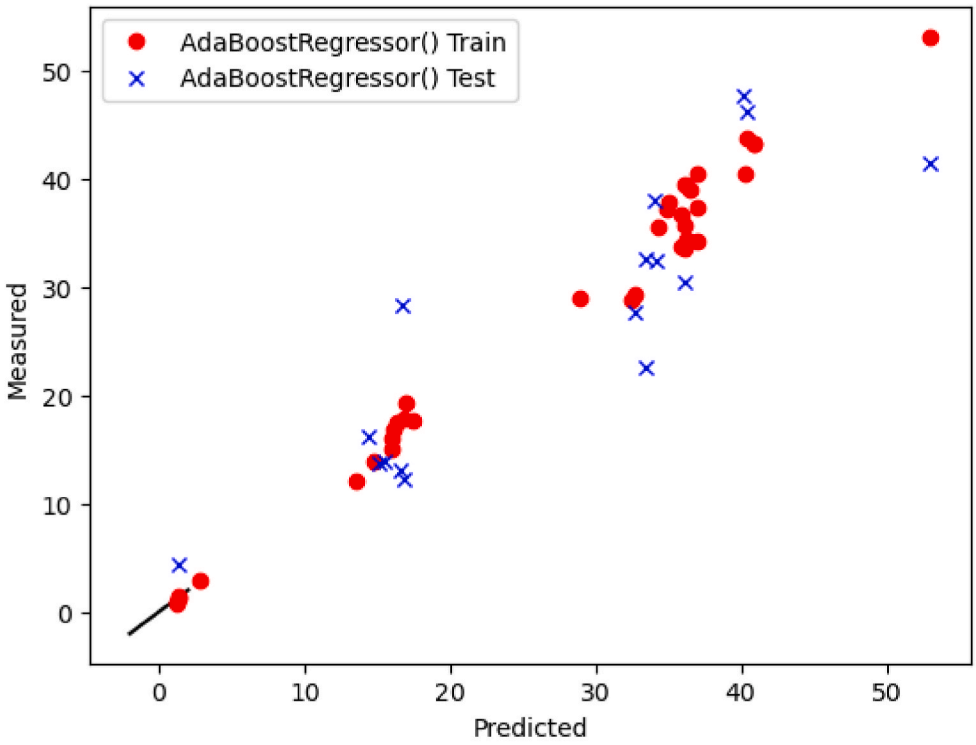
Fig. 12. (a–i): Scatter plot of model for training and testing from other literature in Table 3.



(g)Scatter plot of Random Forest regression



(h) Scatter plot of Random Forest regression



(i): Scatter plot of AdaBoost regression

Fig. 12. (*continued*).

significantly impact the carbon emissions of concrete mixtures. GGBS-OPC mixtures reached an optimal point of low GHG emissions compared to PFA-OPC mixtures under certain conditions.

5. Conclusion and future directions

The essence of computer model development is to enhance accessibility and application by the user and stakeholder. The growing demand for a sustainable lightweight low carbon concrete presents the challenge of mix optimization and have resulted to wastage and economic loss due to trial mixes not meeting the target design strength as well as embodied carbon. Considering the importance of effective dataset in model prediction, this study uses dataset from experimental work of lightweight using ground granular blast

Table 7
Influence of Compressive strength on embodied carbon for low carbon concrete.

Number of concrete mix in dataset	Concrete ingredients	Affiliation/location of study	Compressive strength vs embodied carbon	Suggestion for trends	Literature
512	CEM I Cement binder, Pulverised Fuel Ash (PFA) as a cement replacement material, Super-plasticiser as a liquid additive, Aggregate (uncrushed or crushed), Water (implicit in the water: binder ratio)	Institute for Resilient Infrastructure, School of Civil Engineering, University of Leeds, Leeds, LS2 9JT, UK	Optimal concrete strength range of 50–70 MPa for minimizing eCO2 per unit of structural performance	The use of a superplasticiser can reduce overall eCO2 by approximately 26 % by achieving the desired workability with reduced water content, leading to a corresponding reduction in binder content	(Purnell and Black [81],
105	Ordinary Portland Cement (OPC), natural pozzolans (NP), limestone, ground granulated blast furnace slag (GGBS), and Class F fly ash (PFA) as cement replacement	Department of Civil and Environmental Engineering at the University of California, Davis	higher compressive strength requirements in concrete mixtures can lead to increased carbon emissions.	GGBS-OPC mixtures reached an optimal point of low GHG emissions when the s/c ratio increased from 0.5 to 1.57 compared to PFA-OPC.	(Fan and Miller [82],
22	Cement, Water, Recycled aggregates, Carbonation pressure, Carbonation duration	School of Engineering, Design and Built Environment, Western Sydney University, Australia, and the Engineering Institute of Technology, Perth, Australia	Compressive strength increases with carbonation pressure	Injection of CO2 into recycled aggregate concrete led to an increase in compressive strength compared to samples without carbonation	(Tam et al. [83],
12	Portland cement with a density of 3.14 g/cm3, PFA with a density of 2.11 g/cm3, Fine aggregate (sea sand and recycled sand), Micro-steel fiber (MSF) as reinforcing fiber, Polycarboxylic acid-based water reducing agent as an admixture	Korea Institute of Civil Engineering and Building Technology (KICT) in South Korea	Compressive strength increases with embodied carbon	The relationship between compressive strength and carbon emissions in fiber-reinforced cementitious composites (FRCCs) is strongly informed by the binder index (Bi) and CO_2 index (Ci).	(Lee et al. [84],
210 datasets	Temperature, water-to-binder ratio (w/b), GGBS-to-binder ratio (GGBFS/b), fine aggregate (FA), coarse aggregate (CA), and superplasticizer content (SP)		Compressive strength increases with embodied carbon	By optimizing the mix design, incorporating supplementary cementitious materials like GGBFS, and using alternative binders, it is possible to achieve desired compressive strength levels while reducing the embodied carbon footprint of concrete	(Ahmad Khalil Mohammed, Hassan and Ahmed Salih Mohammed [85],
64	Concrete age, OPC, Lytag, sand, 60 % GGBS, 70 % GGBS, 80 % GGBS, superplasticizer, silica fume, alkaline binder ratio, and water to cement ratio	University of the West of England, Bristol, UK	Compressive strength increases with embodied carbon	Addition of silica fume and superplasticizer to the geopolymer concrete enhance good concrete performance at low embodied carbon	Present study

furnace slag (GGBS) at 60–80 % cement replacement activated with alkaline precursor and lytag as coarse aggregate. Machine learning techniques was developed using multi output regression for prediction of compressive strength and embodied carbon of lightweight low carbon concrete. Several machine learning models were applied on training data set for choice of optimal model. A general user interface was developed to aid the application of the model to enhance waste minimization and cost effective in lightweight concrete production. Based on the results obtained in terms of model performance and the output interface developed in the study on lightweight low carbon concrete prediction using machine learning, the following conclusions can be drawn:

1. Model Performance:

XGBRF Regressor outperformed the other models with the lowest RMSE of 7.08, MAE of 5.26, MAPE of 11.76, and highest % Explained Variance of 0.97 and R-squared values of 0.96. The Neural Network had the highest errors and lowest % Explained Variance and R-squared values among the models compared. This indicates superior predictive accuracy compared to other regression models. Decision Tree Regression also showed promising performance with a low MSE and relatively high R-squared value, suggesting it effectively captured the variability in the data. Support Vector Machine (SVM) with a linear kernel exhibited unusual results with a negative R-squared value, indicating potential issues in fitting the data well. Neural Network model demonstrated poor performance with high MSE and negative R-squared values, suggesting limitations in its suitability.

The intuitive and user-friendly output interface for interpreting the model predictions. This interface facilitated easier understanding and utilization of the prediction results by stakeholders in the construction industry. The output interface included user-friendly displays of prediction outcomes for both compressive strength and embodied carbon which enhances the usability and accessibility of the concrete sustainability as well as structural integrity. By developing a user-friendly output interface, we were able to bridge the gap between complex machine learning algorithms and practical applications in the construction sector, enabling stakeholders to make informed decisions based on the model predictions. The developed lightweight prediction tool when compared with concrete of similar mix ingredients performed more than 95 % in predicting compressive strength and embodied carbon.

2. Embodied carbon vs compressive strength:

The importance of considering the mix design, choice of materials, and construction practices to optimize concrete mixtures for both compressive strength requirements and reduced embodied carbon is strongly suggested.

3. Future direction

Further studies should incorporate other advance learning algorithms by exploring the integration of advanced machine learning algorithms, such as Deep Learning and Reinforcement Learning, to enhance the accuracy and efficiency of carbon footprint predictions in construction projects. These advanced algorithms have the potential to uncover complex patterns in large datasets and improve the predictive capabilities of models over time. Also, the inclusion of long-term prediction models should be considered by developing predictive models that can forecast carbon emissions over the entire lifecycle of construction projects, including the operational phase and end-of-life considerations. By incorporating temporal aspects and considering the dynamic nature of carbon emissions, these long-term prediction models can provide valuable insights for sustainable decision-making and resource optimization.

Author agreement statement

We the undersigned declare that this manuscript is original, has not been published before and is not currently being considered for publication elsewhere. We confirm that the manuscript has been read and approved by all named authors and that there are no other persons who satisfied the criteria for authorship but are not listed. We further confirm that the order of authors listed in the manuscript has been approved by all of us. We understand that the Corresponding Author is the sole contact for the Editorial process. He is responsible for communicating with the other authors about progress, submissions of revisions and final approval of proofs.

CRediT authorship contribution statement

Promise D. Nukah: Writing – original draft, Software, Resources, Methodology, Conceptualization. **Samuel J. Abbey:** Supervision, Funding acquisition. **Colin A. Booth:** Validation, Project administration.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgement

The authors would like to specially thank the University of the West of England, Bristol for funding this PhD Research project under the grant number 7911422.

References

[1] M. Lu, J. Lai, Review on carbon emissions of commercial buildings, Renew. Sustain. Energy Rev. 119 (2020) 109545, https://doi.org/10.1016/j.rser.2019.109545.

[2] B. Sizirici, Y. Fseha, C.-S. Cho, I. Yildiz, Y.-J. Byon, A review of carbon footprint reduction in construction industry, from design to operation, Materials 14 (20) (2021) 6094, https://doi.org/10.3390/ma14206094 [online].

[3] M. Adams, V. Burrows, S. Richardson, J. Drinkwater, C. Gamboa, Bringing Embodied Carbon Upfront, World Green Building Council, London, 2019.

[4] N. Mahasenan, S. Smith, K. Humphreys, Greenhouse Gas Control Technologies-6th International Conference. *Proceeding*, Elsevier, Kyoto, Japan, 2003, p. 50157, 4.

[5] Promise D. Nukah, et al., Mapping and synthesizing the viability of cement replacement materials via a systematic review and meta-analysis, Construct. Build. Mater. 405 (1 Nov) (2023) 133290, https://doi.org/10.1016/j.conbuildmat.2023.133290, 133290.

[6] Demi Fang, et al., Reducing embodied carbon in structural systems: a review of early-stage design strategies, J. Build. Eng. 76 (1 Oct) (2023) 107054, https://doi.org/10.1016/j.jobe.2023.107054. www.sciencedirect.com/science/article/abs/pii/S2352710223012330.

[7] Patryk Ziolkowski, Computational complexity and its influence on predictive capabilities of machine learning models for concrete mix design, Materials 16 (17) (2023) 5956, https://doi.org/10.3390/ma16175956. www.mdpi.com/1996-1944/16/17/5956. (Accessed 21 December 2023).

[8] Omar Y. Al-Jarrah, et al., Efficient machine learning for big data: a review, Big Data Research 2 (3) (Sept. 2015) 87–93, https://doi.org/10.1016/j.bdr.2015.04.001.

[9] Flavio Barboza, et al., Machine learning models and bankruptcy prediction, Expert Syst. Appl. 83 (Oct. 2017) 405–417, https://doi.org/10.1016/j.eswa.2017.04.006.

[10] Amandalynne Paullada, et al., Data and its (Dis)Contents: a survey of dataset development and use in machine learning Research, Patterns 2 (11) (Nov. 2021) 100336, https://doi.org/10.1016/j.patter.2021.100336.

[11] M.A. DeRousseau, E. Laftchiev, J.R. Kasprzyk, B. Rajagopalan, W.V. Srubar, A comparison of machine learning methods for predicting the compressive strength of field-placed concrete, Construct. Build. Mater. 228 (2019) 116661, https://doi.org/10.1016/j.conbuildmat.2019.08.042.

[12] P.S.M. Thilakarathna, S. Seo, K.S.K. Baduge, H. Lee, P. Mendis, G. Foliente, Embodied carbon analysis and benchmarking emissions of high and ultra-high strength concrete using machine learning algorithms, J. Clean. Prod. 262 (2020) 121281.

[13] Hamed Naseri, et al., Designing sustainable concrete mixture by developing a new machine learning technique, J. Clean. Prod. 258 (June 2020) 120578, https://doi.org/10.1016/j.jclepro.2020.120578.

[14] D'Amico, V. Anthony, et al., Artificial Neural Networks to Assess Energy and Environmental Performance of Buildings: an Italian Case Study, vol. 239, 2019, p. 117993, https://doi.org/10.1016/j.jclepro.2019.117993, 117993.

[15] Promise D. Nukah, et al., Optimisation of embodied carbon and compressive strength in low carbon concrete, Materials 15 (23) (2022) 8673, https://doi.org/10.3390/ma15238673, 5 Dec.

[16] W. Luo, M. Sandanayake, G. Zhang, Direct and indirect carbon emissions in foundation construction – two case studies of driven precast and cast-in-situ piles, J. Clean. Prod. 211 (2019) 1517–1526, https://doi.org/10.1016/j.jclepro.2018.11.244.

[17] Y. Chen, J. Zeng, J. Jia, Jabli Mahjoub, N. Abdullah, Samia Elattar, Amine Kadimallah Mohamed, Marzouki Riadh, A. Hashm, Assilzadeh Hamid, A fusion of neural, genetic and ensemble machine learning approaches for enhancing the engineering predictive capabilities of lightweight foamed reinforced concrete beam, Powder Technol. 440 (2024) 119680, https://doi.org/10.1016/j.powtec.2024.119680, 119680.

[18] F. Pomponi, M.L. Anguita, M. Lange, B. D'Amico, E. Hart, Enhancing the practicality of tools to estimate the whole life embodied carbon of building structures via machine learning models, Front. Built Environ. 7 (2021), https://doi.org/10.3389/fbuil.2021.745598.

[19] Prakhash Neelamegam, Bhuvaneshwari Muthusubramanian, Evaluating embodied energy, carbon impact, and predictive precision through machine learning for pavers manufactured with treated recycled construction and demolition waste aggregate, Environ. Res. 248 (2024) 118296, https://doi.org/10.1016/j.envres.2024.118296, 118296.

[20] Y. Fang, X. Lu, H. Li, A random forest-based model for the prediction of construction-stage carbon emissions at the early design stage, J. Clean. Prod. 328 (2021) 129657, https://doi.org/10.1016/j.jclepro.2021.129657.

[21] Umehara Mitsutaro, H.S. Stein, D. Guevarra, P.A. Newhouse, D.A. Boyd, J.M. Gregoire, Analyzing machine learning models to accelerate generation of fundamental materials insights, npj Comput. Mater. 5 (1) (2019), https://doi.org/10.1038/s41524-019-0172-5.

[22] W. Zhu, W. Feng, X. Li, Z. Zhang, Analysis of the embodied carbon dioxide in the building sector: a case of China, J. Clean. Prod. 269 (2020) 122438, https://doi.org/10.1016/j.jclepro.2020.122438.

[23] L. Farahzadi, M. Kioumarsi, Application of machine learning initiatives and intelligent perspectives for CO2 emissions reduction in construction, J. Clean. Prod. (2022) 135504, https://doi.org/10.1016/j.jclepro.2022.135504 [online].

[24] I.H. Sarker, AI-based modeling: techniques, applications and Research issues towards automation, intelligent and smart systems, SN Computer Science 3 (2) (2022), https://doi.org/10.1007/s42979-022-01043-x.

[25] J. Rosłon, Materials and technology selection for construction projects supported with the use of artificial intelligence, Materials 15 (4) (2022) 1282, https://doi.org/10.3390/ma15041282.

[26] S. Ronghui, N. Liangrong, An intelligent fuzzy-based hybrid metaheuristic algorithm for analysis the strength, energy and cost optimization of building material in construction management, Eng. Comput. (2021), https://doi.org/10.1007/s00366-021-01420-9.

[27] T. Proske, S. Hainer, M. Rezvani, C.-A. Graubner, Eco-friendly concretes with reduced water and cement contents — mix design principles and laboratory tests, Cement Concr. Res. 51 (2013) 38–46, https://doi.org/10.1016/j.cemconres.2013.04.011.

[28] I. Nunez, M.L. Nehdi, Machine learning prediction of carbonation depth in recycled aggregate concrete incorporating SCMs, Construct. Build. Mater. 287 (2021) 123027, https://doi.org/10.1016/j.conbuildmat.2021.123027.

[29] R.M. Khalifa, S. Yacout, S. Bassetto, Developing machine-learning regression model with logical analysis of data (LAD), Comput. Ind. Eng. 151 (2021) 106947, https://doi.org/10.1016/j.cie.2020.106947 [online].

[30] X. Guo, X. Gui, H. Xiong, X. Hu, Y. Li, H. Cui, Y. Qiu, C. Ma, Critical role of climate factors for groundwater potential mapping in arid regions: insights from random forest, XGBoost, and LightGBM algorithms, J. Hydrol. 621 (2023) 129599, https://doi.org/10.1016/j.jhydrol.2023.129599, 129599.

[31] A. Mukherjee, S. Nag Biswas, Artificial neural networks in prediction of mechanical behavior of concrete at high temperature, Nucl. Eng. Des. 178 (1) (1997) 1–11, https://doi.org/10.1016/s0029-5493(97)00152-0.

[32] H. Ling, C. Qian, W. Kang, C. Liang, H. Chen, Combination of Support Vector Machine and K-Fold cross validation to predict compressive strength of concrete in marine environment, Construct. Build. Mater. 206 (2019) 355–363, https://doi.org/10.1016/j.conbuildmat.2019.02.071.

[33] H.-B. Ly, T.-A. Nguyen, H.-V. Thi Mai, V.Q. Tran, Development of deep neural network model to predict the compressive strength of rubber concrete, Construct. Build. Mater. 301 (2021) 124081, https://doi.org/10.1016/j.conbuildmat.2021.124081.

[34] S. Jueyendah, M. Lezgy-Nazargah, H. Eskandari-Naddaf, S.A. Emamian, Predicting the mechanical properties of cement mortar using the support vector machine approach, Construct. Build. Mater. 291 (2021) 123396, https://doi.org/10.1016/j.conbuildmat.2021.123396.

[35] J. Liu, K. Yan, X. Zhao, Y. Hu, Prediction of autogenous shrinkage of concretes by support vector machine, Int. J. Pavement Res. Tech. 9 (3) (2016) 169–177, https://doi.org/10.1016/j.ijprt.2016.06.003.

[36] M.R. Kaloop, D. Kumar, P. Samui, J.W. Hu, D. Kim, Compressive strength prediction of high-performance concrete using gradient tree boosting machine, Construct. Build. Mater. 264 (2020) 120198, https://doi.org/10.1016/j.conbuildmat.2020.120198.

[37] Abisek Rajakarunakaran Surya, Raja Lourdu Arun, S. Muthusamy, H. Panchal, A. Jawad, Musa Jaber Mustafa, M.K. Ali, Iskander Tlili, Andino Maseleno, A. Majdi, Ali, Prediction of strength and analysis in self-compacting concrete using machine learning based regression techniques, Adv. Eng. Software 173 (2022) 103267, https://doi.org/10.1016/j.advengsoft.2022.103267, 103267.

[38] M.A. Shafiq, Predicting the compressive strength of concrete using neural network and kernel ridge regression. https://doi.org/10.1109/ftc.2016.7821698, 2016.
[39] P.D. Nukah, S.J. Abbey, C.A. Booth, J. Oti, Development of low carbon concrete and prospective of geopolymer concrete using lightweight coarse aggregate and cement replacement materials, Construct. Build. Mater. 428 (2024) 136295, https://doi.org/10.1016/j.conbuildmat.2024.136295, 136295.
[40] BS EN 12390-3, Testing Hardened Concrete—Part 3: Compressive Strength of Test Specimens, Management Centre, Brussels, 2009.
[41] BS EN15978, Sustainability of Construction Works-Assessment of Environmental Performance of Building-Calculation Method, British Standard Institution, London, 2011, 2011.
[42] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, Scikit-learn: machine learning in Python, J. Mach. Learn. Res. 12 (2011) 2825–2830.
[43] A. Hafeez, A.H. Sial, Comparative analysis of data visualization libraries matplotlib and seaborn in Python, Int. J. Adv. Trends Comput. Sci. Eng. 10 (1) (2021) 277–281, https://doi.org/10.30534/ijatcse/2021/391012021.
[44] E. Bressert, SciPy and NumPy: an Overview for Developers, 2012.
[45] C.-C. Tsai, T.-H. Yiu, Investigation of laser ablation quality based on data science and machine learning XGBoost classifier, Appl. Sci. 14 (1) (2023) 326, https://doi.org/10.3390/app14010326, 326.
[46] D.Y. Chen, Pandas for Everyone: Python Data Analysis, Addison-Wesley Professional, 2017.
[47] I. Idris, Python Data Analysis Cookbook, Packt Publishing Ltd, 2016.
[48] J. Faouzi, H. Janati, pyts: a python package for time series classification, J. Mach. Learn. Res. 21 (46) (2020) 1–6.
[49] K. Khawar, S. Munawar, N. Naveed, Fuzzy logic-based expert system for assessing programming course performance of E-learning students, J. Inform. Commun. Tech. Robotic Appl. (2020) 54–64.
[50] U. Patkar, P. Singh, H. Panse, S. Bhavsar, C. Pandey, Python for web development, Int. J. Comput. Sci. Mobile Comput. 11 (4) (2022) 36.
[51] J. Gopalakrishnan, Lectures on Mathematical Computing with Python, 2020.
[52] M. Feurer, F. Hutter, Hyperparameter Optimization. *Automated Machine Learning*, 2019, pp. 3–33. *Methods, systems, challenges*.
[53] M. Azimi-Pour, H. Eskandari-Naddaf, A. Pakzad, Linear and non-linear SVM prediction for fresh properties and compressive strength of high volume fly ash self-compacting concrete, Construct. Build. Mater. 230 (2020) 117021, https://doi.org/10.1016/j.conbuildmat.2019.117021 [online].
[54] Y. Freund, R.E. Schapire, Experiments with a new boosting algorithm, icml 96 (1996, July) 148–156.
[55] Y. Freund, R.E. Schapire, A decision-theoretic generalization of on-line learning and an application to boosting, J. Comput. Syst. Sci. 55 (1) (1997) 119–139, https://doi.org/10.1006/jcss.1997.1504 [online].
[56] D.P. Solomatine, D.L. Shrestha, AdaBoost. RT: a boosting algorithm for regression problems, July, in: 2004 IEEE International Joint Conference on Neural Networks (IEEE Cat. No. 04CH37541), vol. 2IEEE, 2004, pp. 1163–1168.
[57] M. Czajkowski, M. Kretowski, The role of decision tree representation in regression problems – an evolutionary perspective, Appl. Soft Comput. 48 (2016) 458–475, https://doi.org/10.1016/j.asoc.2016.07.007.
[58] A.M. Ahmed, A. Rizaner, A.H. Ulusoy, A decision tree algorithm combined with linear regression for data classification, in: 2018 International Conference on Computer, Control, Electrical, and Electronics Engineering (ICCCEEE), IEEE, 2018, August, pp. 1–5.
[59] Y. Li, H. Zeng, M. Zhang, B. Wu, Y. Zhao, X. Yao, T. Cheng, X. Qin, F. Wu, A county-level soybean yield prediction framework coupled with XGBoost and multidimensional feature engineering, Int. J. Appl. Earth Obs. Geoinf. 118 (2023) 103269, https://doi.org/10.1016/j.jag.2023.103269, 103269.
[60] H. Mo, H. Sun, J. Liu, S. Wei, Developing window behavior models for residential buildings using XGBoost algorithm, Energy Build. 205 (2019) 109564, https://doi.org/10.1016/j.enbuild.2019.109564.
[61] D.F. Specht, The general regression neural network—rediscovered, Neural Network. 6 (7) (1993) 1033–1034, https://doi.org/10.1016/s0893-6080(09)80013-0.
[62] D.F. Specht, A general regression neural network, IEEE Trans. Neural Network. 2 (6) (1991) 568–576.
[63] H.K. Cigizoglu, M. Alp, Generalized regression neural network in modelling river sediment yield, Adv. Eng. Software 37 (2) (2006) 63–68, https://doi.org/10.1016/j.advengsoft.2005.05.002.
[64] D.M. Hawkins, X. Yin, A faster algorithm for ridge regression of reduced rank data, Computat. Statist. Data Analy. 40 (2) (2002) 253–262, https://doi.org/10.1016/s0167-9473(02)00034-8.
[65] C. Saunders, A. Gammerman, V. Vovk, Ridge regression learning algorithm in dual variables, in: 15th International Conference on Machine Learning, ICML '98, Morgan Kaufmann, 1998, pp. 515–521.
[66] R. Alhamzawi, H.T.M. Ali, The Bayesian adaptive lasso regression, Math. Biosci. 303 (2018) 75–82.
[67] Ergül Yaşar, Payam Shafigh, Alaettin Kılıç, Hasan Gulsen, Strength properties of lightweight concrete made with basaltic pumice and fly ash, Mater. Lett. 57 (15) (2003) 2267–2270, https://doi.org/10.1016/s0167-577x(03)00146-0.
[68] L. Domagała, Size effect in compressive strength tests of cored specimens of lightweight aggregate concrete, Materials 13 (5) (2020) 1187, https://doi.org/10.3390/ma13051187.
[69] M.J. Shannag, Characteristics of lightweight concrete containing mineral admixtures, Construct. Build. Mater. 25 (2) (2011) 658–662, https://doi.org/10.1016/j.conbuildmat.2010.07.025.
[70] P. Sikora, T. Rucinska, D. Stephan, S.-Y. Chung, M. Abd Elrahman, Evaluating the effects of nanosilica on the material properties of lightweight and ultra-lightweight concrete using image-based approaches, Construct. Build. Mater. 264 (2020) 120241, https://doi.org/10.1016/j.conbuildmat.2020.120241.
[71] J.-I. Sim, K.-H. Yang, H.-Y. Kim, B.-J. Choi, Size and shape effects on compressive strength of lightweight concrete, Construct. Build. Mater. 38 (2013) 854–864, https://doi.org/10.1016/j.conbuildmat.2012.09.073.
[72] P. Shafigh, M.Z. Jumaat, H.B. Mahmud, N.A.A. Hamid, Lightweight concrete made from crushed oil palm shell: tensile strength and effect of initial curing on compressive strength, Construct. Build. Mater. 27 (1) (2012) 252–258, https://doi.org/10.1016/j.conbuildmat.2011.07.051.
[73] K. Sahoo, A.K. Samal, J. Pramanik, S.K. Pani, Exploratory data analysis using Python, Int. J. Innovative Technol. Explor. Eng. 8 (12) (2019) 4727–4735, https://doi.org/10.35940/ijitee.l3591.1081219.
[74] N.H. Augustin, E.-A. Sauleau, S.N. Wood, On quantile quantile plots for generalized linear models, Comput. Stat. Data Anal. 56 (8) (2012) 2404–2409, https://doi.org/10.1016/j.csda.2012.01.026.
[75] H. Nguyen, T. Vu, T.P. Vo, H.-T. Thai, Efficient machine learning models for prediction of concrete strengths, Construct. Build. Mater. 266 (2021) 120950, https://doi.org/10.1016/j.conbuildmat.2020.120950.
[76] F. Kanavaris, O. Gibbons, E. Walport, E. Shearer, A. Abbas, J. Orr, B. Marsh, Reducing the carbon footprint of lightweight aggregate concrete, in: LowC3 2020, 2020.
[77] W. Arnold, M. Cook, D. Cox, O. Gibbons, J. Orr, Setting carbon targets: an introduction to the proposed SCORS rating scheme, Struct. Eng. 98 (10) (2020) 8–12, https://doi.org/10.56330/sqdi8782.
[78] R. Jones, The institution of structural engineers, The institution of structural engineers 102 (5) (2024). May.
[79] X. Zhu, Y. Zhang, Z. Liu, H. Qiao, F. Ye, Z. Lei, Research on carbon emission reduction of manufactured sand concrete based on compressive strength, Construct. Build. Mater. 403 (2023) 133101, https://doi.org/10.1016/j.conbuildmat.2023.133101 [online].
[80] P.S.M. Thilakarathna, S. Seo, K.S.K. Baduge, H. Lee, P. Mendis, G. Foliente, Embodied carbon analysis and benchmarking emissions of high and ultra-high strength concrete using machine learning algorithms, J. Clean. Prod. 262 (2020) 121281, https://doi.org/10.1016/j.jclepro.2020.121281.
[81] P. Purnell, L. Black, Embodied carbon dioxide in concrete: variation with common mix design parameters, Cement Concr. Res. 42 (6) (2012) 874–877, https://doi.org/10.1016/j.cemconres.2012.02.005 [online].
[82] C. Fan, S.A. Miller, Reducing greenhouse gas emissions for prescribed concrete compressive strength, Construct. Build. Mater. 167 (2018) 918–928, https://doi.org/10.1016/j.conbuildmat.2018.02.092.

[83] V.W.Y. Tam, A. Butera, K.N. Le, L.C.F.D. Silva, A.C.J. Evangelista, A prediction model for compressive strength of CO2 concrete using regression analysis and artificial neural networks, Construct. Build. Mater. 324 (2022) 126689, https://doi.org/10.1016/j.conbuildmat.2022.126689 [online].
[84] J.-W. Lee, Y.-I. Jang, W.-S. Park, H.-D. Yun, The effect of fly ash and recycled aggregate on the strength and carbon emission impact of FRCCs, Int. J. Concrete Struct. Mater. 14 (1) (2020), https://doi.org/10.1186/s40069-020-0392-6.
[85] Khalil Mohammed Ahmad, Hassan, Salih Mohammed Ahmed, Predicting the compressive strength of green concrete at various temperature ranges using different soft computing techniques, Sustainability 15 (15) (2023) 11907, https://doi.org/10.3390/su151511907, 11907.