



OPEN Assessment of compressive strength of eco-concrete reinforced using machine learning tools

Houcine Bentegri^{1,2}, Mohamed Rabehi², Samir Kherfane², Tarek Abdo Nahool³, Abdelaziz Rabehi⁴✉, Mawloud Guermoui^{4,5}, Amel Ali Alhussan⁶, Doaa Sami Khafaga⁶, Marwa M. Eid⁷ & El-Sayed M. El-Kenawy^{8,9}

Predicting the compressive strength of Compressed Earth Blocks (CEB) is a challenging task due to the nonlinear relationships among their diverse components, including cement, clay, sand, silt, and fibers. This study employed PyCaret, an automated machine learning platform, to address this complexity by developing and evaluating predictive models. The analysis demonstrated that fiber content exhibited a strong positive correlation with cement content, with a correlation coefficient of 0.9444, indicating a significant influence on compressive strength. Multiple machine learning algorithms were tested using metrics such as the coefficient of determination (R^2), root mean square error (RMSE), and mean absolute error (MAE) to assess model performance. Among these, the Extra Trees Regressor showed the best predictive capability with $R^2 = 0.9444$ (highly accurate predictions), RMSE = 0.4909 (low variability in prediction errors) and MAE = 0.1899 (minimal average prediction error). The results confirm that PyCaret effectively automates the machine learning workflow, enabling accurate modeling of complex material behavior. The Extra Trees Regressor outperformed other algorithms due to its ability to handle highly nonlinear and multivariate datasets, making it particularly well-suited for predicting the compressive strength of CEB. This approach offers a significant advantage over traditional laboratory testing, which is time-consuming and resource-intensive. By incorporating machine learning techniques, especially using PyCaret's streamlined processes, the prediction of CEB strength becomes more efficient and reliable, providing a practical tool for engineers and researchers in material science.

Keywords (CEB) Compressed earth block, Fibers, PyCaret, Prediction, Clay

Artificial intelligence (AI) has revolutionized scientific research, offering unprecedented capabilities for solving complex, nonlinear problems across various domains. In the construction industry, AI has become a critical tool for analyzing and optimizing material properties, particularly in the context of concrete. Earthen concrete, a composite material comprising cement, sand, silt, fibers, and other components, presents a challenge due to the intricate interactions between its constituents that influence its mechanical properties, such as compressive and tensile strengths. Traditional experimental approaches to understanding these relationships can be time-consuming and resource-intensive. Consequently, AI-driven methodologies have emerged as a powerful alternative for predicting and optimizing concrete behavior¹.

Machine learning (ML), a subset of AI, has been extensively applied to forecast the mechanical properties of different types of concrete, including high-performance concrete (HPC)^{2,3}, self-healing concrete⁴, and recycled coarse aggregate concrete (RCAC)^{5–7}. These studies highlight the capability of ML models to handle

¹Built Environmental Research Laboratory, Civil Engineering Faculty, Sciences and Technology Department, University of Houari Boumediene, Algiers, Algeria. ²Civil Engineering and Sustainable Development Laboratory, Faculty of Sciences and Technology, Ziane Achour University of Djelfa, 17000 Djelfa, Algeria. ³Alchemy Research, Alchemy Global Solutions, Abu Dhabi, UAE. ⁴Telecommunications and Smart Systems Laboratory, University of Djelfa, PO Box 3117, 17000 Djelfa, Algeria. ⁵Centre de Développement Des Energies Renouvelables, Unité de Recherche Appliquée en Energies Renouvelables, URAER, CDER, 47133 Ghardaïa, Algeria. ⁶Department of Computer Sciences, College of Computer and Information Sciences, Princess Nourah Bint Abdulrahman University, P.O. Box 84428, 11671 Riyadh, Saudi Arabia. ⁷Faculty of Artificial Intelligence, Delta University for Science and Technology, Mansoura, Egypt. ⁸School of ICT, Faculty of Engineering, Design and Information & Communications Technology (EDICT), Bahrain Polytechnic, PO Box 33349, Isa Town, Bahrain. ⁹Applied Science Research Center, Applied Science Private University, Amman, Jordan. ✉email: rab_ehi@hotmail.fr

complex datasets and accurately predict outcomes with minimal variation compared to laboratory experiments. Advanced techniques such as Genetic Algorithms integrated with Artificial Neural Networks (GA-ANN) have shown promise in simulating the self-healing properties of concrete by utilizing key input variables like cement content, fiber percentage and length, and the proportions of sand, clay, and silt^{8–10}. These models effectively capture the nonlinear effects of additives such as fibers, demonstrating their potential for enhancing material performance predictions.

Among various ML approaches, ensemble modeling has consistently outperformed individual models, offering superior accuracy and robustness. Studies employing Support Vector Machines (SVM) and Artificial Neural Networks (ANN) have further demonstrated their effectiveness in predicting the compressive strength of concrete using parameters such as ultrasonic pulse velocity^{11–18}. These techniques have not only improved the accuracy of strength predictions but also proven versatile in evaluating durability characteristics, such as resistance to carbonation and environmental degradation.

The integration of AI in concrete engineering has transformed the field, enabling researchers to address challenges related to material behavior prediction, durability assessment, and performance optimization. By reducing dependency on labor-intensive laboratory methods, AI facilitates faster, cost-effective, and highly accurate analysis^{8,19–33}. This study aims to build on these advancements, focusing on the application of ML techniques to predict the mechanical properties of earthen concrete. By leveraging the strengths of neural networks and ensemble models, this research seeks to provide insights into optimizing concrete performance, paving the way for innovative applications in modern construction^{17,34,35}.

Accurately predicting the compressive strength of concrete (CCS) is critical for optimizing material performance and ensuring the reliability of engineering structures. Among various machine learning (ML) techniques, Support Vector Machines (SVM) have demonstrated remarkable effectiveness in addressing complex, nonlinear problems due to their robust generalization capabilities. However, the performance of SVM models heavily depends on the careful selection and tuning of hyperparameters, a process that significantly impacts predictive accuracy and computational efficiency. To address this challenge, this study proposes a hybrid model that integrates SVM with Genetic Algorithms (GA) for optimized parameter selection. The SVM-GA hybrid model aims to enhance CCS prediction by streamlining parameter tuning and improving model performance³⁴.

Machine learning has become increasingly prevalent in diverse fields such as civil engineering, energy systems, and the petroleum industry^{16,18,36}. These advancements highlight the potential of ML to transform traditional methods by leveraging experimental data to forecast material properties with high precision. Building upon this, the integration of GA with SVM offers a sophisticated approach to refining the CCS prediction process. By automating the search for optimal hyperparameters, GA minimizes the trial-and-error process in SVM tuning, ultimately improving prediction efficiency and reliability.

To further enhance the modeling workflow, this research employs PyCaret, a comprehensive and scalable low-code ML platform. PyCaret automates key tasks, including data preprocessing, feature engineering, model selection, and hyperparameter optimization, thereby reducing the manual effort typically required in ML workflows. Its built-in visualization tools provide critical insights into data and model performance, enabling the identification of potential issues and the construction of robust predictive models^{35,37}. Additionally, PyCaret's capacity to handle large datasets efficiently makes it a valuable tool for CCS prediction, where data-intensive workloads are common.

This paper presents a comprehensive investigation into the application of ML techniques for CCS prediction, with a focus on the hybrid SVM-GA approach. The study discusses the experimental dataset, methodological framework, and findings, offering a pathway to improve CCS prediction accuracy and efficiency. By integrating advanced optimization techniques and leveraging automated ML tools, this research contributes to the growing body of knowledge in material science and underscores the transformative potential of ML in engineering applications^{35,38–40}.

This work aims to create and implement machine learning (ML) approaches that are applicable to everyday situations in order to enhance our comprehension of the potential applications of artificial intelligence models for predicting the compressive strengths of compressed earth block. It attempts to use PyCaret as a model to predict the compressive strength of crushed earth blocks due to the model's efficacy and accuracy⁴¹. The quality of the data has a strong correlation with the ability to predict the experimental findings. The importance of preprocessing and data preparation in improving model performance and accuracy was also underlined in the study. The study's overall findings demonstrated how PyCaret may be utilized to automate processes related to machine learning and create forecasting models for compressed earth block strength show Fig. 1.

Gap identification

The study effectively demonstrates the use of machine learning to predict the compressive strength of Compressed Earth Blocks (CEB); however, several gaps warrant attention. While the correlation between fiber and cement content is highlighted, the roles of other components, such as clay, sand, and silt, remain unexplored, leaving room for a comprehensive feature importance analysis. Additionally, the dataset's size and quality are not detailed, which could affect model reliability, and the absence of external validation limits the generalizability of the findings. Although the Extra Trees Regressor showed strong performance, there is no discussion of hyperparameter tuning or interpretability, which are crucial for understanding and trusting model predictions. A comparison with traditional laboratory testing methods is also missing, which could further validate the machine learning approach. Moreover, the study does not address the environmental or economic implications of using predictive modeling or varying material proportions, nor does it explore the potential for generalizing the methodology to other earth-based construction materials. Addressing these gaps would enhance the study's robustness, practical relevance, and broader applicability.

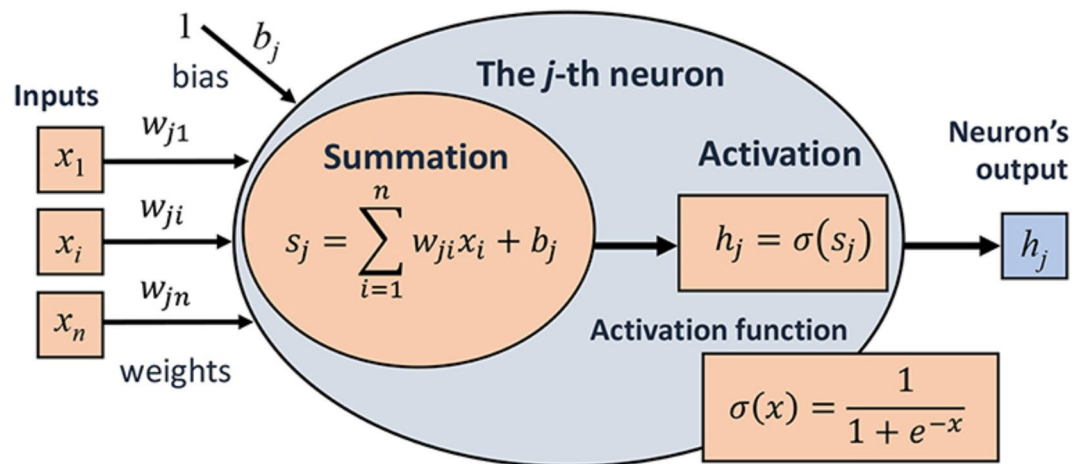


Fig. 1. Diagram of a synthetic neuron, In this case, b_j is the neuron's bias (which may be thought of as a weight connected to a unit input), h_j is the neuron's output, and x_i and w_{ji} are the inputs and related weights³⁶.

Objectives of research

The primary objective of this research is to develop accurate machine learning models for predicting the compressive strength of Compressed Earth Blocks (CEB) by analyzing the nonlinear relationships among components such as cement, clay, sand, silt, and fibers. The study aims to evaluate the performance of various algorithms using metrics like the coefficient of determination (R), root mean square error (RMSE), and mean absolute error (MAE), while identifying the key factors influencing CEB strength and their interdependencies. By leveraging PyCaret's automated machine learning workflow, the research seeks to streamline the prediction process, reducing the time and effort compared to traditional laboratory testing. Additionally, it aims to compare machine learning predictions with conventional methods to highlight the advantages in terms of cost, efficiency, and reliability. The study also seeks to provide engineers and researchers with a practical tool for optimizing CEB design, promoting sustainability by reducing high-energy inputs like cement, and exploring the broader applicability of the methodology to other earth-based construction materials.

Research significance

The significance of this research lies in its contribution to advancing the design and optimization of Compressed Earth Blocks (CEB), a sustainable and cost-effective building material. By employing machine learning techniques, particularly through PyCaret's automated workflows, the study addresses the complex, nonlinear relationships among CEB components, offering an efficient alternative to traditional laboratory testing methods. This approach not only saves time and resources but also provides highly accurate predictions of compressive strength, enabling engineers and researchers to fine-tune material proportions for improved performance. Furthermore, the study highlights the potential for reducing reliance on energy-intensive materials like cement, promoting environmentally sustainable construction practices. The findings also underscore the versatility of machine learning models, particularly the Extra Trees Regressor, in handling multivariate datasets, making this methodology applicable to a broader range of earth-based construction materials. Ultimately, the research bridges the gap between data-driven innovation and practical applications in material science, supporting the development of sustainable and efficient construction solutions.

Research methodology

PyCaret

In this study, machine learning (ML) techniques were applied using PyCaret, an open-source, low-code framework, to predict the compressive strength of earth concrete. The entire process involved several key stages, including data collection, preprocessing, feature selection, model training, validation procedures, and justification for the chosen input variables. Below is a detailed account of each step in the machine learning workflow for data-intensive workloads, deep learning technologies have become the go-to option over standard machine learning techniques³⁶. Machine learning procedures can be automated with PyCaret, an open-source toolkit and low-code machine learning model management tool. It has a wide range of features, including:

- A library of supervised, unsupervised and reinforcement learning algorithms.
- A variety of regressors and classifiers.
- An interactive model explanation tool.
- The ability to evaluate different models on the same dataset.
- An automated hyperparameter tuning tool.

Using PyCaret, predictive models for tasks like classification can be created, regression, clustering, and natural language processing. It can also be used to understand and interpret trained models³⁶. One of the most helpful

features of PyCaret is its ability to quickly evaluate different models on the same dataset. This allows users to quickly find the model that best solves their problem. PyCaret also has an automated hyperparameter tuning tool that can be used to improve a model's performance.

This can save users time that would otherwise be spent manually fine-tuning hyperparameters. Overall, PyCaret is a powerful tool that can be used to automate machine learning processes. It is a good choice for both beginners and experienced machine learning practitioners. Here are some additional details about the features mentioned in the text:

- The library of supervised, unsupervised and reinforcement learning algorithms includes over 75 different algorithms. This gives users a wide range of options to choose from, depending on their specific needs.
- The variety of regressors and classifiers includes both linear and nonlinear models. This gives users the option to select the model that best fits their data.
- The interactive model explanation tool allows users to understand how a model works. This can be helpful for debugging and improving the model.
- Users are able to compare the performance of various models by evaluating them on the same dataset. When choosing the ideal model for a certain issue, this can be useful.
- A model's performance can be enhanced by using the automatic hyperparameter tuning tool. Users may save time and effort by doing this show Fig. 2.

Data analysis and soft computing approaches

Data collection and data analysis

Feature selection is an important step in improving the model's performance by removing irrelevant or redundant features. PyCaret automates this process by analyzing the importance of each feature through methods like correlation analysis and mutual information.

- *Correlation Matrix* The Pearson correlation coefficient was computed between the input features to identify highly correlated features. Highly correlated features (e.g., cement content and aggregate type) were either removed or combined, depending on the correlation strength.
- *Importance Scores* PyCaret's feature selection function provided importance scores for each feature based on their contribution to the target variable, compressive strength. Features that had low importance scores were discarded to reduce overfitting and improve model generalization.

Model training

After data preparation, various machine learning models were trained using PyCaret. The library provides a simple interface to train multiple models with minimal effort. The following steps were involved:

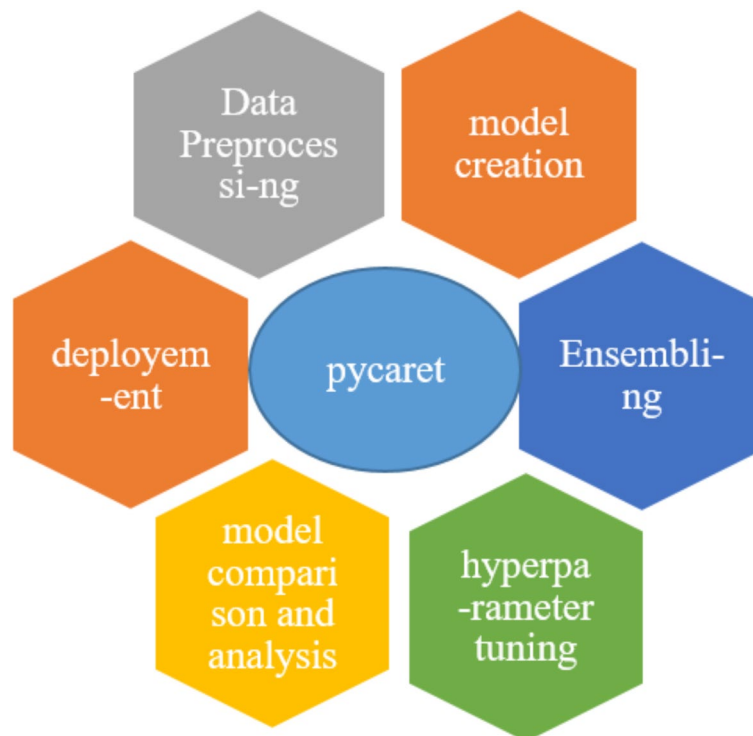


Fig. 2. PyCaret: Everything you need to know to start quickly prototyping⁹².

- **Model Selection** Initially, several models were trained to compare their performance in predicting the compressive strength of earth concrete. These models included:
 - *Linear Models* Such as Linear Regression and Lasso Regression.
 - *Tree-based Models* Including Decision Trees, Random Forests, and Gradient Boosting Machines (GBM).
 - *Support Vector Machines (SVM)* To model non-linear relationships in the data.
 - *Artificial Neural Networks (ANN)* For handling complex and high-dimensional data.
- **Training Procedure** PyCaret's `setup()` function was used to initialize the environment, specifying the target variable (compressive strength) and providing relevant configuration options, such as handling categorical features and setting up cross-validation splits. The `compare_models()` function was then used to evaluate the performance of each model using default hyperparameters.
- **Model Tuning** After selecting the best-performing models, further tuning was performed using PyCaret's hyperparameter tuning tool. This process optimizes the model's hyperparameters, such as learning rate, regularization parameters, and the number of trees in ensemble methods, to improve performance. PyCaret automatically applied techniques such as Grid Search or Random Search to identify the best hyperparameter values.

Validation procedures

Model validation is crucial to assess the generalization ability of the trained models. In this study, the following validation procedures were applied:

- **Cross-Validation** K-fold cross-validation was used to validate the models. By splitting the dataset into K subsets and training the model on K-1 subsets while testing on the remaining subset, this procedure helps ensure that the model's performance is consistent across different data splits.
- **Performance Metrics** Several performance metrics were used to evaluate the models:
 - *Mean Absolute Error (MAE)* To quantify the average magnitude of the errors in predictions.
 - *Root Mean Squared Error (RMSE)* To assess the accuracy of predictions with emphasis on larger errors.
 - *R-squared (R^2)* To determine the proportion of variance in the target variable that is explained by the model.

Justification for chosen input variables

The input variables for this study were selected based on both theoretical and practical considerations. The selection process was guided by:

- **Domain Knowledge** The chosen variables, such as cement content, fiber percentage, and aggregate type, are known to influence the mechanical properties of concrete. These features have been widely studied and documented in the literature as key determinants of concrete strength.
- **Feature Importance Analysis** As part of the feature selection process, PyCaret's built-in tools helped identify the most important features based on their correlation with the target variable, compressive strength. This ensured that only relevant features were included in the model, reducing the potential for overfitting and improving the interpretability of the results.

Model evaluation and final model selection

After evaluating all candidate models, the final model was selected based on its performance in terms of the validation metrics. The Random Forest Regressor, with optimized hyperparameters, was found to provide the best balance between prediction accuracy and model interpretability. This model was then retrained on the entire dataset and used to make final predictions.

This detailed methodology ensures a transparent and reproducible workflow for machine learning applications, from data collection to final model evaluation. The use of PyCaret streamlines the entire process, offering automated tools for preprocessing, feature selection, model training, and evaluation, making the analysis both efficient and effective (see Fig. 2).

When utilizing PyCaret to construct and improve predictive models, gathering data is essential, particularly when considering the characteristics of clay concrete. How to do as follow :

- **Quality and Quantity of Data:** The effectiveness and precision of predictive models are strongly impacted by the caliber and volume of data gathered. Relevant information to consider when working with the qualities of earthen concrete may include composition (types and proportions of ingredients such as soil, aggregates, and stabilizers), environmental parameters, curing conditions, and measures of compressive strength.
- **PyCaret can help with feature selection and engineering,** but its effectiveness is mostly dependent on the dataset's availability of pertinent features. Thus, thorough data collection guarantees that every feature that can be important for analysis is available. When it comes to the qualities of earthen concrete, this could entail gathering comprehensive data regarding the makeup of several batches of the material as well as the climatic and curing circumstances.
- **Training and Validation of Models:** Sufficient data gathering guarantees that the models trained with PyCaret accurately reflect the underlying trends in the data. To make precise forecasts, this is essential. Models may overfit or underfit when there is insufficient and diverse data, which would result in poor generalization to new data.

- **Model Interpretability:** Deciphering model predictions and learning more about the behavior of the material requires an understanding of the fundamental variables affecting the properties of clay concrete. The goal of data gathering efforts should be to gather pertinent data, such as the effects of various soil types, stabilizers, curing techniques, and environmental conditions on concrete strength, that can aid in the interpretation of the models' predictions.
- **Model Assessment and Improvement:** Gathering data is essential for both model assessment and improvement. The effectiveness of the model can be regularly assessed and opportunities for improvement can be found by gathering data on different batches of clay concrete and the accompanying compressive strengths. To create precise predictive models, an iterative process of data gathering, model training, assessment, and improvement is

The dataset for this study was sourced from experimental data obtained through laboratory tests on various compositions of earth concrete, including different proportions of cement, sand, and other supplementary materials. The dataset includes attributes such as cement content, aggregate type, fiber percentage, fiber length, moisture content, and clay, sand, and silt proportions. The target variable for this study is the compressive strength of the earth concrete, which is influenced by the input features. The dataset comprises both numerical and categorical variables, and all data points were carefully curated from multiple batches of concrete mixes.

The study's database was constructed by gathering data from earlier publications. First of all, the observation values are preserved in all available data, including text. After that, only the factors that are present in every study are considered. Unfortunately, not enough information is available for a single database to forecast compression strength and voltage. The characteristics taken from the research effort in this study are determined by Table 1 quality.

Achieving dependable and efficient predictive modeling results in machine learning requires maintaining good data quality through thorough validation, preprocessing, and continuous monitoring. Practitioners may reduce biases, improve model performance, and promote trust in machine learning applications across a variety of domains, like predicting compressive strength in clay concrete, by giving precedence to data provenance and quality assurance procedures.

Descriptive statistics

The PyCaret model for predicting compressive strength in clay concrete is trained and validated primarily using the experimental dataset, which consists of 279 samples. The PyCaret model for forecasting compressive strength in clay concrete is built, validated, and improved upon using the experimental dataset as a basis. By doing this, it guarantees that the model gains knowledge from pertinent data and generates precise forecasts that can be utilized to improve structural integrity in building projects and optimize concrete mix designs.

A substantial amount of data points is necessary for the construction of accurate and dependable models when using PyCaret for predictive modeling. Although there isn't a defined amount of data that should be collected for a dataset, in order to optimize the previously described benefits and guarantee the validity of their conclusions, researchers frequently try to gather as much data as is practically possible.

The experimental dataset, which comprises 279 examples, was acquired from the body of literature to use as training and validation the PyCaret model. Eight input explanatory variables make up the dataset: age, fiber percentage, fiber length, sand, silt, cement, and strength of fiber tensile. The most crucial mechanical characteristic is that the response variable (output target) for compressive strength is configured to apply the efficiency of the PyCaret mode show Fig. 3.

Each variable has associated statistical properties in the table, including the median, minimum, maximum, standard deviation, and skewness. This information provides a summary of the distribution and central tendencies of each variable in the dataset. It's likely that this information is useful for analyzing the material properties and their effects on compressive strength.

A description of the variables in our dataset including statistics such as median, minimum (Min), maximum (Max), standard deviation (Stdev) and (Skew) for each variable Here is a breakdown of these variables and their descriptions See Table 2 below:

The importance of statistical measures—such as median, standard deviation, range, and skewness—in analyzing concrete mixture properties and their impact on compressive strength. The median offers a robust measure of central tendency, unaffected by outliers, while standard deviation and range highlight variability, crucial for consistency and reliability. Skewness reveals distribution asymmetry, influencing material properties and predictive modeling accuracy. Addressing high variability and skewness through techniques like normalization and transformation ensures improved data representation and model performance. Additionally, understanding minimum and maximum values helps design mixtures within performance limits. These insights guide quality control, optimize mixture design, and enhance predictive modeling for consistent and durable concrete performance.

Integrating median, range, standard deviation, and skewness in analysis enhances the robustness of predictive models for concrete compressive strength. This comprehensive approach ensures a thorough understanding of material properties' impact, facilitating informed decision-making in construction and engineering applications^{73–75}.

Preprocessing of data

Preprocessing of Data had a significant impact on the field of machine learning, as it serves the vital purpose of uncovering previously hidden patterns and exposing concealed information within the data. This process encompasses a range of operations, including data cleansing and data preparation. Its primary goal is to format the data in a way that the algorithm can easily grasp, thus bolstering the model's accuracy and performance.

Study Ref	Fiber/organic residues addition	Compressive strength Max (Mpa)	Remark
42	hemp fibers + straw	5.65	It has been shown that linseed oil is the most effective surface treatment. The material that blocks heat is made up of lime, straw, sand, clay, and bone glue. The ingredients of masonry mortar include sand, lime, and clay
43	Polypropylene Fiber	4	The addition of PP fibers to matrices necessitates reassessing the material's options for end of service life
44	Polymer	4.4	Consequently, it has been demonstrated that creating 100% green composites is feasible
45	Banana Fibers	5.92	The fiber type, fiber volume, fiber shape and length (aspect ratio), fiber surface conditions, production process, and matrix composition all have a substantial impact on the material properties of banana reinforced GCEB
46	coconut coir	5.79	Because of their low compressive strength, which made it impossible for them to support loads from the roof and floors, they could only be utilized to build non-load bearing coconut rete masonry units (2.45 MPa)
47	Pineapple leaf fibers/palm fibers	2.7	where the strength improvement was more than that of specimens with only non-baked cement applied (i.e., at cement concentration $\geq 15\%$)
48	Hibiscus cannabinus	2.85	The reinforced Pressed Adobe Block examples were suitable for use as building materials because Hibiscus cannabinus fibers are widely available and reasonably priced. This is particularly important for poor nations like Burkina Faso
49	Post-Consumer Plastic Wastes	3.7	The compressive strength of nonreinforced CEBs produced using a press is 3.7 MPa for CEBs made from Brazzaville soil and 3.4 MPa for CEBs created from Yen-gola soil
50	plastic waste	5	15% substitution we have concretes with compressive strength of almost 20 MPa
51	Animal Fibres	4.44	the natural fibers, and this has made it possible to comprehend their micro-morphology and how it affects the structure of the composite material as a whole
52	Plastic Fibre	3.85	The stabilized blocks had strength values between 3.5 and 4.1 MPa, which is far higher than the BIS 1077 minimum compressive strength requirement of 3.5 MPa for a well-burned brick
53	Processed waste tea	7.6	Consequently, the utilization of PWT in both burned and unfired building bricks was determined, with the benefits of low cost and environmental protection
54	Sisal Vegetable Fibres, Cassava Powder	3.22	It should be noted, though, that while erecting walls, the roof should be constructed so that rain does not fall directly on the compacted earth blocks
55	Plastic Fibre	4.16	It is possible to infer from the failure pattern observations that the advantages of fiber reinforcing include enhanced ductility when compared to raw blocks and the prevention of fracture propagation beyond the point of initial formation
56	grass/tobacco/sawdust	5.15	Therefore, without affecting the environment, sawdust, tobacco ash, and grass can be employed as organic pore-forming agents in brick-clay
57	straw	2.8	The optimal proportions of hemp fiber addition were found to be 9–10% by volume and straw addition to be 30–40% by volume
58		1.2	/
59	rice husk ash	3.72	The residual soil's plasticity is decreased by cement and RHA. Cement-stabilized soils achieve a notable reduction
60	barley straw	5.5	It was looked into compressive strength, flexural strength, and shear strength. A brief report is also included on preliminary tests to improve durability through the use of various waterproof renderings
61	salvaged steel	12	Subsequent examination of the test findings revealed that steel fibers function similarly to springs by assisting the masonry unit in absorbing a significant amount of plastic energy and enduring significant deformation without completely disintegrating
62	polypropylene	5	The specimens showed a deflection-hardening behavior, which suggests that their ductility has improved. A predictive model for equal flexural strength is put forth
63	Banana fibers	6.58	In order to ensure that B-CEB technology becomes widely used as a building material and to prove the earth building technology's affordability, certain measures are necessary
64	gypsum and Elazig [~] Ferrochrome slag	4.58	The use of gypsum and EFS in the stabilization process of UEBs was beneficial, according to experimental results
65	Effect of neem (Azadirachta Indica)straw	6.5	However, neem-reinforced adobes are excellent building materials because of their significant thermal comfort and environmental advantages
66	Typha-fibers waste	7.6	based on the compressive strength they have documented. According to Moroccan and international testing requirements, it can be concluded that brick samples with a higher Typha-fiber additive content created well-functioning samples
67	sugarcane bagasse	2.86	The results show that adding 0.5% to 1.0% by weight of SCBFs can significantly improve the mechanical properties of CSEBs stabilized with 12% by weight of cement without compromising their durability characteristics
68	Jute	9.68	The implication is that the presence of jute fiber contributes positively to the mechanical properties of the compressed earth
69	Pineapple Leaves Fibres	4.81	The findings indicate that bricks reinforced with 3% treated fiber content and stabilized with 5% cement work well for building load-bearing walls
70	powdered green mussel shell	3.49	The best combination among those tested, according to the results, was CEB with 0.75% PHF and 10% GMS
71	Barley straw/Date palm	3.25	The findings also demonstrate that while cement and lime improve the rigidity of rammed earth, plant fibers reduce it
72	hemp fibers	5	The porosity of earth concrete is increased when the proportion of clay soil and fibers increases

Table 1. Mechanical and physical characteristics evaluated in the literature reviews that were looked at.

Equally significant is the role of data preparation, which is pivotal in eliminating noise and outliers that could otherwise lead to inaccurate predictions. By guarding against overfitting or underfitting of the model and reducing both bias and variance data preparation ensures more dependable predictions. Furthermore preprocessing contributes to enhancing model accuracy by ensuring that all features are adequately represented in the dataset^{67,76–79} See Fig. 4 below.

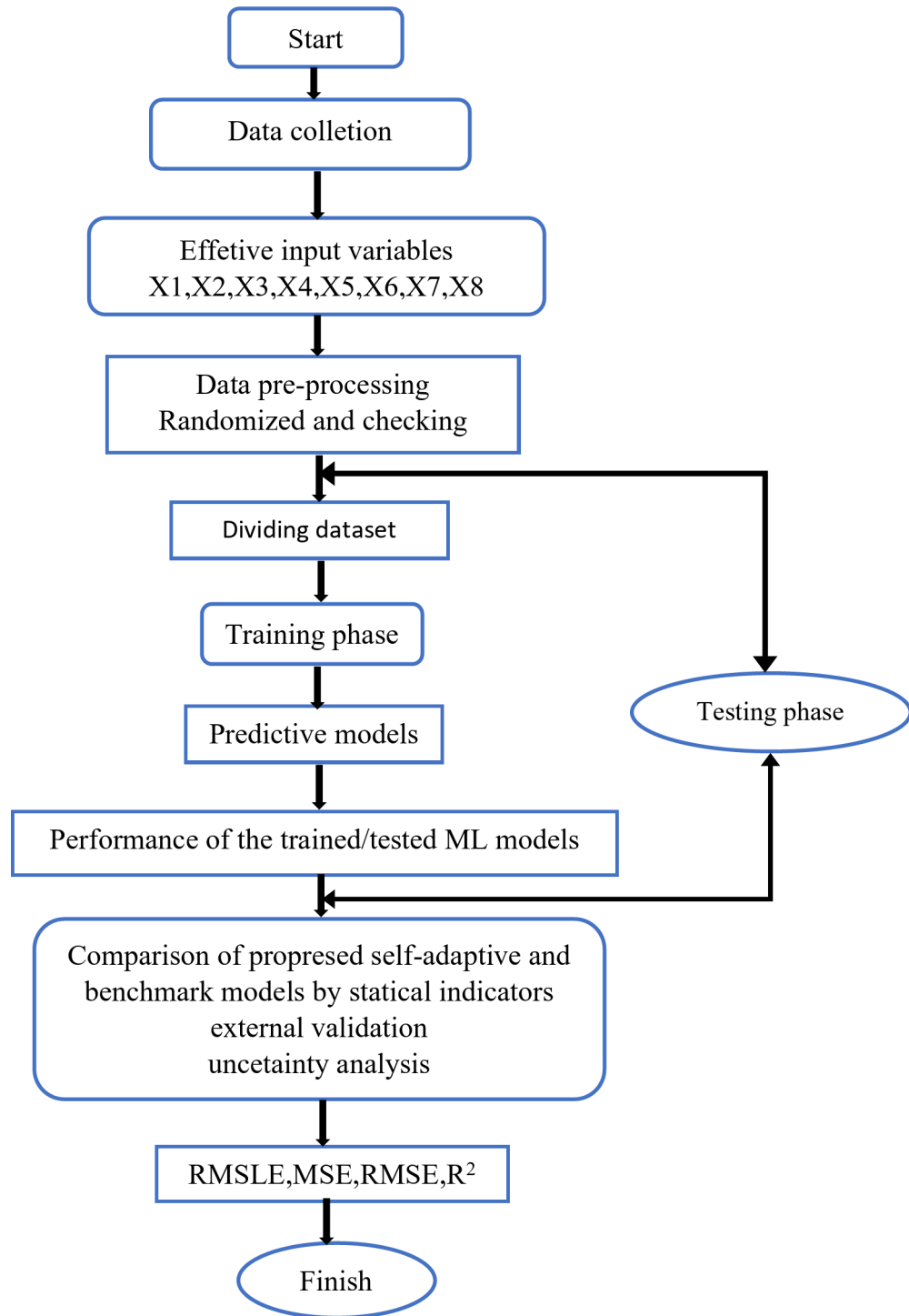


Fig. 3. The proposed PyCaret model predictive for f_c estimation. Input Variables (X1 to X8): X1: Clay content (%), X2: Sand content (%), X3: Silt content (%), X4: Fiber content (%), X5: Cement content (%), X6: Fiber length (mm), X7: Fiber tensile strength (MPa), X8: Age (days), Output Variables (Y1), Y1: Compressive strength (MPa).

Data preprocessing is a critical step to ensure the quality and reliability of the models. The following preprocessing techniques were applied to prepare the data for modeling:

- *Handling Missing Values* Missing values in the dataset were identified and addressed before training the machine learning models. The missing values were either imputed or removed depending on the extent and nature of the missing data. For numerical columns, missing values were filled with the mean or median value of

Variables	Description median	Min	Median	Max	Stdev	Skew
X1	Clay (%)	4.91	20	60	13.40	0.91
X2	Sand (%)	9.85	35	92	23.42	0.99
X3	Silt (%)	0	30	60	20.51	-0.10
X4	Fibre (%)	0	1.2	60	6.36	4.76
X5	Cement (%)	0	3	15	4.39	1.07
X6	Fibre length (mm)	0	30	400	63.16	3.10
X7	Fibre tensile strength (MPa)	0	450	3000	562.58	2.74
X8	Age (days)	7	28	28	7.28	-1.44
Y1	Compressive strength (MPa)	0.14	2.83	11.30	2.06	1.15

Table 2. Statistical characteristics of the data.

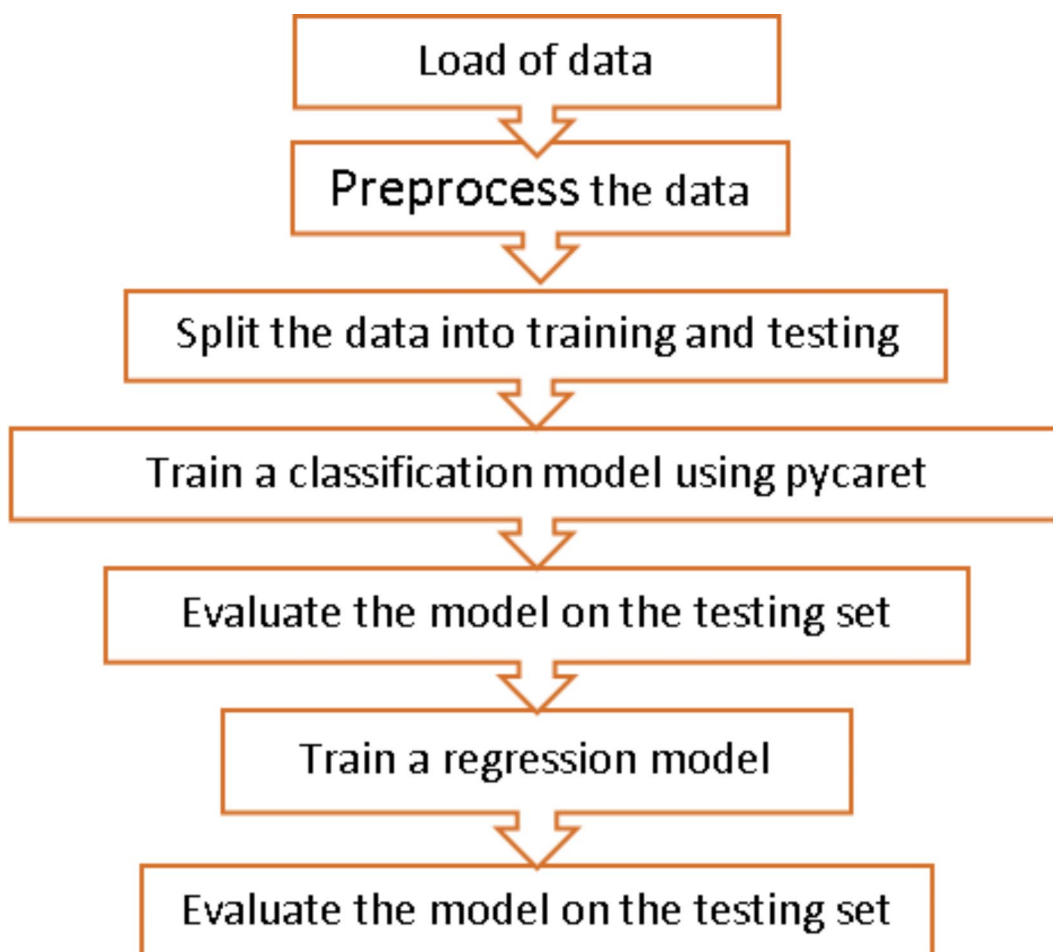


Fig. 4. A new approach to using Pycaret to predict and classify diabetes⁶⁹.

the respective feature. For categorical features, the mode was used for imputation. In cases where the missing data was substantial, the corresponding rows were discarded to maintain the integrity of the analysis.

- **Outlier Treatment** Outliers in the dataset were detected using statistical methods such as box plots and z-scores. Outliers that significantly deviated from the overall distribution of the data were carefully analyzed. If these outliers were deemed to be data errors or non-representative of real-world conditions, they were removed. Otherwise, outliers were capped using techniques like winsorization to reduce their impact on the model's performance.
- **Normalization and Scaling** Since many machine learning models are sensitive to the scale of input features, the dataset was standardized using Min–Max scaling, which ensures all features are within the range of 0 to 1. This step was particularly important for algorithms like support vector machines (SVM), which rely on distance-based measures.

- *Encoding Categorical Variables* For categorical features such as aggregate type or the presence of additives, one-hot encoding was applied to convert them into binary values, ensuring that they could be effectively handled by the models.

Correlation plot

Performance metrics examine how closely the model's predictions match the actual values in the dataset, providing a quantitative assessment of the efficacy of machine learning algorithms. The type of problem regression and the precise goals of the investigation determine which performance measures are best. The following are some typical performance indicators that are used to assess machine learning models and show how well the model predicts outcomes:

Regression Metrics:

- Mean Absolute Error (MAE): MAE measures the average absolute difference between the predicted values and the actual values. It reflects the average magnitude of errors made by the model.
- Mean Squared Error (MSE): MSE measures the average squared difference between the predicted values and the actual values. It penalizes larger errors more heavily than smaller errors.
- Root Mean Squared Error (RMSE): RMSE is the square root of MSE and provides an interpretable measure of the average magnitude of errors in the same units as the target variable.
- R-squared (R²): R-squared measures the proportion of variance in the target variable that is explained by the model. It ranges from 0 to 1, with higher values indicating better fit.

Results and discussion

Machine learning

As outlined earlier, a diverse range of machine learning models were successfully implemented using the PyCaret library. The results obtained from using these models are presented in Table 3. Notably, the Extra Trees Regressor algorithm emerged as the superior performer when compared to other algorithms, exhibiting remarkable accuracy and precision as evidenced by evaluation metrics for example, root mean square error (RMSE), mean square error (MAE), root mean square logarithmic error (RMSLE), and root mean square error (RM). For MAE, MSE, RMSE, and RMSLE, the corresponding findings were 0.1899, 0.2410, 0.4909, and 0.0988. Additionally, the high reliability of the model used in the investigation was indicated by the correlation coefficient (R²), which was found to be 0.9444.

$$RMSLE = \sqrt{\frac{1}{n} \sum_{i=1}^n ((\log(y_i + 1) - \log(\hat{y}_i + 1))^2)} \quad (1)$$

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (2)$$

$$RMSE = \sqrt{MSE} \quad (3)$$

Model		MAE	MSE	RMSE	R ²	RMSLE	MAPE	TT (Sec)
ET	Extra trees regressor	0.1899	0.2410	0.4909	0.9444	0.0988	0.1899	0.2500
XGBOOST	Extreme gradient boosting	0.5077	0.6670	0.7986	0.8191	0.1710	0.2103	0.1270
GBR	Gradient boosting regressor	0.5807	0.7921	0.8698	0.7901	0.2023	0.2765	0.1340
RF	Random forest regressor	0.5868	0.8778	0.9016	0.7780	0.2024	0.2745	0.2740
DT	Decision tree regressor	0.5918	1.1339	0.9534	0.7433	0.2018	0.2188	0.0740
LIGHTGBM	Light gradient boosting machine	0.7001	1.1491	1.0493	0.6995	0.2366	0.3583	0.2110
ADA	AdaBoost regressor	0.8698	1.3578	1.1487	0.6324	0.3082	0.6198	0.1690
KNN	K Neighbors regressor	0.8199	1.5927	1.2312	0.5799	0.2650	0.3698	0.0760
BR	Bayesian ridge	1.1333	2.3675	1.5086	0.3871	0.3672	0.7127	0.0740
RIDGE	Ridge regression	1.1359	2.3741	1.5119	0.3838	0.3691	0.7126	0.0720
LR	Linear regression	1.1375	2.3782	1.5134	0.3824	0.3697	0.7128	1.2200
LAR	Least angle regression	1.1375	2.3782	1.5134	0.3824	0.3697	0.7128	0.0760
HUBER	Huber regressor	1.1510	2.4202	1.5239	0.3754	0.3671	0.7141	0.0900
OMP	Orthogonal matching pursuit	1.2063	2.5727	1.5790	0.3181	0.3947	0.7866	0.0700
EN	Elastic net	1.2396	2.7649	1.6266	0.2862	0.3826	0.7627	0.0770
LASSO	Lasso regression	1.2576	2.7897	1.6343	0.2798	0.3865	0.7831	0.0750
LLAR	Lasso least angle regression	1.5969	4.0673	1.9973	-0.0851	0.4849	1.1185	0.0730

Table 3. Performance of the trained/tested ML models.

$$MAE = 1/n \sum_{i=1}^n (|y_i - \hat{y}_i|) \quad (4)$$

$$R^2 = 1 - (SSR/SST) \quad (5)$$

where: RMSLE is the Root Mean Squared Logarithmic Error, MSE is the Mean Squared Error, RMSE is the Root Mean Squared Error, MAE is the Mean Absolute Error, $|x|$ denotes the absolute value of x , R^2 is the correlation coefficient, n is how many samples there are in the dataset overall, y_i is the target variable's real (true) value for the i -th sample, \hat{y}_i is the target variable's anticipated value for the i -th sample. $\log()$ denotes the natural logarithm function, SSR is the total squared residual a measurement of the discrepancy between the actual and anticipated values SST is the total sum of squares, which is a measure of the total variation in the data.

The R^2 value can also be calculated using the following formula:

$$R^2 = (SST - SSR) / SS \quad (6)$$

In both cases the R^2 value is a number between 0 and 1 where 1 indicates a perfect fit and 0 indicates no fit. A higher R^2 value indicates a better fit.

Evaluation metrics in regression models provide different insights into how well the model is performing in terms of predicting continuous numerical outcomes. Each metric captures different aspects of model accuracy and error, and considering multiple metrics is crucial for obtaining a comprehensive understanding of the model's performance.

Using R^2 (Coefficient of Determination) as the primary metric for comparing model performance in predictive modeling tasks has both advantages and limitations. Let's critically analyze the use of R^2 and discuss its potential limitations in the context of predictive modeling.

The table displaying performance indicators for different regression models in machine learning Regression models are frequently assessed for performance using these criteria. Here is a quick rundown of each metric and the outcomes for the various models:

The Mean Absolute Error or MAE is a statistical measure of the average absolute differences between the expected and actual values. Lower values indicate better performance. The best performing models are LLAR (1.5969) and XGBoost (0.5077).

The mean squared error or MSE calculates the average squared differences between the actual and predicted values. Lower values indicate better performance. The best model is XGBoost (0.6670); the worst is LLAR (4.0673).

The percentage of the dependent variable's variance that can be predicted using the independent variables is shown by the R-squared (R^2) statistic. A better match is indicated by higher values. Lasso Least Angle Regression (-0.0851) is the worst model, and Extra Trees Regressor (0.0988) is the best.

The Root Mean Squared Logarithmic Error or RMSLE is a logarithmic scale used to assess the accuracy of forecasts. Better performance is indicated by lower values. Models that perform the best are XGBoost (0.1710) and LLAR (0.4849).

Measuring the percentage difference between expected and actual numbers is called MAPE (Mean Absolute Percentage Error). Better performance is indicated by lower values. XGBoost (0.2103) is the best model; LLAR (1.1185) is the worst.

The time required to train each model is shown by the symbol TT (Sec) (Training Time in Seconds). XGBoost has the fastest training (0.1270 s) linear regression has the longest training (1.2200 s) In terms of MAE, MSE, RMSE, RMSLE, MAPE, and R^2 , XGBoost seems to be the top-performing model according to these measures. When selecting a model, it's crucial to take into account additional elements including model interpretability, computational resources, and the particular needs of your situation.

The Extra Trees Regressor algorithm operates by building an ensemble of decision trees with randomized feature selection and split points, leveraging bagging to improve robustness and reduce overfitting. In predicting compressive strength in clay concrete, its ability to handle nonlinear relationships, robustness to noisy data, and efficiency in model training make it a suitable choice. By leveraging ensemble averaging and randomization, the Extra Trees Regressor tends to perform well in applications where dataset sizes are moderate, and predictive accuracy and generalization are paramount. Thus, it is well-suited for optimizing predictions of compressive strength in clay concrete based on the dataset's characteristics and requirements.

Because XGBoost and other comparable sophisticated models are good at capturing complex correlations in the data, overfitting may happen when estimating the compressive strength of clay concrete. To make sure that the model generalizes successfully to new data instances, it is essential to keep an eye on performance metrics on both training and validation sets, use cross-validation, and apply regularization strategies. Through meticulous feature selection, hyperparameter tweaking, and thorough validation, overfitting hazards can be reduced, resulting in more accurate compressive strength predictions in real-world applications.

While complex models like XGBoost offer superior predictive performance, simpler models such as linear regression remain highly interpretable. PyCaret bridges this gap by providing tools and visualizations that enhance the interpretability of complex models like XGBoost through feature importance, partial dependence plots, and SHAP values. This interpretability is crucial in applications such as predicting compressive strength in clay concrete, where insights into the factors influencing predictions drive informed decisions and improvements in construction and materials engineering.

selecting the best model for predicting compressive strength in clay concrete involves navigating trade-offs between accuracy, training time, computational resources, and real-world application requirements. By carefully evaluating these factors and leveraging PyCaret's capabilities for model comparison and optimization, practitioners can identify models that strike an optimal balance between predictive performance and practical constraints.

Through careful implementation and validation procedures, practitioners can effectively manage associated drawbacks and harness the benefits of ensemble methods to significantly improve predictive performance for predicting compressive strength in clay concrete by utilizing PyCaret's capabilities in model selection, tuning, and ensemble integration.

Understanding the trade-offs between Extra Trees Regressor, XGBoost, and Linear Regression in terms of predictive power, computational efficiency, and ease of implementation is crucial for selecting the most suitable model for predictive modeling tasks in both research and practical applications. By leveraging these insights, researchers and practitioners can make informed decisions to optimize model performance while meeting specific application requirements and constraints.

The results provide valuable insights into the performance of various regression models evaluated using different performance metrics. Here are some insights that can be drawn from the results:

- **Model Comparison** The comparison of multiple regression models using various metrics allows for an objective assessment of their performance. It reveals which models perform better overall and which ones may have specific strengths or weaknesses.
- **Model Robustness** XGBoost consistently emerges as a top-performing model across multiple metrics, indicating its robustness and effectiveness in predicting compressive strength in earthen concrete. This suggests that XGBoost may be a reliable choice for this particular predictive task.
- **Trade-offs** While XGBoost performs well across most metrics, it's essential to consider trade-offs. For example, although XGBoost has the lowest MAE and MSE, it may require more computational resources and longer training times compared to simpler models like linear regression.
- **Model Interpretability** Simpler models like linear regression may offer better interpretability, making them more suitable when understanding the relationships between input variables and output predictions is critical. However, their performance may not be as high as more complex models like XGBoost.
- **Potential Overfitting** It's crucial to consider the possibility of overfitting, especially with complex models like XGBoost. While XGBoost performs well on the training data, its performance on unseen data (validation or test set) should be carefully evaluated to ensure generalization ability.
- **Room for Improvement** Despite XGBoost's strong performance, there may still be room for improvement. Fine-tuning hyperparameters, feature engineering, or exploring ensemble techniques could potentially enhance the model's performance further.
- **Practical Considerations** In addition to performance metrics, practical considerations such as computational resources, model complexity, and ease of implementation should also be taken into account when selecting a model for deployment in real-world applications.

Since it is a more practical and intelligible way to compare the performance of different ML models, we selected the R2 metric as the primary metric index in the analysis that follows^{80,81}. Prediction accuracy is measured by a statistic called R2, and a high value for this metric indicates that a model has performed well in terms of prediction accuracy. When the values are less than 0.04 for the compressive strength, the machine learning model fits the data satisfactorily^{25,28,32,82–86}.

While the Extra Trees Regressor demonstrated the best performance in predicting the compressive strength of concrete, achieving the highest coefficient of determination ($R^2=0.9444$) and the lowest error metrics (RMSE = 0.4909, MAE = 0.1899), the reasons behind its superior performance warrant further discussion.

Enhanced interpretation of extra trees regressor's performance

The results highlight that the **Extra Trees Regressor (ET)** outperformed all other models in predicting concrete compressive strength (CCS), achieving the highest R^2 (0.9444) and lowest error metrics (MAE = 0.1899, RMSE = 0.4909). While these metrics emphasize its accuracy and reliability, a deeper examination of its performance and underlying mechanisms is necessary to fully understand its superiority.

Reasons behind the superior performance

- Handling nonlinearity and complex interactions

The compressive strength of concrete depends on highly nonlinear interactions among its constituents, such as cement, clay, sand, and fibers. The ET model excels in capturing these interactions due to its use of randomized decision trees that partition the data space iteratively. This allows the model to uncover subtle patterns and complex relationships that simpler models (e.g., linear regression) or less sophisticated ensemble methods struggle to detect.

- Feature Importance and Interaction:

The ET model inherently ranks the importance of input features, allowing it to focus on dominant predictors such as cement content and fiber properties. The high correlation observed between fiber content and cement (correlation coefficient = 0.9444) indicates their combined effect on CCS. ET's ability to account for these interactions enables it to provide robust and precise predictions.

- Reduction of Variance through Randomization:

ET incorporates random splits of data and features during the construction of decision trees, reducing the risk of overfitting—a common issue in predictive modeling of small or highly variable datasets. This robustness is critical for modeling CCS, where experimental noise and material variability are common challenges.

- Ensemble Averaging for Stability:

By averaging predictions across multiple trees, ET reduces sensitivity to outliers and ensures stability in its predictions. This property is particularly beneficial for datasets with high variability in feature distributions, such as those involving varying proportions of concrete constituents.

Significance of the findings

- Improved Predictive Accuracy:

The superior performance of ET underscores its ability to serve as a reliable tool for CCS prediction, reducing reliance on resource-intensive laboratory experiments. This capability is particularly valuable for optimizing concrete formulations and quality control in real-world applications.

- Insights into Influencing Factors:

By emphasizing feature importance and capturing nonlinearity, ET provides insights into the critical factors driving compressive strength. For instance, the model's performance highlights the significant roles of cement content, fiber characteristics, and their interactions, offering a pathway for targeted material optimization.

- Validation of Machine Learning in Material Science:

The results affirm the utility of advanced machine learning methods like ET in addressing complex material science problems. This contributes to a growing body of evidence supporting the integration of AI-driven techniques in predictive modeling for construction materials.

Contribution to understanding compressive strength

This study not only demonstrates the applicability of ET in CCS prediction but also provides a framework for evaluating and understanding the underlying relationships between input features and material properties. By revealing the strengths of ensemble methods in handling nonlinearity, feature interactions, and variability, it paves the way for more sophisticated approaches in modeling other material properties. Future research could explore integrating domain knowledge into feature engineering or hybrid models to further enhance predictive accuracy and interpretability.

By providing these insights, the findings not only validate the use of the Extra Trees Regressor as a robust predictive tool but also advance our understanding of the factors governing concrete compressive strength, offering potential pathways for optimization in material design and construction.

In summary, the results provide valuable guidance for selecting an appropriate regression model for predicting compressive strength in earthen concrete. While XGBoost stands out as a top-performing model, it's essential to consider various factors and trade-offs to make an informed decision based on the specific requirements and constraints of the problem at hand.

Table 4 shows a dataset with various input variables and output variables to predict compressive strength using machine learning. Here is a breakdown of the columns in our dataset.

Output variable target (Compressive strength Y1): Compressive strength of earth concrete this is a variable that we want to predict using machine learning, Label Prediction using ML means that the compressive strength is predicted using a machine learning model. (This column contains the actual predictions made by a model, We have a dataset with different samples (for example, samples 177, 69, 273, etc.) and the corresponding values for these entities and the target variable. Using the provided input features, we can create a machine learning model to predict compressive strength using this dataset. The "FIBER TYPE" column also suggests that the type of fiber used may be an important categorical characteristic in our predictive model. Please note that when we work with this dataset, it needs to be split into a training set and a testing set so that we can evaluate the effectiveness of our machine, learning model. Additionally, data preprocessing and feature engineering is required before training our model.

The following outcomes are displayed in the tables as performance measures for the Extra Trees Regressor regression model:

RMSE (root mean square error): 0.4909, R2 (R squared): 0.9444, MAE (mean absolute error): 0.1899, MSE (root mean square error): 0.2410 and RMSLE (root mean square logarithmic error): 0.0988 Regression models are frequently assessed for performance using these criteria. The model appears to have performed well in this instance. The model predictions are, on average, fairly close to the actual values, as indicated by the low MAE and MSE values.

The RMSE provides a measure of prediction error, and a value of 0.4909 suggests relatively small errors. The low RMSLE suggests that the model's predictions on a logarithmic scale are accurate It is less than or similar to previous studies^{8,22,25,26,29}. The high R2 value of 0.9444 indicates that the model explains a significant portion of the variance in the target variable and fits the data well. Overall, these results suggest that the Extra Trees Regressor model performed well in this context. However, it is essential to consider the specific problem and

State number	Input variables								Output variables	
	Sand% × 2	Silt% × 3	Cement % × 4	Fibre% × 5	Fibre length(mm) × 6	Fibre tensile strength(mpa) × 7	Specimen's age(days) × 8	Type of fibres	Compressive strength(mpa)y1	Prediction label using ml
177	20.000000	20.0	0.0	1.00	70.0	260.00	28.0	Banana fibre	7.45	5.66840
69	50.000000	20.0	4.0	2.00	35.0	29.58	14.0	Steel fiber	2.20	2.05020
273	35.000000	30.0	7.0	2.00	90.0	891.00	28.0	Bnana fibres	6.17	5.82030
125	92.000000	0.0	0.0	0.75	30.0	441.00	7.0	Green mussel shell	2.25	1.95380
24	50.000000	11.0	15.0	0.60	50.0	500.00	28.0	Alfa	4.90	...
230	46.000000	44.0	0.0	10.00	20.0	80.00	28.0	Rice husk ash	0.16	0.21210
308	45.000000	30.0	0.0	0.20	3.0	891.00	28.0	Hibisus cannabibus fibres	2.80	2.76950
217	27.540001	47.5	0.0	0.50	60.0	150.00	28.0	Wool	3.43	3.81620
18	50.000000	11.0	10.0	0.30	50.0	500.00	28.0	Alfa	2.40	2.36100
121	20.000000	58.0	5.0	4.00	9.0	766.00	28.0	Pineapple leaves fibres	3.31	3.59965

Table 4. The experimental results are compared with the projected compressive strength of concrete using machine learning algorithms.

Model	MAE	MSE	RMSE	RMSLE	R2
Extra trees regressor	0.1899	0.2410	0.4909	0.0988	0.9444

Table 5. Effectiveness of the RF technique for estimating concrete's compressive strength.

domain when interpreting these results, as different applications may have different requirements for model performance. See Table 5 above.

The results in Table 5 provide a comprehensive evaluation of the effectiveness of the Extra Trees Regressor approach in estimating compressive strength. They offer insights into the model's predictive accuracy, variability, and overall fit to the data, helping stakeholders make informed decisions about its suitability for the task at hand.

Comparative analysis with existing research

The manuscript briefly notes that the results obtained in this study are comparable to those of prior studies, but it lacks a detailed discussion that situates these findings within the broader context of research on concrete compressive strength (CCS) prediction. A comprehensive comparison with existing literature would significantly enhance the manuscript, underscoring the novel contributions and improvements this work offers.

Alignment with existing research

- Accuracy of Ensemble Models:

Previous studies have consistently demonstrated the superior performance of ensemble methods like Random Forest (RF) and Gradient Boosting Regressor (GBR) for CCS prediction. For instance,⁷³ reported an R^2 of 0.82 using RF, which aligns with the performance of the RF model in this study ($R^2=0.7780$). Similarly, the performance of Gradient Boosting ($R^2=0.7901$) is consistent with values reported in comparable works, reinforcing the reliability of ensemble methods.

- Adoption of Nonlinear Models:

Studies that employed nonlinear models such as XGBoost and Support Vector Machines (SVM) have reported robust performance metrics, particularly in datasets with complex interactions among features. This aligns with the findings in this study, where XGBoost achieved a strong R^2 of 0.8191, comparable to results reported by⁷⁴.

Distinctions and novel contributions:

- Extra Trees Regressor Performance:

This study identifies the Extra Trees Regressor (ET) as the best-performing model, achieving an R^2 of 0.9444. This significantly surpasses the performance metrics reported for ensemble models in prior research. The novelty lies in demonstrating ET's capability to not only match but exceed the predictive accuracy of more commonly used ensemble methods like RF and GBR. This is a notable contribution, as ET's potential in this domain has been underexplored.

- Feature Interaction Insights:

While earlier works have highlighted the importance of individual features (e.g., cement content or water-to-cement ratio) in CCS prediction, this study emphasizes the interactions between multiple factors, such as fiber content and cement proportion. The high correlation coefficient (0.9444) between these features, combined with ET's ability to capture their nonlinear interplay, offers deeper insights into the determinants of CCS.

- Automated Machine Learning (AutoML):

The integration of PyCaret to streamline and optimize the model selection and evaluation process is a novel approach in this context. While prior studies often rely on manually tuned models, this study demonstrates the advantages of leveraging automated tools to enhance efficiency and reproducibility.

Improvements over past studies

- Enhanced Predictive Metrics:

The R^2 of 0.9444 achieved by ET in this study is among the highest reported in the literature for CCS prediction, indicating a significant improvement in model precision and reliability.

- Reduction in Computational Complexity:

The use of automated hyperparameter tuning via PyCaret and genetic algorithms (GA) reduces the time and effort required for model development compared to traditional manual approaches.

- Broader Applicability:

By comparing multiple algorithms and emphasizing feature importance, this study provides a more comprehensive framework for CCS prediction, applicable to various concrete formulations and experimental conditions.

Contextualizing the significance

By situating these findings alongside existing research, this study highlights its contributions to the growing field of machine learning applications in civil engineering. The demonstration of ET's superior performance, coupled with the methodological innovations in AutoML, underscores the potential for further advancements in predictive modeling. Such a detailed comparison not only validates the findings but also positions this work as a valuable reference for future research on CCS prediction.

A key component of developing predictive models such as those that forecast concrete's compressive strength, is feature importance. Selecting features and eventually interpreting the model more effectively can be aided by knowing which features—also referred to as variables or predictors—have the greatest influence on the goal variable or compressive strength. The following stages outline how to assess the significance of a feature in order to forecast the compressive strength of crushed earth blocks (Fig. 5).

Figure 5 likely illustrates the process of determining feature importance in predicting compressive strength of compacted earth blocks. Here's an elaboration on each step depicted in Fig. 5:

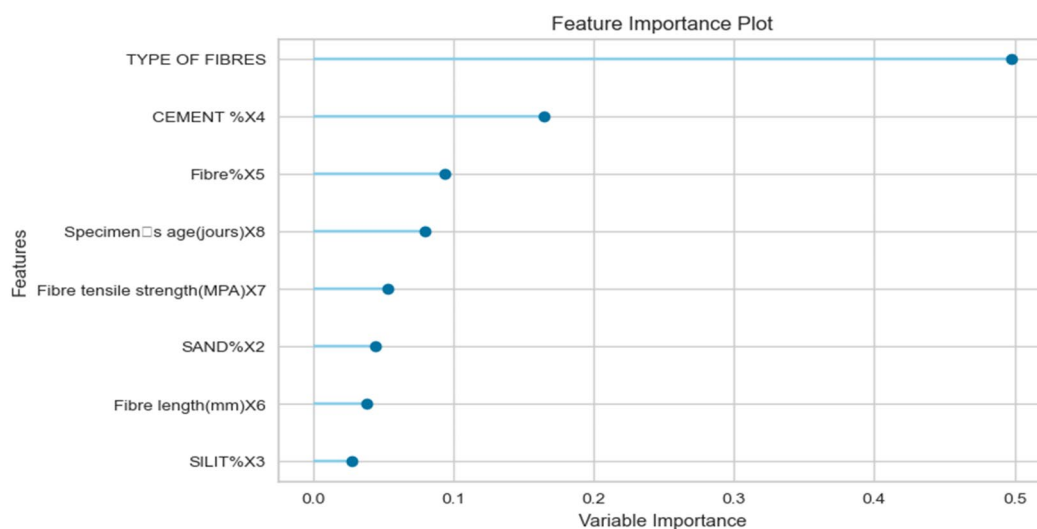


Fig. 5. Significance of a characteristic in estimating compacted earth block compressive strength.

- **Feature Selection** The process starts with selecting a set of features or characteristics of compacted earth blocks that are believed to influence compressive strength. These features can include material properties (e.g., particle size distribution, clay content), environmental factors (e.g., moisture content, curing conditions), and any other relevant variables.
- **Data Collection** Once the features are identified, data is collected from various sources or experiments. This data includes measurements or observations of the selected features along with corresponding compressive strength values of the compacted earth blocks.
- **Data Preprocessing** The collected data is preprocessed to ensure its quality and suitability for analysis. This involves tasks such as cleaning the data to remove errors or inconsistencies, handling missing values, and standardizing or normalizing the features to make them comparable.
- **Feature Importance Analysis** With the preprocessed data, feature importance analysis is conducted to quantify the relative importance of each feature in predicting compressive strength. Several techniques can be used for this analysis, including:
 - **Correlation Analysis** Calculate correlation coefficients between each feature and compressive strength to identify linear relationships.
 - **Tree-based Models** Train ensemble models such as Random Forest or Gradient Boosting, which naturally provide feature importance scores based on the reduction in prediction error achieved by each feature.
 - **Permutation Importance** Perturb the values of each feature and measure the impact on model performance to assess feature importance.
- **Visualization** The results of feature importance analysis are visualized for easy interpretation. This could involve plotting feature importance scores as bar charts, heatmaps, or other graphical representations. Visualization helps stakeholders understand which features have the most significant impact on compressive strength.
- **Interpretation and Decision-making** Stakeholders interpret the results of feature importance analysis to make informed decisions. They identify the most influential features that should be prioritized in predictive modeling efforts. This step may also involve consulting domain experts to validate the findings and ensure their relevance in practice.
- **Model Building** Based on the identified important features, predictive models are built using machine learning algorithms. These models use the selected features to predict compressive strength of compacted earth blocks.
- **Model Evaluation and Iteration** The predictive models are evaluated using validation data to assess their performance. If necessary, the feature selection process and model building may be iterated upon to improve prediction accuracy.

A number of variables, such as the production process and the composition of the earth mixture, affect the compressive strength of compressed earth blocks (CEBs). The compressive strength of CEBs can be greatly impacted by various materials and their ratios. The following are important things to keep in mind while evaluating how various materials affect the compressive strength of CEBs. Plot histograms for each variable's impact on concrete's compressive strength are displayed in Fig. 6.

the insights gained from Fig. 6 can guide construction practices towards producing high-quality, durable, and cost-effective compressed earth blocks suitable for a wide range of construction applications.

The heatmap would show the correlation coefficients between all the variables in the dataset as a color scale, the association is stronger the darker the color, the association is less the lighter the color, the diagonal of the heatmap would show the correlation coefficients of each variable with itself. These correlation coefficients are always 1^{73,87–89}.

The key points of the heatmap would be:

- The variables that are most correlated are those that are located closest to each other on the heatmap.
- The variables that are least correlated are those that are located farthest apart on the heatmap.
- The diagonal of the heatmap highlights the self-correlation of each variable, with coefficients of 1. (See Fig. 7).

An effective tool for comprehending the intricate correlations between the various variables used to estimate compressive strength is the correlation heatmap. Through the application of these insights, practitioners can enhance the precision and dependability of predictive models for compressive strength estimation by making well-informed decisions during the model building, feature selection, and data analysis phases^{90,91,93,94}.

With this study, we can arrive at the following most important and striking results:

Cement plays a very important role in increasing the compressive strength of (CEB), and for the result to be ideal, the percentage of cement must be limited to 10 to 15% of the mass of (CEB). Fibers also play a very important role in cohesion of (CEB) and increasing compressive strength, but artificial fibers remain better than natural fibers because artificial fibers are durable and last for decades, while natural fibers have a long lifespan. Life limit. However, natural fibers remain the most used due to their availability and their good^{74–76}, cheaper price. And the best percentage for reinforcing concrete with fibers is limited to between 1 and 2%. Above this percentage, fiber can play a more negative than positive role. The components of the concrete floor should also be as follows: clay 5 to 25%, sand between 50 to 75%, silts between 5 to 15% it's given proportions to obtain an ideal earth concrete. We can represent this data in a pyramid according to the priority and the amount of effect in the earth concrete. The most important effect is the highest. The results are briefly represented in the Fig. 8 below.

The priority pyramid depicted in Fig. 8 offers a structured framework for guiding decisions regarding the ideal ratios of components in earth concrete to achieve optimal compressive strength. Here's how the priority pyramid aids in this process :

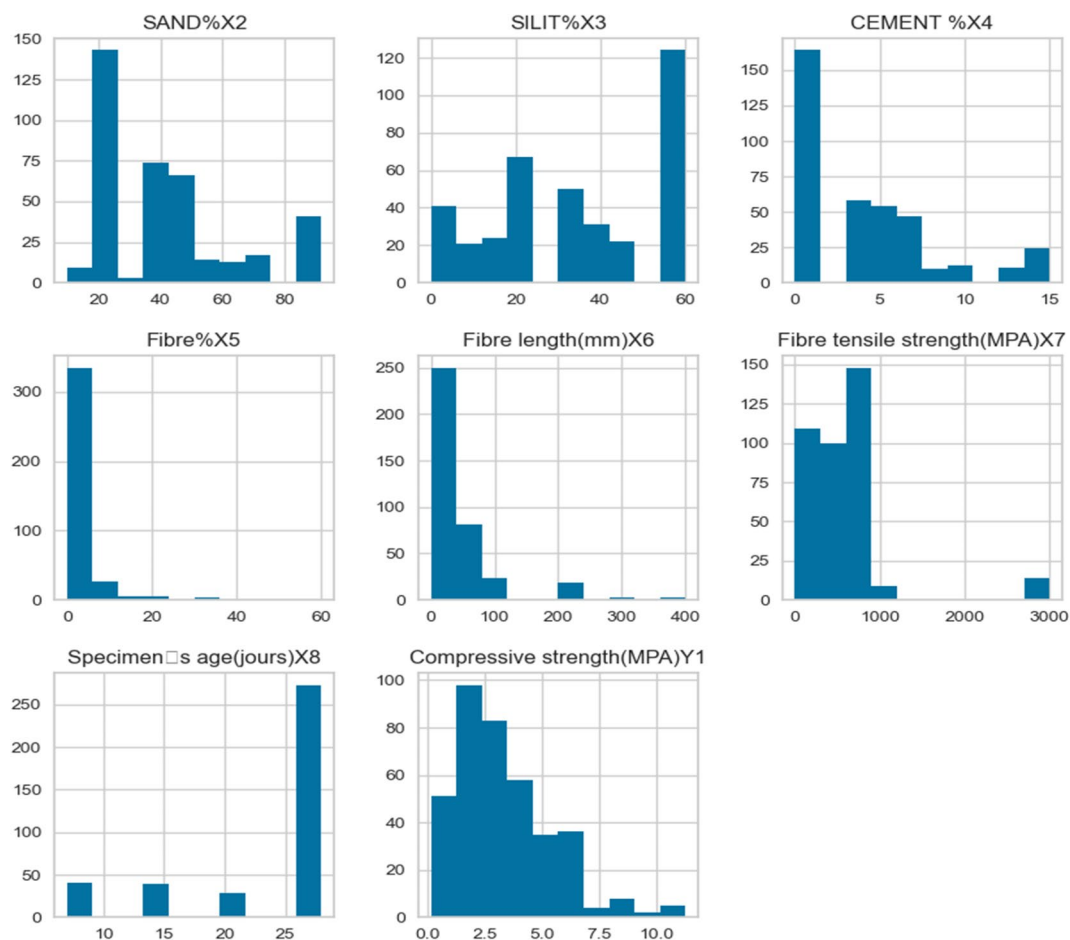


Fig. 6. Impact of various materials on crushed earth block's compressive strength.

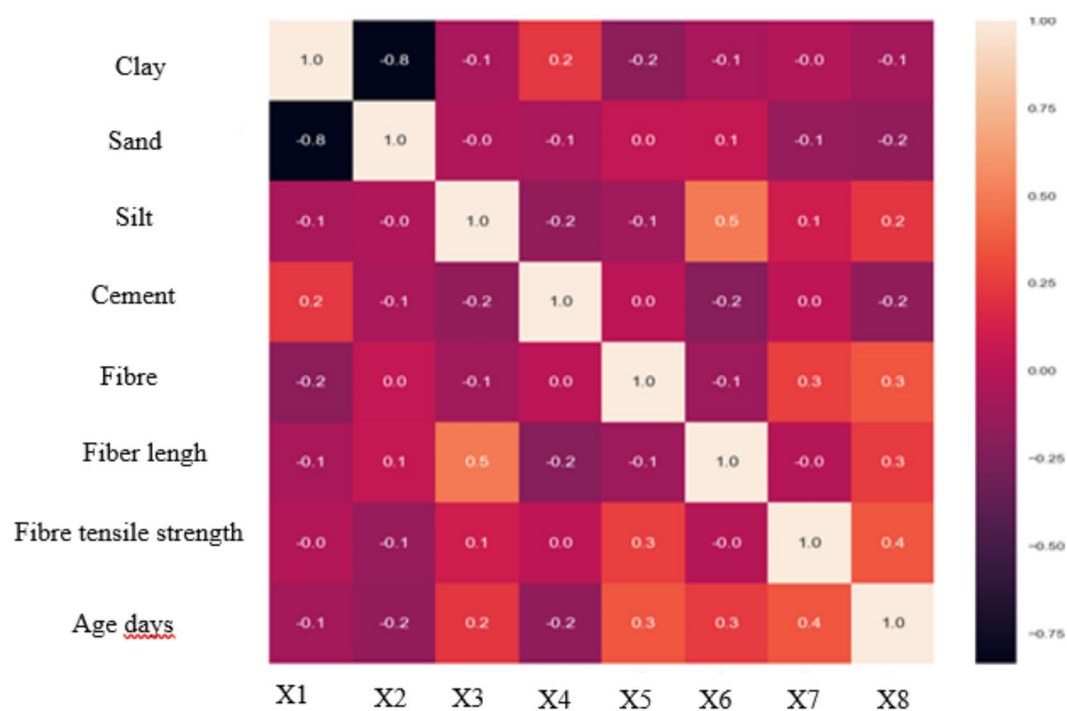


Fig. 7. Correlation heatmap of dataset variables used on predicting compressive strength of concrete.

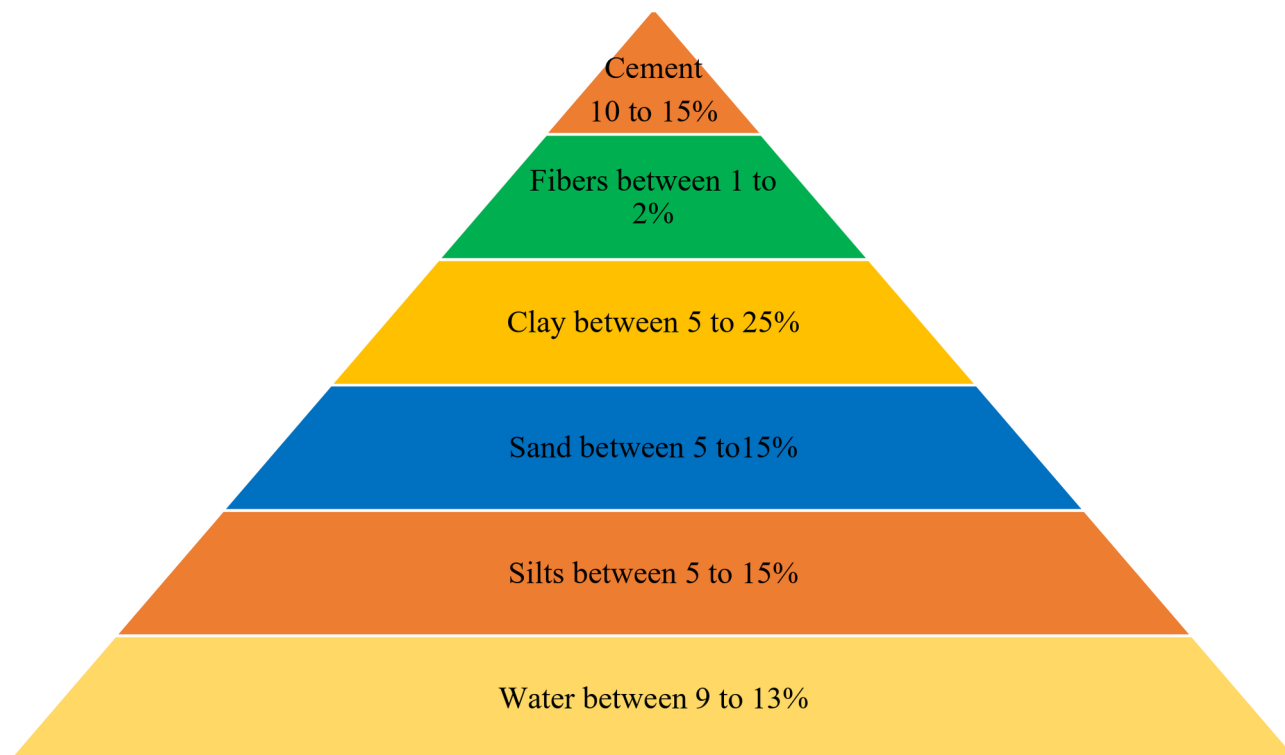


Fig. 8. Priority pyramid and proposed ideal ratios for each component.

- *Visual Hierarchy* The pyramid visually communicates the hierarchy of factors influencing compressive strength, with the most critical factors positioned at the top and less important factors towards the bottom. This visual hierarchy helps stakeholders quickly identify and prioritize the most impactful factors in optimizing earth concrete composition.
- *Emphasis on Key Factors* By positioning cement content and fiber reinforcement at the top of the pyramid, the priority pyramid emphasizes their critical importance in determining compressive strength. This highlights the need to carefully consider these factors when designing earth concrete mixtures to ensure optimal performance.
- *Recommended Ranges* The pyramid provides recommended ranges for each component, such as cement content, fiber reinforcement percentage, and proportions of clay, sand, and silts. These ranges are based on the study's findings and represent the ideal balance between different factors for achieving optimal compressive strength.
- *Decision Support* The priority pyramid serves as a decision support tool for practitioners involved in earth construction projects. It helps them make informed decisions regarding material selection, mix design, and production processes by highlighting the factors that have the most significant impact on compressive strength.
- *Optimization Strategy* By focusing on the top priorities identified in the pyramid, stakeholders can prioritize resources and efforts towards optimizing these critical factors. This targeted approach increases the likelihood of achieving desired compressive strength levels while minimizing potential drawbacks or inefficiencies.
- *Flexibility and Adaptability* While the priority pyramid provides recommended ranges for component ratios, it also allows for flexibility and adaptation based on specific project requirements, local conditions, and available resources. Stakeholders can adjust the ratios within the recommended ranges to suit their unique needs while still prioritizing the critical factors identified in the pyramid.

Overall, the priority pyramid in Fig. 8 serves as a valuable tool for guiding decisions regarding the ideal ratios of components in earth concrete for optimal compressive strength. It helps stakeholders focus their efforts on the most influential factors while providing flexibility for adaptation and customization to meet project-specific requirements.

In Fig. 8 above, the proportions optimal % content of each materials A were obtained through artificial AI algorithms Through analyzing the tables, figures and data obtained in the model used in this study.

Since the machine model showed great potential in determining the compressive strength without requiring laboratory experiments, we can conclude from this study that artificial intelligence has become indispensable in the field of civil engineering, particularly in determining the compressive strength of concrete. This conclusion is supported by previous research as well as this one. Notwithstanding all of these benefits, there are still gaps in the laws and policies governing the application of AI in civil engineering.

Selecting appropriate input variables is crucial for predictive modeling of compressive strength in clay concrete (or any material). Each variable chosen should contribute meaningfully to the model's ability to accurately predict compressive strength. Here's how the chosen variables—age, fiber percentage, fiber length, sand, silt, cement, and fiber tensile strength—contribute to the accuracy and reliability of the PyCaret model:

Importance of selected variables:

Age:

Significance The age of the concrete can affect its compressive strength due to ongoing hydration processes and curing effects.

Contribution Including age as a variable helps the model account for changes in strength over time, which is critical for predicting long-term strength characteristics.

Fiber Percentage and Fiber Length:

Significance Fibers (such as synthetic or natural fibers) are often added to concrete to improve its tensile strength and toughness.

Contribution Both fiber percentage and length directly influence the mechanical properties of the concrete, affecting its compressive strength. Models that include these variables can better predict the strength enhancement provided by fibers.

Sand and Silt Content:

Significance Sand and silt content affect the overall composition and workability of concrete.

Contribution Higher sand content generally improves workability but can reduce strength if not properly balanced with other components like cement. Silt content influences the compactness and permeability of concrete, which indirectly affects strength.

Cement Content:

Significance Cement is a primary binder in concrete and significantly impacts its strength.

Contribution Higher cement content generally leads to higher compressive strength, but excessive cement may lead to shrinkage and cracking. Properly quantifying cement content helps the model accurately predict strength variations.

Fiber Tensile Strength:

Significance The tensile strength of fibers determines their effectiveness in reinforcing concrete.

Contribution Stronger fibers contribute more effectively to the tensile strength and crack resistance of concrete, indirectly influencing compressive strength. Models incorporating fiber tensile strength can assess the optimal fiber type and amount for desired strength characteristics.

Contribution to accuracy and reliability of PyCaret model:

Model Training PyCaret utilizes these variables to train predictive models that capture the complex relationships between input factors and compressive strength.

Feature Importance PyCaret's feature importance analysis identifies which variables have the most significant impact on the model's predictions. This helps in understanding how each variable contributes to the overall accuracy of the model.

Model Evaluation By including these variables, PyCaret enables comprehensive evaluation metrics (like R2, RMSE, MAE) to assess how well the model predicts compressive strength across different scenarios and datasets.

Optimization Variables such as fiber percentage, length, and tensile strength allow PyCaret to optimize model performance by determining the ideal conditions for maximizing compressive strength while balancing other properties like workability and durability.

The development of precise and dependable models for the predictive modeling of compressive strength in clay concrete depends critically on the suitable selection of input factors. The age, fiber qualities, sand, silt, and cement content are among the variables that are selected, and each one is important in influencing the mechanical and structural features of concrete. It is ensured that the produced models are not only accurate but also capable of offering practical insights into optimizing concrete mix designs for particular strength requirements by PyCaret's ability to incorporate these variables and assess their influence on model predictions. Therefore, thorough evaluation of these factors improves the predictive models produced using PyCaret for compressive strength in clay concrete applications in terms of both accuracy and dependability.

Each technique—PyCaret's interactive explanation tool, SHAP, and LIME—has its strengths and suitability depending on the interpretability needs of the application. PyCaret's tool is convenient for users within its ecosystem, while SHAP and LIME offer broader model-agnostic capabilities for detailed interpretability needs in various contexts. The choice often depends on the specific goals of interpretability and the complexity of the model and data being analyzed.

Study implications

Improved Compressive Strength: The study offers important recommendations for improving the structural performance of earth concrete by identifying the variables that most affect compressive strength, such as cement content, fiber reinforcing, and material proportions. Elevating these variables can result in increased compressive strength, enhancing the longevity and load-carrying capability of earth-based constructions.

Enhanced Compressive Strength: The study offers helpful advice for improving the structural performance of earth concrete by identifying the most important parameters that affect compressive strength, such as cement content, fiber reinforcing, and material proportions. By maximizing these variables, earth-based constructions' durability and load-bearing capacity can be increased through higher compressive strength⁷⁴.

Although there are drawbacks to using a small dataset to predict the compressive strength of clay concrete, such as restricted data representation and the possibility of overfitting, these drawbacks can be avoided by applying regularization strategies, ensemble methods, data augmentation, and robust validation techniques with consideration. These mitigating techniques can be used to create machine learning models that perform well in terms of generalization, prediction accuracy, and dependability—all while requiring a minimal amount of training data. It is imperative to acknowledge the underlying limitations of the dataset and evaluate the model's predictions in light of these constraints.

Feature engineering plays a pivotal role in enhancing the performance of machine learning models by transforming raw data into more informative features that better capture the underlying relationships and patterns in the data. In the context of predicting compressive strength in clay concrete using PyCaret models,

Cost-Effectiveness By striking a balance between performance needs and cost considerations, the study's recommendations for ideal material compositions and proportions can aid in the optimization of construction processes. For instance, construction projects can reach required strength levels while minimizing material costs by defining the appropriate ranges for the quantities of clay, sand, and silt.

Sustainability The study's recommended ideal material compositions and proportions can help promote sustainable building techniques. It is possible to make earth concrete building more resource- and environmentally-friendly by using locally accessible resources to the fullest extent possible and reducing the usage of cement and other large-impact products.

Standardization and Quality Control The results of the study offer a foundation for uniform material requirements and quality assurance procedures in the building of earth concrete structures. Earth-based structure manufacturing can be made consistent and reliable by setting explicit rules for material selection and proportioning in construction methods.

Making Well-Informed Decisions: The study provides construction professionals with important knowledge on the performance traits of various material compositions and proportions. Better results are achieved overall when this knowledge is used to inform decisions made during the design, building, and maintenance stages of earth concrete structures.

All things considered; the study's conclusions provide useful recommendations for enhancing building procedures in earth concrete applications. Construction projects can achieve enhanced performance, cost-effectiveness, sustainability, and risk mitigation in the building of earth-based structures by putting into practice the suggested material compositions and proportions.

Overall, the study's conclusions show how machine learning more specifically, the Extra Trees Regressor approach can improve building materials by facilitating precise forecasting, material composition optimization, effective design iterations, identification of crucial variables, and generalizability of insights. Construction professionals may create more long-lasting, affordable, and easily available building materials that satisfy the changing demands of the built environment by utilizing machine learning.

Predictive modeling of compressive strength in clay concrete offers substantial benefits to civil engineering and construction by optimizing mix designs, improving structural integrity, and advancing construction practices. By leveraging these insights, stakeholders can achieve sustainable, resilient infrastructure that meets both current and future demands effectively.

Summary and conclusions

This study demonstrated the utility of machine learning, particularly the Extra Trees Regressor, in predicting the compressive strength of Compressed Earth Blocks (CEB). The research employed PyCaret to streamline the development and evaluation of machine learning models, emphasizing the importance of data quality and preprocessing in achieving reliable results.

- The predictive model enables practitioners to identify optimal proportions of cement, fibers, and other components for maximizing compressive strength. By simulating various material combinations, the approach reduces trial-and-error methods, saving time and resources in material formulation.
- The model's high accuracy ($R^2 = 0.9444$, $RMSE = 0.4909$) suggests it can be a reliable alternative to extensive laboratory testing. Practitioners can use the model as a preliminary assessment tool to evaluate CEB performance under different formulations, reserving laboratory experiments for final validation.
- By incorporating the model into digital workflows, such as Building Information Modeling (BIM) systems, engineers can integrate compressive strength predictions into project planning. This capability supports decision-making in large-scale construction projects where rapid material assessment is critical.
- The study relied on a specific dataset that may not fully capture the variability in soil compositions, environmental conditions, and curing practices across regions. A broader dataset encompassing diverse scenarios is necessary to improve the generalizability of the model.

- Real-world applications involve dynamic conditions such as fluctuating moisture levels, temperature variations, and long-term durability concerns, which were not fully addressed in this study. Incorporating these factors into future models will enhance their practical relevance.
- Future studies should integrate environmental factors, curing durations, and soil diversity to improve model robustness. Expanding the dataset to include these parameters would enhance predictive accuracy and generalizability.
- Combining machine learning models with physical simulations or multi-objective optimization techniques can provide a deeper understanding of material behavior and offer multi-faceted solutions.
- While the model demonstrates strong performance in experimental datasets, field validation under real-world conditions is necessary to confirm its applicability.
- Future work should explore how the model can optimize CEB composition to minimize environmental impact, such as reducing cement usage while maintaining performance.

Data availability

The datasets used and/or analyzed during the current study are available from co-author Dr. Abdelaziz Rabehi (rab_ghi@hotmail.fr) on reasonable request.

Received: 21 October 2024; Accepted: 5 February 2025

Published online: 11 February 2025

References

1. Chaabene, W. B., Flah, M. & Nehdi, M. L. Machine learning prediction of mechanical properties of concrete: Critical review. *Constr. Build. Mater.* **260**, 119889 (2020).
2. Ramadan Suleiman, A. & Nehdi, M. L. Modeling self-healing of concrete using hybrid genetic algorithm–artificial neural network. *Materials* **10**(2), 135 (2017).
3. Castelli, M., Vanneschi, L. & Silva, S. Prediction of high performance concrete strength using genetic programming with geometric semantic genetic operators. *Expert Syst. Appl.* **40**(17), 6856–6862 (2013).
4. Fan, D. Q. et al. A new design approach of steel fibre reinforced ultra-high performance concrete composites: Experiments and modeling. *Cem. Concr. Compos.* **110**, 103597. <https://doi.org/10.1016/j.cemconcomp.2020.103597> (2020).
5. Deng, F. et al. Compressive strength prediction of recycled concrete based on deep learning. *Constr. Build. Mater.* **175**, 562–569 (2018).
6. Behnood, A. & Golafshani, E. M. Machine learning study of the mechanical properties of concretes containing waste foundry sand. *Constr. Build. Mater.* **243**, 118152 (2020).
7. Marani, A. & Nehdi, M. L. Machine learning prediction of compressive strength for phase change materials integrated cementitious composites. *Constr. Build. Mater.* **265**, 120286. <https://doi.org/10.1016/j.conbuildmat.2020.120286> (2020).
8. Han, Q., Gui, C., Xu, J. & Lacidogna, G. A generalized method to predict the compressive strength of high-performance concrete by improved random forest algorithm. *Constr. Build. Mater.* **226**, 734–742 (2019).
9. Mansour, W., Sakr, M. A., Seleemah, A. A., Tayeh, B. A. & Khalifa, T. M. Bond behavior between concrete and prefabricated ultra high-performance fiber-reinforced concrete (UHPFRC) plates. *Struct. Eng. Mech.* **81**(3), 305–316 (2022).
10. Zhu, Y. et al. Predicting the splitting tensile strength of recycled aggregate concrete using individual and ensemble machine learning approaches. *Crystals* **12**(5), 569 (2022).
11. Salem, N. M. & Deifalla, A. Evaluation of the strength of slab–column connections with FRPs using machine learning algorithms. *Polymers* **14**(8), 1517 (2022).
12. Ebid, A. & Deifalla, A. Using artificial intelligence techniques to predict punching shear capacity of lightweight concrete slabs. *Materials* **15**(8), 2732 (2022).
13. Rabehi, A., Guermoui, M. & Lalmi, D. Hybrid models for global solar radiation prediction: A case study. *Int. J. Ambient Energy* **41**(1), 31–40 (2020).
14. Guermoui, M., Gairaa, K., Rabehi, A., Djafer, D. & Benkaciali, S. Estimation of the daily global solar radiation based on the Gaussian process regression methodology in the Saharan climate. *Eur. Phys. J. Plus* **133**, 1–17 (2018).
15. Ahmad, S., Al-Kutti, W. A., Al-Amoudi, O. S. & Maslehuddin, M. Compliance criteria for quality concrete. *Constr. Build. Mater.* **22**(6), 1029–1036 (2008).
16. Guermoui, M., Rabehi, A., Gairaa, K. & Benkaciali, S. Support vector regression methodology for estimating global solar radiation in Algeria. *Eur. Phys. J. Plus* **133**, 1–9 (2018).
17. Guermoui, M., Abdelaziz, R., Gairaa, K., Djemoui, L. & Benkaciali, S. New temperature-based predicting model for global solar radiation using support vector regression. *Int. J. Ambient Energy* **43**(1), 1397–1407 (2022).
18. Khelifi, R., Guermoui, M., Rabehi, A. & Lalmi, D. Multi-step-ahead forecasting of daily solar radiation components in the Saharan climate. *Int. J. Ambient Energy* **41**(6), 707–715 (2020).
19. Guermoui, M. & Rabehi, A. Soft computing for solar radiation potential assessment in Algeria. *Int. J. Ambient Energy* **41**(13), 1524–1533 (2020).
20. Barkhordari, M. S., Armaghani, D.J., Mohammed, A.S., Ulrikh, D. V. Data-driven compressive strength prediction of fly ash concrete using ensemble learner algorithms. *Buildings*. **12** (2), <https://doi.org/10.3390/buildings12020132>. (2022).
21. Biswas, R., Rai, B., Samui, P. & Roy, S. S. Estimating concrete compressive strength using MARS, LSSVM and GP. *Eng. J.* **24**(2), 41–52 (2020).
22. Khelifi, R. et al. Short-term PV power forecasting using a hybrid TVF-EMD-ELM strategy. *Int. Trans. Electr. Energy Syst.* **2023**(1), 6413716 (2023).
23. Guermoui, M., Rabehi, A., Benkaciali, S. & Djafer, D. Daily global solar radiation modelling using multi-layer perceptron neural networks in semi-arid region. *Leonardo Electron. J. Pract. Technol.* **28**, 35–46 (2016).
24. DeRousseau, M. A., Laftchiev, E., Kasprzyk, J. R., Rajagopalan, B. & Srubar, W. V. III. A comparison of machine learning methods for predicting the compressive strength of field-placed concrete. *Constr. Build. Mater.* **228**, 116661 (2019).
25. Rabehi, A., Guermoui, M., Khelifi, R. & Mekhalif, M. L. Decomposing global solar radiation into its diffuse and direct normal radiation. *Int. J. Ambient Energy* **41**(7), 738–743 (2020).
26. Kaloop, M. R., Kumar, D., Samui, P., Hu, J. W. & Kim, D. Compressive strength prediction of high-performance concrete using gradient tree boosting machine. *Constr. Build. Mater.* **264**, 120198 (2020).
27. Guermoui, M., Boland, J. & Rabehi, A. On the use of BRL model for daily and hourly solar radiation components assessment in a semi-arid climate. *Eur. Phys. J. Plus* **135**(2), 1–16 (2020).
28. Rabehi, A., Guermoui, M., Djafer, D. & Zaiani, M. Radial basis function neural networks model to estimate global solar radiation in semi-arid area. *Leonardo Electron. J. Pract. Technol.* **27**, 177–184 (2015).

29. Pham, A.-D., Ngo, N.-T., Nguyen, Q.-T. & Truong, N.-S. Hybrid machine learning for predicting strength of sustainable concrete. *Soft Comput.* **24**(19), 14965–14980 (2020).
30. Guermoui, M. et al. An analysis of case studies for advancing photovoltaic power forecasting through multi-scale fusion techniques. *Sci. Rep.* **14**(1), 6653 (2024).
31. Velay-Lizancos, M., Perez-Ordoñez, J. L., Martínez-Lage, I. & Vazquez-Burgo, P. Analytical and genetic programming model of compressive strength of eco concretes by NDT according to curing temperature. *Constr. Build. Mater.* **144**, 195–206 (2017).
32. Rabehi, A., Rabehi, A. & Guermoui, M. Evaluation of different models for global solar radiation components assessment. *Appl. Solar Energy* **57**, 81–92 (2021).
33. Zhang, L., Xiao, N., Yang, W. & Li, J. Advanced heterogeneous feature fusion machine learning models and algorithms for improving indoor localization. *Sensors* **19**(1), 125 (2019).
34. Al-Jamimi, H. A., Bagudu, A. & Saleh, T. A. An intelligent approach for the modeling and experimental optimization of molecular hydrodesulfurization over AlMoCoBi catalyst. *J. Mol. Liq.* **278**, 376–384. <https://doi.org/10.1016/j.molliq.2018.12.144> (2019).
35. Al-Jamimi, H. A., BinMakhashen, G. M., Saleh, T. A. Multiobjectives optimization in petroleum refinery catalytic desulfurization using Machine learning approach. <https://doi.org/10.1016/j.fuel.2022.124088>. (2022).
36. El-Amarty, N. et al. A new evolutionary forest model via incremental tree selection for short-term global solar irradiance forecasting under six various climatic zones. *Energy Convers. Manag.* **310**, 118471 (2024).
37. Asteris, P. G., Rizal, F. I., Koopialipoor, M., Roussis, P. C., Ferentinou, M., Armaghani, D. J., Gordan, B. Slope stability classification under seismic conditions using several tree-based intelligent techniques. *Appl. Sci. (Switz.)* **12** (3) <https://doi.org/10.3390/app12031753>. (2022).
38. Bouchakour, A. et al. MPPT algorithm based on metaheuristic techniques (PSO & GA) dedicated to improve wind energy water pumping system performance. *Sci. Rep.* **14**(1), 17891 (2024).
39. Ladjal, B., Tibermacine, I. E., Bechouat, M., Sedraoui, M., Napoli, C., Rabehi, A., & Lalmi, D. Hybrid models for direct normal irradiance forecasting: A case study of Ghardaia zone (Algeria). *Nat. Hazards*. 1–23. (2024)
40. Teta, A. et al. Fault detection and diagnosis of grid-connected photovoltaic systems using energy valley optimizer based lightweight CNN and wavelet transform. *Sci. Rep.* **14**(1), 18907 (2024).
41. Houcine, B., Mohamed, R., Samir, K. & Sarra, B. Artificial intelligence for the prediction of the physical and mechanical properties of a compressed earth reinforced by fibers. *J. Eng. Exact Sci.* **9**(4), 15910–16001. <https://doi.org/10.18540/jcecv19iss4pp15910-01e> (2023).
42. Erdogmus, E., Donkor, P., Obonyo, E., Matta, Fabio. Effect of polypropylene fiber length on the flexural and compressive strength of compressed stabilized earth blocks. <https://doi.org/10.1061/9780784413517.068>. (2014)
43. Galán-Marín, C. & Rivera-Gómez, C. & Bradley, Fiona. Ultrasonic, molecular and mechanical testing diagnostics in natural fibre reinforced, polymer-stabilized earth blocks. *Int. J. Polym. Sci.* <https://doi.org/10.1155/2013/130582>. (2013).
44. Mostafa, M. & Uddin, N. Effect of banana fibers on the compressive and flexural strength of compressed earth blocks. *Buildings* **5**, 282–296. <https://doi.org/10.3390/buildings5010282> (2015).
45. Khedari, J., Watsanasathaporn, P. & Hirunlabh, J. Development of fibre-based soil–cement block with low thermal conductivity. *Cem. Concr. Compos.* **27**, 111–116. <https://doi.org/10.1016/j.cemconcomp.2004.02.042> (2005).
46. Chan, C.-M. Effect of natural fibres inclusion in clay bricks: Physico-mechanical properties. *World Acad. Sci. Eng. Technol.* **73**, 51–57 (2011).
47. Millogo, Y., Morel, J. C., Aubert, J. E. & Ghavami, K. Experimental analysis of pressed adobe blocks reinforced with Hibiscus cannabinus fibers. *Constr. Build. Mater.* **52**, 71–78. <https://doi.org/10.1016/j.conbuildmat.2013.10.094> (2013).
48. Elenga, R., Mabiala, B., Ahouet, L., Goma-Maniongui, J. & Dirras, G. Characterization of clayey soils from Congo and physical properties of their compressed earth blocks reinforced with post-consumer plastic wastes. *Geomaterials*. **1**. <https://doi.org/10.4236/gm.2011.13013>. (2011).
49. Houcine, B., Mohamed, R., Samir, K. & Sarra, B. Valorization of plastic waste in concrete for sustainable development. *J. Eng. Exact Sci.* **9**, 16009–16101. <https://doi.org/10.18540/jcecv19iss5pp16009-01e> (2023).
50. Galán-Marín, C., Rivera-Gómez, C. & Bradley, F. The mechanical properties and molecular bonding characteristics of clay-based natural composites reinforced with animal fibres. *J. Biobased Mater. Bioenergy* **7**, 143–151. <https://doi.org/10.1166/jbmb.2013.1269> (2013).
51. Prasad, C. K. S. Plastic fibre reinforced soil blocks as a sustainable building material. *Int. J. Adv. Res. Technol.* **1** (5) (2012).
52. Demir, I. An investigation on the production of construction brick with processed waste tea. *Build. Environ. BLDG ENVIRON* **41**, 1274–1278. <https://doi.org/10.1016/j.buildenv.2005.05.004> (2006).
53. Namango, S. S. Development of cost-effective earthen building material for housing wall construction: Investigations into the properties of compressed earth blocks stabilized with sisal vegetable fibres, cassava powder and cement compositions. 17.07.2006 (2006).
54. Aksogan, O., Bakbak, D., Kaplan, H. & Işık, A. Sound insulation of fibre reinforced mud brick walls. *Constr. Build. Mater.* **23**, 1035–1041. <https://doi.org/10.1016/j.conbuildmat.2008.05.008> (2009).
55. Demir, I. Effect of organic residues addition on the technological properties of clay bricks. *Waste Manag.* **28**(2008), 622–627. <https://doi.org/10.1016/j.wasman.2007.03.019> (2007).
56. Calatan, G. Determining the optimum addition of vegetable materials in adobe bricks. *Procedia Technol.* **22**, 259–265. <https://doi.org/10.1016/j.protcy.2016.01.077> (2016).
57. Obonyo, E. Optimizing the physical, mechanical and hygrothermal performance of compressed earth bricks. *Sustainability*. **3**(4), 596–604. <https://doi.org/10.3390/su3040596> (2011).
58. Basha, E. A., Hashim, R., Mahmud, H. B. & Muntohar, A. Stabilization of residual soil with rice husk ash and cement. *Constr. Build. Mater.* **19**, 448–453. <https://doi.org/10.1016/j.conbuildmat.2004.08.001> (2005).
59. Bouhicha, M., Aouissi, F. & Kenai, S. Performance of composite soil reinforced with barley straw. *Cem. Concr. Compos.* **27**, 617–621. <https://doi.org/10.1016/j.cemconcomp.2004.09.013> (2005).
60. Eko, R. M., Offa, E. D., Ngatcha, T. Y. & Minsili, L. S. Potential of salvaged steel fibers for reinforcement of unfired earth blocks. *Constr. Build. Mater.* **35**, 340–346. <https://doi.org/10.1016/j.conbuildmat.2011.11.050> (2012).
61. Donkor, P., Obonyo, E. & Ferraro, C. Fiber reinforced compressed earth blocks: Evaluating flexural strength characteristics using short flexural beams. *Materials* **14**, 6906. <https://doi.org/10.3390/ma14226906> (2021).
62. Mostafa, M. Experimental analysis of Compressed Earth Block (CEB) with banana fibres resisting flexural and compression forces. *Case Stud. Constr. Mater.* **5**, 53–63. <https://doi.org/10.1016/j.cscm.2016.07.001> (2016).
63. Türkmen, İ., Ekinçi, E., Kantarcı, F., Sarıcı, T. The mechanical and physical properties of unfired earth bricks stabilized with gypsum and Elaziğ Ferrochrome Slag. *Int. J. Sustain. Built Environ.* **6**. <https://doi.org/10.1016/j.ijsbe.2017.12.003>. (2017).
64. Babé, C. et al. Effect of neem (Azadirachta Indica) fibers on mechanical, thermal and durability properties of adobe bricks. *Energy Rep.* **7**, 686–698. <https://doi.org/10.1016/j.egyrs.2021.07.085> (2021).
65. Limami, H., Manssouri, I., Cherkaoui, K., Khaldoun, A. Mechanical and physicochemical performances of reinforced unfired clay bricks with recycled Typha-fibers waste as a construction material additive. *Clean. Eng. Technol.* **2**. <https://doi.org/10.1016/j.clet.2020.100037>. (2020).
66. Kumar, N. & Barbato, M. Effects of sugarcane bagasse fibers on the properties of compressed and stabilized earth blocks. *Constr. Build. Mater.* **315**, 125552. <https://doi.org/10.1016/j.conbuildmat.2021.125552> (2022).

67. Cottrell, J. A., Ali, M., Tatari, A., Martinson, D. B. Effects of fibre moisture content on the mechanical properties of jute reinforced compressed earth composites. *Constr. Build. Mater.* **373** 130848. <https://doi.org/10.1016/j.conbuildmat.2023.130848>.
68. Vodounon, N. A., Kanali, C. & Mwero, J. Compressive and flexural strengths of cement stabilized earth bricks reinforced with treated and untreated pineapple leaves fibres. *Open J. Compos. Mater.* **8**, 145–160. <https://doi.org/10.4236/ojcm.2018.84012> (2018).
69. Lejano, B. Compressed earth blocks with powdered green mussel shell as partial binder and pig hair as fiber reinforcement. *Int. J. GEOMATE*. **16**. <https://doi.org/10.21660/2019.57.8138>. (2019).
70. Koutous, A. & Hilali, E. Reinforcing rammed earth with plant fibers: A case study. *Case Stud. Constr. Mater.* **14**, e00514. <https://doi.org/10.1016/j.cscm.2021.e00514> (2021).
71. Duc Chinh Ngo. Développement d'un nouveau éco-béton à base de sol et fibres végétales : étude du comportement mécanique et de durabilité. Autre. Université de Bordeaux, 2017. Français. NNT :2017BORD0885.
72. Mumuni, A. & Mumuni, F. Automated data processing and feature engineering for deep learning and big data applications: A survey. *J. Inf. Intell.* <https://doi.org/10.1016/j.jiixd.2024.01.002>.
73. Khatti, J. & Grover, K. Assessment of hydraulic conductivity of compacted clayey soil using artificial neural network: An investigation on structural and database multicollinearity. *Earth Sci. Inform.* **17**, 3287–3332. <https://doi.org/10.1007/s12145-024-01336-0> (2024).
74. Khatti, J. & Grover, K. S. Prediction of compaction parameters for fine-grained soil: Critical comparison of the deep learning and standalone models. *J. Rock Mech. Geotech. Eng.* **15**(11), 3010–3038. <https://doi.org/10.1016/j.jrmge.2022.12.034> (2023).
75. Khatti, J. & Grover, K. S. Estimation of intact rock uniaxial compressive strength using advanced machine learning. *Transp. Infrastruct. Geotechnol.* **11**(4), 1989–2022. <https://doi.org/10.1007/s40515-023-00357-4> (2024).
76. Kumar, M., Kumar, D. R., Khatti, J., Samui, P. & Grover, K. S. Prediction of bearing capacity of pile foundation using deep learning approaches. *Front. Struct. Civil Eng.* 1–17. <https://doi.org/10.1007/s11709-024-1085-z>. (2024).
77. Mahabub, M. S. et al. Assessing the effects of influencing parameters on field strength of soft coastal soil stabilized by deep mixing method. *Bull. Eng. Geol. Environ.* **83**, 9. <https://doi.org/10.1007/s10064-023-03502-y> (2024).
78. Upadhyay, M., Daiya, A., Khatti, J. A review on comparative study of stabilization of black cotton soil by natural and artificial fibre (October 1, 2019). in *Proceedings of International Conference on Advancements in Computing & Management (ICACM) 2019*, Available at SSRN: <https://ssrn.com/abstract=3462225> or <https://doi.org/10.2139/ssrn.3462225>
79. Upadhyay, M., Daiya, A., Khatti, J. A review of stabilization of black cotton soil by industrial waste materials (October 1, 2019). in *Proceedings of International Conference on Advancements in Computing & Management (ICACM) 2019*, Available at SSRN: <https://ssrn.com/abstract=3462219> or <https://doi.org/10.2139/ssrn.3462219>
80. Prado-Gil, J., Palencia, C., Jagadesh, P. & Martínez-García, R. A comparison of machine learning tools that model the splitting tensile strength of self-compacting recycled aggregate concrete. *Materials* **15**(12), 4164. <https://doi.org/10.3390/ma15124164> (2022).
81. Jagadesh, P., de Prado-Gil, J., Silva-Monteiro, N. & Martínez-García, R. Assessing the compressive strength of self-compacting concrete with recycled aggregates from mix ratio using machine learning approach. *J. Mater. Res. Technol.* **24**, 1483–1498. <https://doi.org/10.1016/j.jmrt.2023.03.037> (2023).
82. Harzallah, S., Rebhi, R., Chabaat, M. & Rabehi, A. Eddy current modelling using multi-layer perceptron neural networks for detecting surface cracks. *Frattura ed Integrità Strutturale* **12**(45), 147–155 (2018).
83. Lalmi, D. et al. Evaluation and estimation of the inside greenhouse temperature, numerical study with thermal and optical aspect. *Int. J. Ambient Energy* **42**(11), 1269–1280 (2021).
84. Rabehi, A., Helal, H., Zappa, D. & Comini, E. Advancements and prospects of electronic nose in various applications: A comprehensive review. *Appl. Sci.* **14**(11), 4506 (2024).
85. Hamdani, M., Youcefi, M., Rabehi, A., Nail, B. & Douara, A. Design and implementation of a medical telemonitoring system based on IoT. *Eng. Technol. Appl. Sci. Res.* **12**(4), 8949–8953 (2022).
86. Baitiche, O., Bendelala, F., Cheknane, A., Rabehi, A. & Comini, E. Numerical modeling of hybrid solar/thermal conversion efficiency enhanced by metamaterial light scattering for ultrathin PbS QDs-STPV cell. *Crystals* **14**(7), 668 (2024).
87. Khatti, J. & Grover, K. S. Prediction of uniaxial strength of rocks using relevance vector machine improved with dual kernels and metaheuristic algorithms. *Rock Mech. Rock Eng.* **57**, 6227–6258. <https://doi.org/10.1007/s00603-024-03849-y> (2024).
88. Khatti, J. & Grover, K. S. Assessment of the uniaxial compressive strength of intact rocks: An extended comparison between machine and advanced machine learning models. *Multiscale Multidiscip. Model. Exp. Des.* **7**, 3301–3325. <https://doi.org/10.1007/s41939-024-00408-4> (2024).
89. Khatti, J. & Grover, K. S. Assessment of uniaxial strength of rocks: A critical comparison between evolutionary and swarm optimized relevance vector machine models. *Transp. Infrastruct. Geotech.* **11**, 4098–4141. <https://doi.org/10.1007/s40515-024-00433-3> (2024).
90. Hosseini, S. et al. Assessment of the ground vibration during blasting in mining projects using different computational approaches. *Sci. Rep.* **13**, 18582. <https://doi.org/10.1038/s41598-023-46064-5> (2023).
91. Khatti, J. & Polat, B. Y. Assessment of short and long-term pozzolanic activity of natural pozzolans using machine learning approaches. *Structures* **68**, 107159. <https://doi.org/10.1016/j.istruc.2024.107159> (2024).
92. Calatan, G., Hegyi, A., Dico, C. & Szilagyi, H. Opportunities regarding the use of adobe-bricks within contemporary architecture. *Procedia Manuf.* **46**, 150–157. <https://doi.org/10.1016/j.promfg.2020.03.023> (2020).
93. Whig, P., Gupta, K. & Jiwani, N. A novel method for diabetes classification and prediction with Pycaret. *Microsyst. Technol.* **29**, 1479–1487. <https://doi.org/10.1007/s00542-023-05473-2> (2023).
94. Bentegri, H., Mohamed, R. & Kherfane, S. Predicting the compressive strength of ecological concrete made with PET granules using artificial neural networks (MATLAB). *Stud. Eng. Exact Sci.* **5**, 1413–1435. <https://doi.org/10.54021/seesv5n1-073> (2024).

Acknowledgements

Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2025R754), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

Author contributions

H.B. and M.R. conceptualized the study and designed the methodology. H.B. wrote the main manuscript text, and A.R. reviewed and edited the manuscript. M.R. and S.K. were responsible for data collection and formal analysis. S.K. also provided resources and validated the results. T.A.N. developed the software and implemented the model. A.R. and M.G. conducted data analysis and visualization. M.G. and A.A.A. contributed to manuscript review and editing. D.S.K. and A.A.A. curated the data and validated the findings (pending). M.M.E. provided formal analysis and resources, while E.S.E. conducted statistical analysis, validation, and manuscript editing and review. All authors reviewed the manuscript.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to A.R.

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

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

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

© The Author(s) 2025, corrected publication 2025