Sathiparan *et al.*
Journal of Engineering and Applied Science (2024) 71:187
https://doi.org/10.1186/s44147-024-00527-2

Journal of Engineering and Applied Science

RESEARCH

Open Access

Prediction of moisture content of cement-stabilized earth blocks using soil characteristics, cement content, and ultrasonic pulse velocity

Navaratnarajah Sathiparan[1*], R. A. N. S. Tharuka[1] and Pratheeba Jeyananthan[2]

*Correspondence: sakthi@eng.jfn.ac.lk

[1] Department of Civil Engineering, Faculty of Engineering, University of Jaffna, Kilinochchi, Sri Lanka
[2] Department of Computer Engineering, Faculty of Engineering, University of Jaffna, Jaffna, Sri Lanka

Abstract

This article investigates the importance of moisture content in cement-stabilized earth blocks (CSEBs) and explores methods for their prediction using machine learning. A key aspect of the research is the development of accurate moisture content prediction models. The study compares the performance of various machine learning models, and XGBoost emerges as the most promising model, demonstrating superior accuracy in predicting moisture content based on factors like soil properties, cement content, and ultrasonic pulse velocity (UPV). The study employs SHAP (SHapley Additive exPlanations) to understand how these features influence the model's predictions. UPV is the most significant factor affecting predicted moisture content, followed by cement content and soil properties like uniformity coefficient. Also, the study explores the possibility of using a reduced set of features for moisture content prediction. They demonstrate that a combination of UPV, cement content, and uniformity coefficient can achieve good accuracy, highlighting the potential for practical applications where obtaining all data points might be challenging.

Keywords: Cement-stabilized earth blocks, Moisture content, Machine learning, Ultrasonic pulse velocity, SHAP analysis

Introduction

Due to its inherent advantages, masonry construction remains a prevalent choice for residential structures. Firstly, masonry offers exceptional durability, withstanding extreme weather conditions and demonstrating longevity over centuries. This translates to superior structural integrity and stability, providing robust resistance against fire, wind, and seismic activity. Furthermore, the inherent thermal mass of masonry materials fosters energy efficiency by regulating indoor temperatures, thereby reducing energy demands for heating and cooling [1]. Two primary types of masonry construction are fired-clay brick and concrete block masonry. However, conventional masonry materials like fired clay bricks and concrete blocks also have limitations. Fired-clay brick production necessitates high kiln temperatures, leading to significant energy consumption and the release of pollutants. While less energy intensive than bricks, concrete block production still contributes to greenhouse gas emissions. The manufacturing processes for bricks and blocks are energy intensive, resulting in a high embodied energy content within the final products [2].

Cement-stabilized earth blocks (CSEBs) offer a promising alternative that addresses these environmental concerns. CSEB production utilizes local, readily available soil, eliminating the need for high-temperature firing or extensive processing. Due to the use of local materials and minimal processing, CSEBs boast a significantly lower embodied energy compared to fired bricks and concrete blocks [3]. Also, CSEBs primarily consist of soil, a naturally renewable resource, minimizing their environmental impact. Therefore, CSEBs present a viable and sustainable option for masonry construction [4]. They address the environmental limitations associated with traditional materials while maintaining the core benefits of durability, strength, and energy efficiency, making them a potentially attractive choice for contemporary building practices.

CSEBs made with cement and lateritic soil contain numerous air pockets or voids. These voids can be filled with water when saturated, with air when dry, and with a combination of air, water vapor, and liquid water at intermediate moisture levels [5]. This moisture content variation significantly affects the strength and durability of CSEBs [6]. Water absorption can cause the cement gel to expand as it hardens. This creates a larger gap between cement gel surfaces, weakening the bond between particles in the CSEB. Consequently, the blocks become more susceptible to breaking under lower external loads. When voids are filled with water, hydraulic pressure is created. This pressure pushes the particles apart, potentially causing cracks to propagate at much lower loads. This translates to a decrease in the compressive strength of the CSEB [7]. If CSEBs absorb water too quickly or in large quantities, there is a higher risk of leaks in masonry construction. Highly absorbent CSEBs might become saturated during rainy seasons and transfer moisture into the building interior. Given these potential drawbacks, it is crucial to understand how moisture affects CSEBs before using them for wall construction in wet climates. As mentioned earlier, the soil type used in the CSEB mix also plays a role. Soils with high clay content typically absorb more water than those dominated by sand.

Several studies have investigated the effects of environmental factors on historical masonry structures [8–10]. These studies consistently show increased moisture content is common in vertical masonry structures. Experimental research by Foraboschi and Vanin [10], Witzany et al. [11], and Amade et al. [12] found that water saturation can significantly reduce the compressive strength of masonry. This decrease can range from a few percent to over 20% compared to the strength of air-dry masonry. Sathiparan et al. [13] reported that the moisture effects on compressive strength for CSEBs. The study was conducted with CSEBs with 1:6 volume ratio of cement to soil, and compressive strength was measured for moisture conditions. Based on this, the authors proposed that the compressive strength (σ_c in MPa) at the particular moisture content (m in percentage) can be given by Eq. (1).

$$\sigma_c = -0.031\mathrm{m} + 5.171 \tag{1}$$

All these studies show that moisture in masonry blocks, especially in CSEBs, affects strength and durability. So, knowing the moisture content in CSEBs at any state helps predict how CSEBs might perform in different weather conditions, allowing for informed decisions about their suitability for a specific project.

By determining the moisture content of CSEBs, manufacturers and builders gain valuable insights into the material's current state and its potential performance. This information helps ensure stronger and more durable CSEBs, reduced risk of leaks, and improved overall quality control and building performance. So, overall, determining the moisture content of CSEBs is crucial in ensuring their structural integrity, long-term durability, and suitability for use in construction projects, particularly in wet climates.

There are several methods for measuring moisture content in masonry, each with its own advantages and limitations. Destructive methods involve taking a core sample of the masonry, drying it in a controlled oven environment, and then measuring the weight difference before and after drying. This provides the most accurate moisture content measurement but damages the structure and requires specialized equipment [14, 15]. Several nondestructive methods, such as microwave method, electrical impedance method, resistance pins, and probe method, are also used to measure the moisture content of the masonry. Destructive methods are generally less expensive but require sample preparation. Nondestructive methods can be more costly due to specialized equipment but offer the advantage of not damaging the structure [16, 17]. The most suitable method depends on the specific needs of the project. Factors such as required accuracy, accessibility, cost, and equipment availability determine the appropriate testing methods. Although nondestructive methods provide some correlation with moisture content, soil properties also have a significant effect on moisture content [18]. So, the present study uses machine learning techniques to predict moisture content based on a combination of soil properties, cement content, and ultrasonic pulse velocity.

Recent research highlights the growing popularity of machine learning algorithms for predicting the characteristics of building materials [19–21]. This is likely due to the complex nature of material properties, which are influenced by numerous variables. Machine learning offers a powerful solution for navigating this complexity, surpassing traditional methods [22, 23]. A wide range of machine learning techniques have been successfully applied to predict the various property building materials. Furthermore, the combination of nondestructive testing measurements with machine learning techniques has proven effective in predicting the properties of concrete [24–27], pervious concrete [28], masonry [29, 30], and timber ([31]). This approach offers a significant advantage by allowing material evaluation without causing damage. While nondestructive testing and machine learning techniques have been successfully combined for various materials, research on predicting the moisture content of masonry, especially CSEBs, using this approach remains limited.

This study aims to bridge this gap by utilizing machine learning techniques to predict the moisture content of CSEBs based on soil characteristics, cement content used for stabilization, and ultrasonic pulse velocity. This will facilitate the assessment of CSEB quality without causing damage to the structures. The study will involve the creation of CSEBs with 27 distinct combinations of cement-to-soil ratios (three ratios) and soil types (9 different soils). Subsequently, various machine learning techniques will explore the relationships between soil properties, cement content, ultrasonic pulse velocity, and moisture content of the CSEBs.

Methods

Materials used

This study utilized ordinary Portland cement (OPC) conforming to Sri Lanka Standard SLS855 (1989) as the binding agent. The OPC's chemical composition comprised 66.6% CaO, 20.6% SiO_2, 4.5% Al_2O_3, 3.6% Fe_2O_3, and 0.4% $K_2O + Na_2O$ (by weight). Additionally, the cement possessed a specific gravity of 3.15, a bulk density of 1362 kg/m^3, and a specific surface area of 325 m^2/kg, with an initial setting time of 140 min and a final setting time of 190 min. Nine lateritic soils were chosen for this experiment, all originating from the northern region of Sri Lanka. These soil samples were then analyzed for their physical characteristics and particle size distribution. The visualization of soils is presented in Fig. 1.

Mix design

This research explores the influence of soil characteristics, cement content, and ultrasonic pulse velocity on the prediction of moisture content of CSEBs. The study used the same cement-to-soil ratios observed at construction sites to achieve this. Blocks were produced with three different cement contents: 8%, 12%, and 16% of the soil weight. Maintaining consistent workability is crucial. Typically, a fixed water-to-cement ratio or a specific slump value is used. However, in this case, the varying particle size distribution of the local soil necessitated a unique approach. Despite clay and silt, the slump test yielded zero slump even with a higher water-to-cement ratio. Consequently, the water content was determined based on achieving the highest possible dry density for the cement-soil mixture. Table 1 details the specific mix proportions used for creating the stabilized earth blocks. Figure 1 shows the physical appearance of the blocks after 28 days of moisture curing. The size of specimens is 150-mm cubic. The mortar cubes made with soils 3 and 6 exhibit a noticeably darker color than the others. This is likely due to the use of red soil in their mixture, resulting in a distinct visual difference. The remaining mortar cubes display a light-yellow coloration.

Testing

Sieve analysis test

Sieve analysis plays a vital role in determining the particle size distribution of granular materials. This fundamental test, employed following ASTM D422 [32], offers valuable insights for various engineering and scientific applications. The procedure commences with obtaining a representative and dry sample. A series of sieves with progressively smaller mesh sizes are meticulously chosen based on the material and the desired level of detail. The sample is then poured onto the topmost sieve, and the entire stack is subjected to a controlled mechanical process lasting 10 min. Following the shaking, any material retained on each sieve is meticulously brushed back onto the stack, meticulously collected, and weighed individually. The mass of material retained on each sieve is then expressed as a percentage of the initial sample mass. This conversion provides a quantitative representation of the particle size distribution. The particle size distribution for each soil used in this study is presented in Fig. 2.

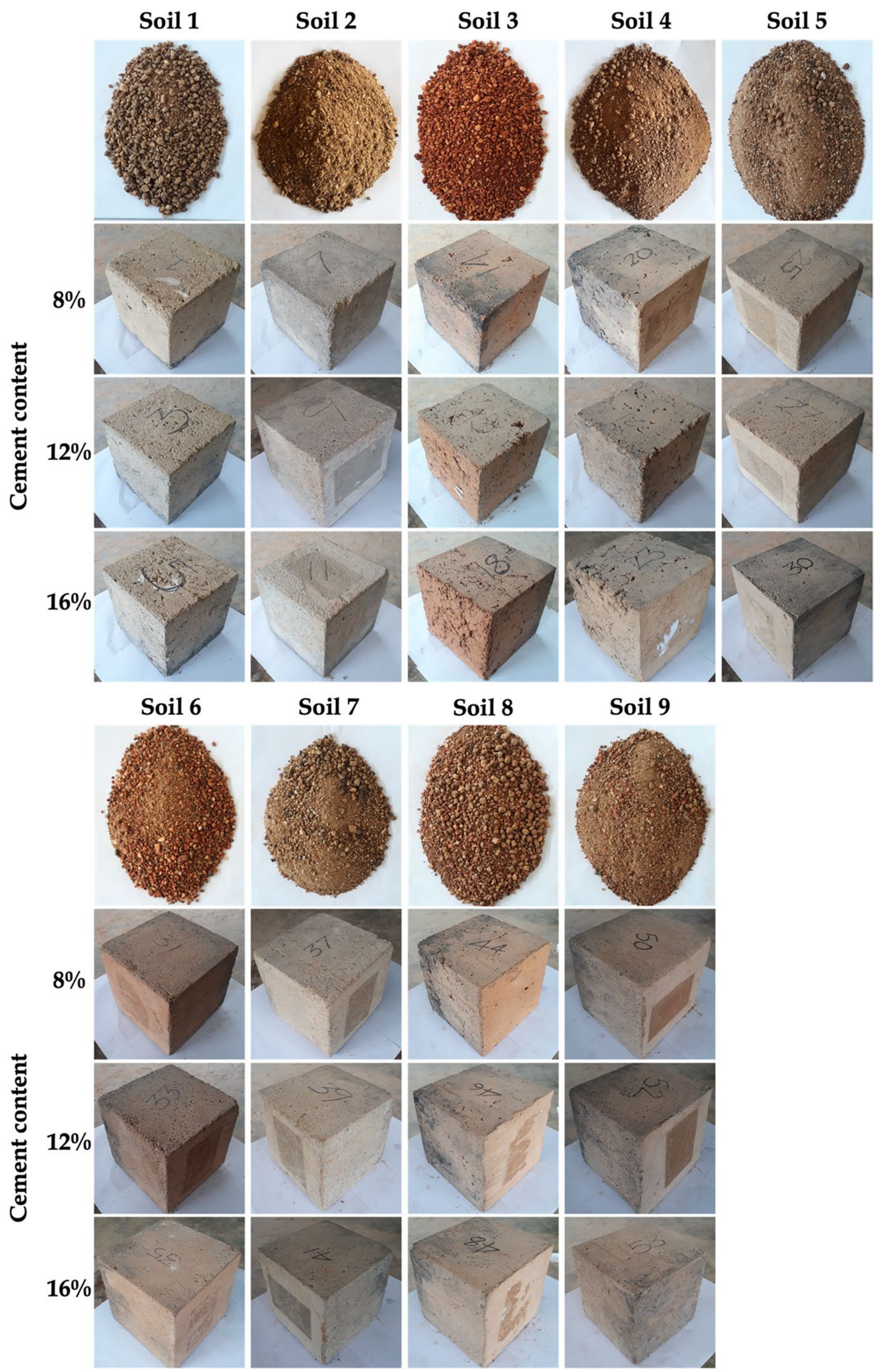


Fig. 1 Physical appearance of soils and mortar cubes casted with various cement contents

Atterberg limit test

Adhering to ASTM D4318 [33], the Atterberg limits test classifies soil behavior based on moisture content. The liquid limit signifies the moisture level at which soil transitions from liquid to plastic. This is determined by repeatedly dropping a soil paste

Table 1 Cement-soil mix used for the casting (material required for 1 kg of soil)

Mix ID	Cement content (%)	Soil type	Soil (g)	Cement (g)	Water (g)
S1-08	8	Soil 1	1000	80	88
S2-08	8	Soil 2	1000	80	96
S3-08	8	Soil 3	1000	80	95
S4-08	8	Soil 4	1000	80	133
S5-08	8	Soil 5	1000	80	88
S6-08	8	Soil 6	1000	80	120
S7-08	8	Soil 7	1000	80	82
S8-08	8	Soil 8	1000	80	102
S9-08	8	Soil 9	1000	80	92
S1-12	12	Soil 1	1000	120	88
S2-12	12	Soil 2	1000	120	96
S3-12	12	Soil 3	1000	120	95
S4-12	12	Soil 4	1000	120	133
S5-12	12	Soil 5	1000	120	88
S6-12	12	Soil 6	1000	120	120
S7-12	12	Soil 7	1000	120	82
S8-12	12	Soil 8	1000	120	102
S9-12	12	Soil 9	1000	120	92
S1-16	16	Soil 1	1000	160	88
S2-16	16	Soil 2	1000	160	96
S3-16	16	Soil 3	1000	160	95
S4-16	16	Soil 4	1000	160	133
S5-16	16	Soil 5	1000	160	88
S6-16	16	Soil 6	1000	160	120
S7-16	16	Soil 7	1000	160	82
S8-16	16	Soil 8	1000	160	102
S9-16	16	Soil 9	1000	160	92

until a specific groove closes within a set distance. The plastic limit, on the other hand, represents the moisture content where the soil transforms from a plastic state to a brittle state. This is achieved by measuring the moisture content of hand-rolled soil threads as they thin to a specific diameter. These tests provide valuable insights into the workability and behavior of soils at varying moisture levels.

Standard proctor compaction test

The standard Proctor compaction test (SPCT) as outlined in ASTM D698 [34] was employed to investigate the influence of moisture content on the packing behavior of lateritic soils. This test establishes a given soil sample's optimal moisture content and maximum achievable dry density. Air-dried soil, sieved to remove large particles, was separated into portions with varying moisture contents through controlled water addition. Each portion was compacted in a mold using a standard rammer with consistent blows per layer. The compacted soil's wet and dry weights, and representative moisture content samples, were determined. The calculated dry density was then plotted against

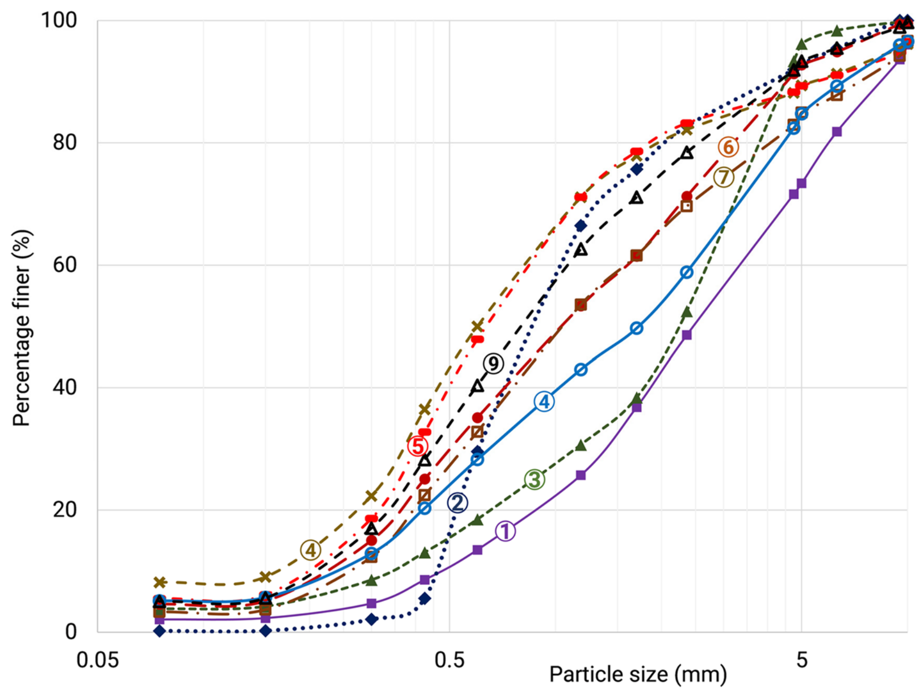


Fig. 2 Particle size distribution of soils (numbers indicate the soil type)

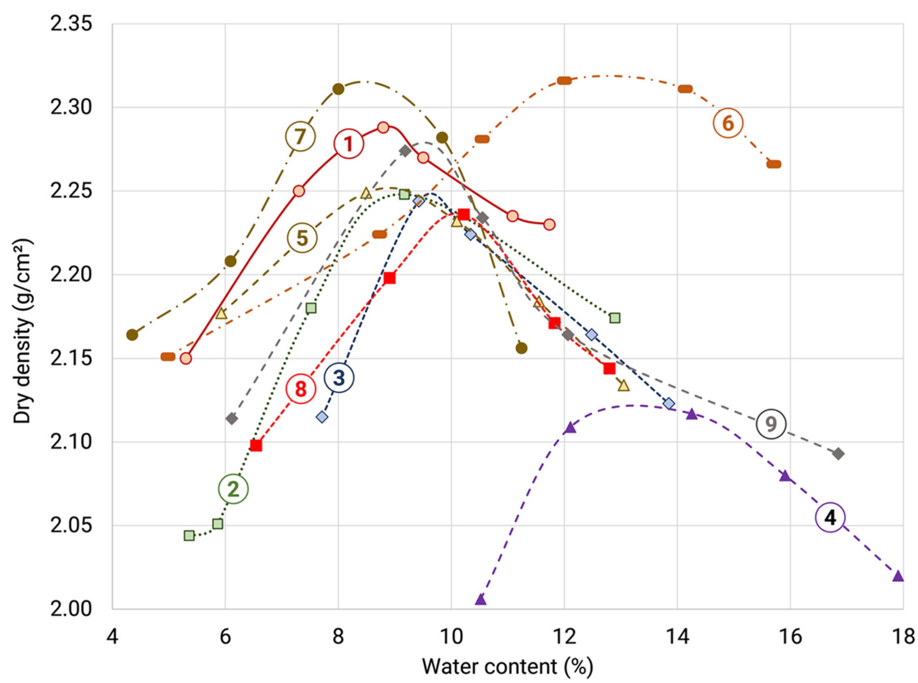


Fig. 3 Standard Proctor compaction test results — dry density variation with moisture content (numbers indicate the soil type)

the corresponding moisture content, revealing the peak that signifies the optimal moisture content at which the soil achieves its maximum dry density. This data, presented in Fig. 3, provides crucial insights into the relationship between moisture content and packing density of the lateritic soils used in this study.

Measuring moisture content and UPV

Following the guidelines outlined in EN 16322 [34], three 150 mm × 150 mm × 150 mm mortar specimens for each category were subjected to a drying test. The initial dry weight of each specimen (W_d) was determined by drying them in an oven at 100 °C for 24 h. After drying, the specimens were cooled in a lab environment. Next, the specimens were fully saturated by submerging them in a water bath for 24 h. Upon reaching saturation (100% moisture content), their weight (W_s) was measured again. Subsequently, the saturated specimens were placed back in the lab environment (30 °C temperature and 80% humidity). At predetermined intervals, each cube's weight (Wt) and ultrasonic pulse velocity (UPV) were recorded. The corresponding moisture content was then calculated using Eq. (2) as described in [31].

$$Moisturecontent\ (\%) = (W_t - W_d) \times 100/(W_s - W_d) \quad (2)$$

To achieve various moisture levels below the natural state, the cubes underwent drying in an oven. The drying process involved exposing the cubes to different temperatures and durations. Moisture content measurements were recorded at each stage.

At same predetermined intervals, the mortar cubes underwent an ultrasound test. This nondestructive technique measured the speed of sound waves traveling through the material. The process adhered to the ASTM C597 standard [35]. A portable ultrasonic testing device was used (Controls 58-E4800 Ultrasonic Pulse Velocity Tester). This equipment has features like a 2-MHz sampling rate, a time measurement range of up to 16 ms with 0.1-s resolution, and a transmitter output of 1200 V. The setup for the test is illustrated in Fig. 4. To ensure accuracy, the UPV for each sample was measured three times, and the final result was determined based on the average UPV of these measurements.

Machine learning modelling

Figure 5 presents a simplified flowchart outlining the modelling process.

Data acquisition

An extensive dataset containing 1620 data points was established based on the experimental program. This data is the foundation for building a robust model capable of

Fig. 4 UPV measurement equipment and setup

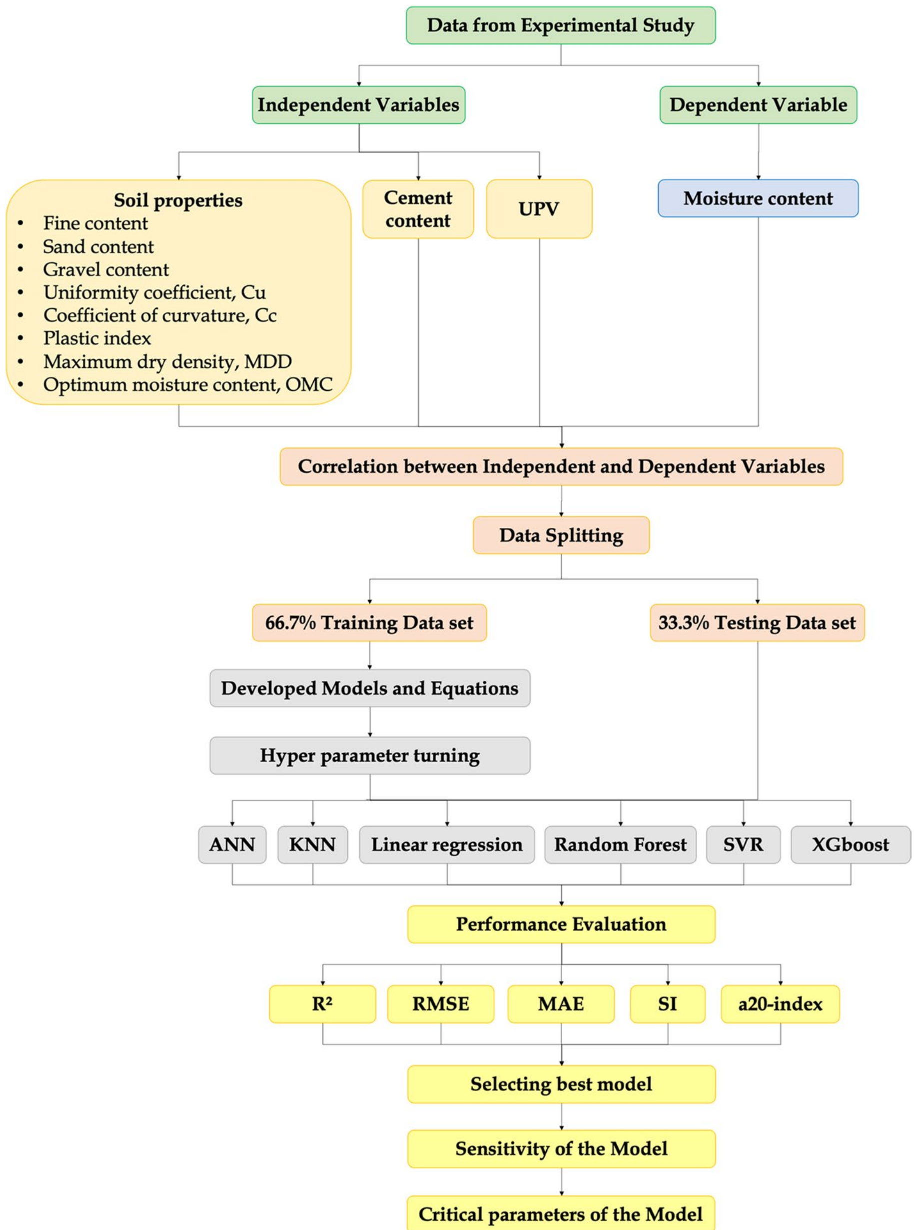


Fig. 5 Machine learning modelling flow chart

predicting sorption behavior. Ten input variables that include soil properties (gravel %, sand %, fine %, uniformity coefficient (Cu), coefficient of gradation (Cc), plasticity index (PI), optimum moisture content (OMC), maximum dry density, (MDD)), cement content, and ultrasonic pulse velocity (UPV) are considered for the model.

This study explored the potential of six distinct machine-learning algorithms:

- Artificial neural network (ANN)
- K-nearest neighbors (KNN)
- Linear regression (LR)
- Random forest regression (RF)

- Support vector regression (SVR)
- Extreme gradient boosting (XGB)

The compiled data was strategically divided into two sets: training set (two-thirds of total data) is used to train and develop the machine learning models, and testing set (one-third of total data) is employed to assess the models' performance on previously unseen data for validation of the model. This step ensures the models' ability to generalize effectively.

Several performance metrics were employed, and their formulation was given by Eqs. (3)–(6) to assess the effectiveness of the models [36].

- Coefficient of determination (R^2) value: Closer to 1 indicates a better fit between predicted and observed values.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(P_i - E_i)^2}{\sum_{i=1}^{n}\left(E_i - \overline{E}\right)^2} \tag{3}$$

- Root-mean-squared error (RMSE) and mean absolute error (MAE): Lower values indicate better model performance.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(P_i - E_i)^2}{n}} \tag{4}$$

$$MAE = \frac{\sum_{i=1}^{n}|P_i - E_i|}{n} \tag{5}$$

- Scatter index (SI): The scatter index (SI) is a normalized metric used to evaluate the accuracy of a model in predicting a specific parameter. It essentially represents the percentage error relative to the mean observed value. A lower value indicated better agreement between expected and measured value.

$$SI = \frac{RMSE}{\overline{E}} \tag{6}$$

- a20 index: The fraction of predicted data points within ± 20% of measured data points. A higher fraction indicates better agreement between predicted and observed values.

Results and discussion

Soil properties

Table 2 summarizes various soil properties measured through different tests for nine soil samples. The amount of fine particles (less than 75 μm) is generally low across all

Table 2 Summary of soil properties used for the casting of CSEBs

Characteristics	Properties	Soil 1	Soil 2	Soil 3	Soil 4	Soil 5	Soil 6	Soil 7	Soil 8	Soil 9
Grain size	Gravel (%)	33.4	8.0	6.6	11.8	11.7	8.7	17.1	17.6	8.1
	Sand (%)	64.3	91.7	89.1	79.1	82.2	86.2	79.2	76.7	86.3
	Fine (%)	2.3	0.2	4.2	9.0	6.1	5.1	3.7	5.7	5.6
	Uniformity coefficient (C_u)	8.0	2.2	7.4	5.0	4.5	7.6	6.4	10.0	5.5
	Coefficient of gradation (C_c)	1.3	0.8	1.6	1.2	1.0	0.7	0.8	0.7	0.9
Atterberg limit	Liquid limit, W_L (%)	15.5	0.0	37.6	31.0	0.0	37.5	0.0	26.8	0.0
	Plastic limit, P_L (%)	32.8	57.0	39.7	32.3	8.8	77.7	44.1	33.2	50.4
	Plasticity index, I_P (%)	17.3	57.0	2.1	1.4	8.8	40.2	44.1	6.4	50.4
Proctor test	Optimum moisture content, OMC (%)	8.8	9.6	9.5	13.3	8.8	12.0	8.2	10.2	9.2
	Maximum dry density, MDD (kg/m^3)	2.288	2.250	2.246	2.126	2.252	2.316	2.314	2.236	2.274
USCS classification		SW	SP	SW	SW-ML	SW-CL	SP-CH	SP	SP-OL	SP-CH

SW well-graded sand, *SP* poorly graded sand, *ML* inorganic silts of slight plasticity, *CL* inorganic clays of low plasticity, *CH* inorganic clay of high plasticity, *OL* organic soil of low plasticity soils, ranging from 0.2 to 9.0%, and sand is the dominant particle size fraction, with values ranging from 64.3 to 91.7%. Most soils exhibit a broader range of particle sizes (Cu > 5) based on the coefficient of uniformity (Cu) values. Exceptions include soils 2 and 5. The coefficient of curvature (Cc) indicates that most soils have a near-uniform (Cc≈1) or slightly nonuniform particle size distribution. Plasticity index (PI) varies significantly. Soils 2, 6, 7, and 9 show high plasticity ($PI > 40$), while the remaining soils have low plasticity ($PI < 10$) except for soil 1 with medium plasticity (PI between 10 and 40). Maximum dry density (MDD) values are relatively consistent across most soils (2.126 to 2.316 g/cm^3). Optimum moisture content (OMC) shows some variation, with slightly higher values for soil 4 and soil 6. The dominance of sand and low fine content in most soils might influence factors like water absorption and retention capacity.

Correlation between moisture content and UPV

Figure 6 illustrates the moisture content variation with UPV for various cement content and soil types used for CSEBs. As shown in figure, the increase in UPV with moisture content in CSEBs can be explained by several factors. In dry CSEBs,

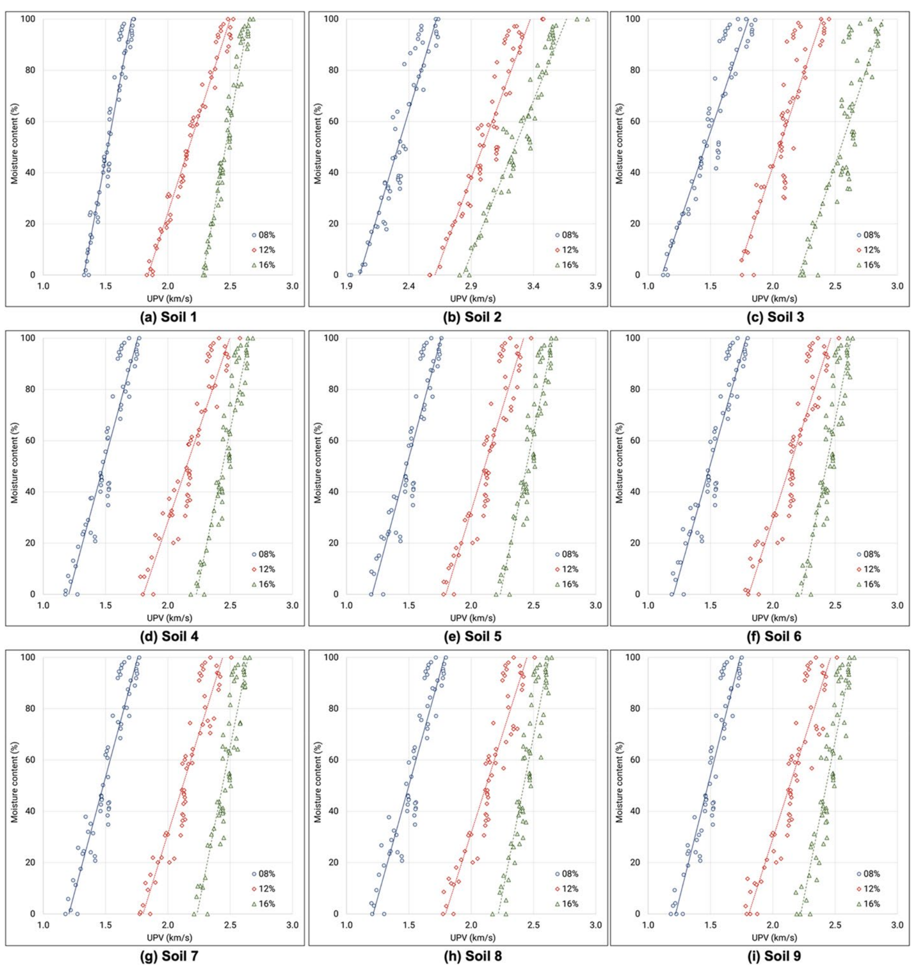


Fig. 6 Variation of moisture content with UPV for CSEBs with various soils and cement content

air occupies the voids or pores within the material. Air has a much lower acoustic impedance (resistance to sound waves) compared to solid materials like cement or soil particles [37]. As moisture content increases, water fills these voids. Water has a significantly higher acoustic impedance than air. Also, the velocity of sound waves is directly related to the material's density and stiffness [38]. With more water filling the pores, the overall density of the CSEB increases. Water particles are more tightly packed than air, leading to a stiffer material. Due to the combined effects of increased density and stiffness, sound waves travel faster through the CSEB as the moisture content rises. This translates to a higher UPV value.

For the particular moisture condition, UPV increases with increase in cement content. Cement acts as a binder, filling the gaps between soil particles and creating a denser matrix. It leads to a reduction in the overall porosity of the CSEB. With less cement, there are more voids or air pockets even in the dry state [39]. Therefore, the increase in UPV with moisture content will be more pronounced due to the significant difference in acoustic impedance between air and water. A denser matrix with less initial porosity due to higher cement content will lead to a smaller increase in UPV with increasing moisture. This is because fewer voids are filled with water, resulting in a smaller relative change in overall density and stiffness [40]. In addition to cement content, the relationship between moisture and UPV was varied for soil to soil. Therefore, it is evident that soil properties also effect the relationship between moisture and UPV.

Prediction models

Hyperparameter optimization

Optimizing the performance of the ANN model involves carefully adjusting its internal structure, also known as hyperparameter tuning [41]. This process involved exploring various configurations for the following:

- Number of layers: Up to two hidden layers were investigated.
- Number of neurons per layer: Each layer could contain up to 12 neurons.

Figure 7a illustrates how the RMSE, a measure of prediction accuracy, changed as these parameters were adjusted. The results revealed that a two-layer network with eight neurons per layer best predicted sorption behavior. Another crucial hyperparameter is the learning rate, which controls how quickly the ANN learns from the training data. Figure 7b presents the investigation of six different learning rates. The optimal learning rate, determined based on the training data performance, was 0.1.

Optimizing the KNN model requires striking a balance between bias and variance [42]. Bias refers to the model's tendency to underfit or overfit the data, while variance reflects the model's sensitivity to changes in the training data. This process involved exploring different k values, representing the number of nearest neighbors considered when making predictions. A technique called cross-validation was employed. Figure 8 depicts how the RMSE changed as the value of k increased. Lower RMSE indicates better prediction accuracy. Analyzing this data revealed that using $k = 6$ neighbors

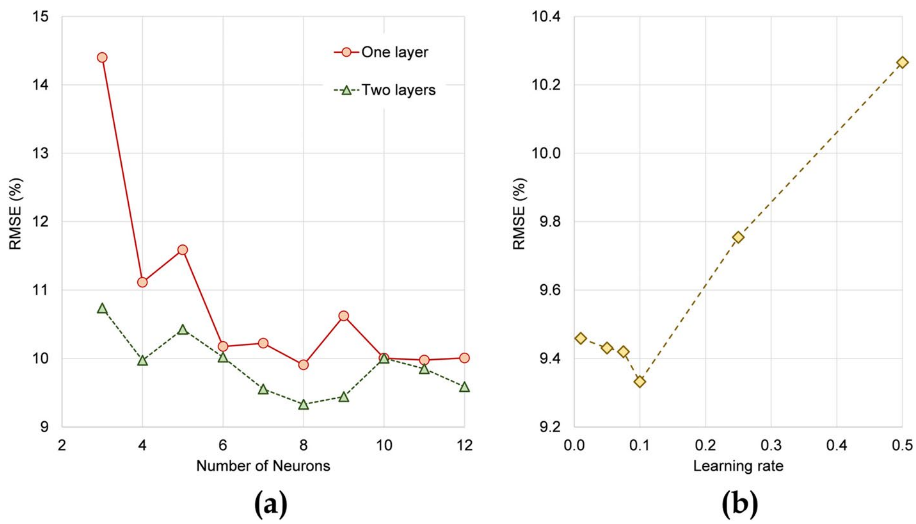


Fig. 7 RMSE variation for ANN models with **a** various numbers of neurons and **b** learning rates

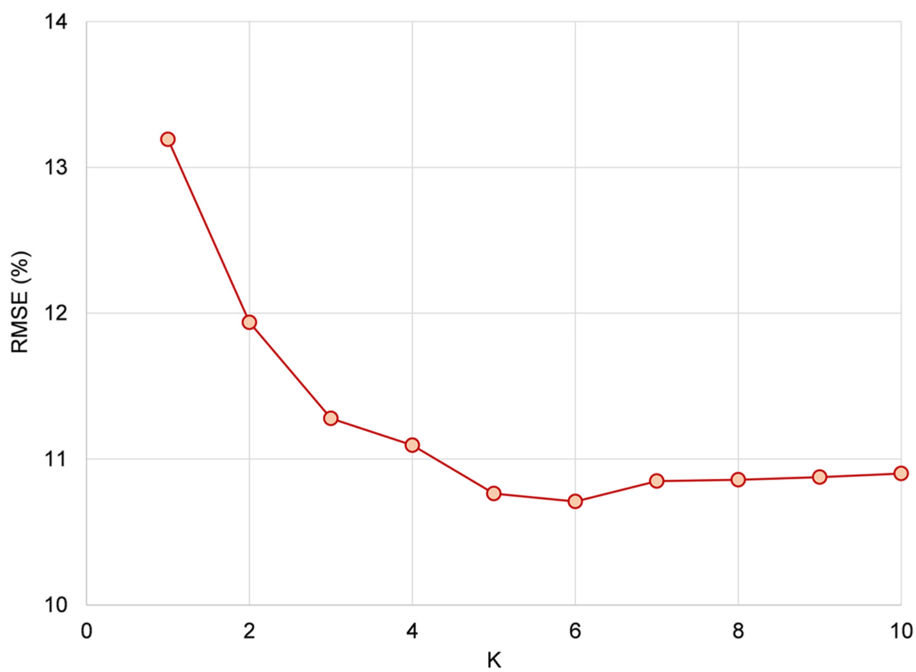


Fig. 8 RMSE variation for KNN models with various k values

resulted in the best performance on the training data. This value was then adopted for further analysis.

Optimizing the performance of random forest regression, support vector regression, and XGBoost models involves adjusting various settings known as hyperparameters [43]. Table 3 summarizes the different values explored for each hyperparameter during the optimization process and the specific hyperparameter values chosen for the model that achieved the best performance on the training data.

Table 3 Hyperparameter optimization for prediction models

Model	Hyperparameters	Analyzed values	No. of runs	Selected value
Artificial neural network	Number of layers	[1, 2]	20	2
	Neurons per layer	[3-12]		8
KNN	K	[1-10]	10	6
Random forest regression	Number of trees	[1-20]		20
	Number of terms sampled per split	[1-8]		8
			500	
	Maximum number of splits per tree	100		100
	Minimum splits per tree	1		1
	Minimum size split	1		1
Support vector regression	Kernel	[Linear, radial basis function]		Radial basis function
	Cost	[0.01–5]		4.8607
	Gamma	[0.001–0.5]		0.4992
XGBoost	Max_depth	[1-5]	500	3
	Subsample	[0.5–1]		0.9834
	Colsample_by tree	[0.5–1]		0.9413
	Min_child_weight	[1-3]		2.9867
	Alpha	[0–0.5]		0.3028
	Lambda	[0–2]		0.7155
	Learning rate	[0.05–0.2]		0.1913
	Iterations	[10–50]		30

Performance of the models

Figure 9 reveals the results for the machine learning model's ability to predict moisture content. The data points cluster relatively close to the diagonal line, signifying a good correlation between predicted and actual moisture values. The XGB model emerged as the most accurate model, with the highest proportion of predictions falling within a 20% error range of the actual measurements. This is evident in both the training data (78%) and test data (69%). The KNN model also achieved comparable overall accuracy (around 75% for training and 66% for testing). This suggests that XGB and KNN models more effectively captures the underlying relationship between the features and the target variable (moisture content). The linear regression model displayed the least impressive performance among the six models evaluated.

Table 4 summarizes the performance indexes for training and validation datasets. Analyzing the performance of various machine learning models for predicting moisture content reveals XGB as the most promising model. XGB boasts superior performance across most metrics, exhibiting the highest training and validation R-squared values (indicating a strong correlation between predictions and actual data) and the lowest training and validation errors (RMSE, MAE, and SI). While KNN also demonstrates good results, particularly in R^2 values, XGB stands out with its consistently lower prediction errors. Linear regression (LR) shows the least desirable performance, highlighting the importance of model selection. However, it is crucial to acknowledge the potential for overfitting (model performing well on training data but not generalizing well) due to

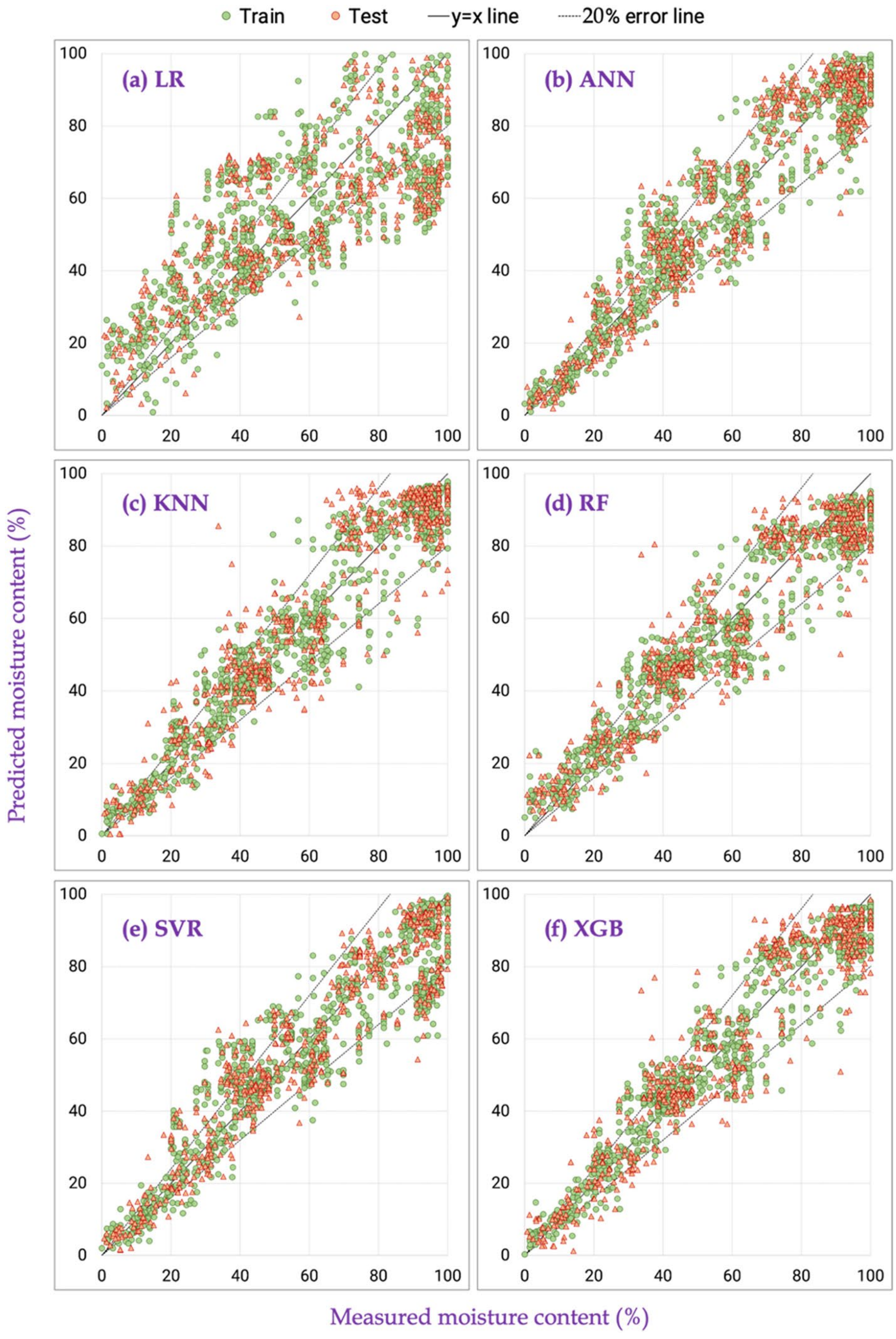


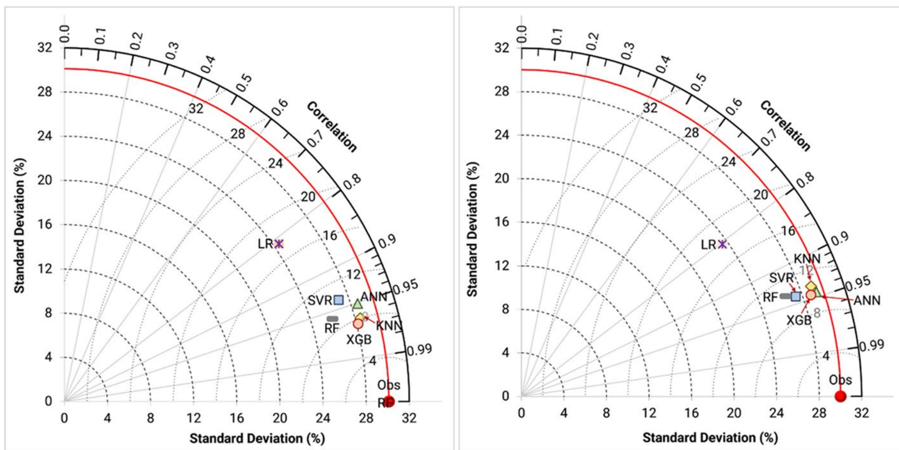
Fig. 9 Predicted vs. measured moisture content of CSEBs for various machine learning techniques (dot line indicates the 20% error)

the observed discrepancy between training and validation performance in some cases, such as KNN model.

XGBoost combines multiple weak learners (decision trees) into a stronger learner, potentially leading to improved accuracy and handling complex relationships in the data. Also, it can mitigate overfitting by penalizing overly complex models, often

Table 4 Performance indicators for various machine learning techniques

	Train					Valid				
	R^2	RMSE	MAE	SI	a20	R^2	RMSE	MAE	SI	a20
ANN	0.904	9.33	7.26	0.17	0.69	0.891	9.91	7.92	0.18	0.68
KNN	0.929	8.03	5.92	0.14	0.75	0.877	10.54	7.95	0.19	0.66
LR	0.661	17.53	14.37	0.32	0.44	0.646	17.87	14.85	0.32	0.44
RF	0.908	9.15	7.31	0.17	0.72	0.876	10.58	8.42	0.19	0.69
SVR	0.882	10.34	7.69	0.19	0.66	0.887	10.11	7.58	0.18	0.69
XGB	0.936	7.60	5.80	0.14	0.78	0.894	9.75	7.60	0.17	0.69



(a) Training dataset **(b) Validation dataset**

Fig. 10 Taylor diagram for **a** training datasets and **b** validation datasets

resulting in better generalization to unseen data. KNN model makes predictions based on the closest data points (neighbors) in the training data, offering some level of explainability. It can be well-suited for datasets with localized patterns and smooth underlying relationships. These factors provided a better chance for XGB and KNN models to perform well over other machine learning models.

A Taylor diagram visually assesses how well machine learning models predict the moisture content. Each model's symbol on the diagram (Fig. 10, red line) agrees with actual data. The closer the symbol is to the red point (obs), the better the match between predictions and measured values. This considers key statistics like standard deviation and correlation [44]. The XGB model sits closest to the red point, indicating the most accurate predictions. Conversely, the LR model is furthest, suggesting the least accurate predictions. The horizontal black line shows the correlation between predicted and actual values. The XGB, positioned closer to this line, has the highest correlation. KNN and ANN models also performed better than other models. Model performance depends on data and task complexity. XGB excels in complex moisture content prediction where input features and outcomes (moisture content) have intricate, nonlinear relationships. XGB, ANN, and KNN can capture these hidden dependencies better than simpler models like LR. Based on the Taylor diagram, the ranking from best- to worst-performing model is as follows: XGB > ANN > KNN > RF > SVR > LR.

Sensitivity analysis

While powerful, artificial neural networks (ANNs) can be like "black boxes." Their intricate internal workings make it difficult to understand how they reach their predictions [45]. This lack of interpretability can be a challenge. Fortunately, a method called SHAP (SHapley Additive exPlanations) offers a solution [46, 47]. SHAP provides a way to see "inside" complex models with many factors. It assigns a unique value (SHAP value) to each input feature for a specific prediction. In simpler terms, SHAP tells that how much each feature "impacts" the model's output [48, 49].

Figure 11 shows the average SHAP values for the ANN model's predictions of moisture content. The UPV has the highest SHAP value. According to the model, UPV is the single most influential factor affecting the predicted moisture content. While UPV can indicate moisture content, it is not a direct measurement. Soil properties inherently impact both UPV and moisture content, making them interrelated factors. Generally, the sound wave travel faster in water than void. So, when void fill with water (moisture content increase), UPV value also increase proportionally. In some cases, UPV might be a good predictor of moisture content, but it depends on the specific mix design and how the UPV relates to the chosen soil in the mortar. Although UPV can be a valuable tool for estimating moisture content in mortar, it is most effective when used in conjunction with other factors, such as known soil properties and a calibrated relationship between UPV and moisture content for the specific mortar mix. However, soil properties themselves play a crucial role in determining the overall moisture content of the mortar. Cement content is the second influence factor on

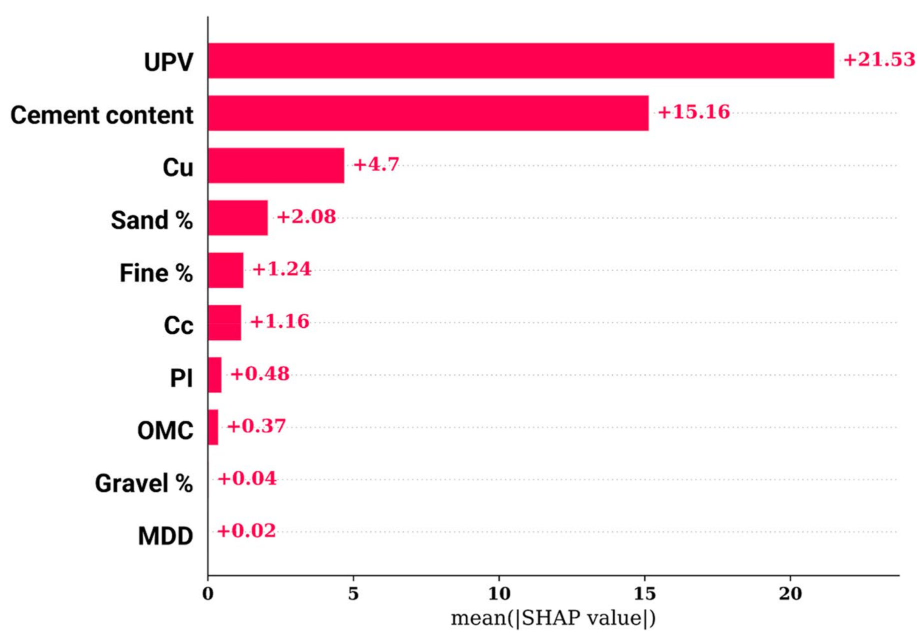


Fig. 11 Mean SHAP values for input variable based on ANN model

prediction of moisture content. Cement acts as a binder in mortar, filling voids and reducing the overall porosity. This can lead to less space for water to be absorbed and retained. As porosity affect the moisture content of the mortar, and the cement content influences on moisture content of the mortar.

In soil characteristics, uniformity coefficient (Cu) have the most influence parameter effects of the moisture content prediction followed by sand content, fine content, and Cc. The gravel content and MDD have no significant effect on prediction of moisture content. By capturing the information about soil gradation, Cu provides valuable input for the machine learning model. The model can learn the relationship between Cu and the observed moisture content data. For well-graded soil (Cu more than 4), the model can expect a lower range of moisture content values due to the limited void space. Conversely, for poorly graded soil (Cu less than 4), the model can anticipate a wider range of moisture content possibilities due to the variable void sizes impacting water retention. Overall, Cu acts as a proxy for various soil properties that influence moisture content. It provides the machine learning model with a single, informative metric that captures the overall effect of particle size distribution on water-holding capacity.

While Cu is significant, it is not the only factor. Other soil properties like sand content and fine content also play a role in moisture retention and can be included as additional features in the machine learning model. Their impact works in opposing ways. Sand particles are relatively large and have minimal surface area. This translates to less space for water to adhere and fewer capillary forces acting on the water molecules. Consequently, higher sand content generally leads to lower overall moisture content. Sandy soils have larger voids between particles, allowing for better drainage and faster water movement. This reduces the amount of water the soil can retain. On the other hand, silt and clay particles are much smaller and have a significantly larger total surface area compared to sand. This expansive surface area provides numerous sites for water adhesion and creates strong capillary forces that draw and hold water molecules. Therefore, higher fine content leads to higher overall moisture content. By including sand content and fine content as features, the model can establish correlations between these features and the observed moisture content data.

Figure 12 showcases a valuable tool called a SHAP summary plot. This plot helps us visualize how different features influence the XGB model's predictions for moisture content. The color spectrum represents the range of possible values for each feature. The horizontal axis shows the SHAP value, which quantifies a feature's contribution to the predicted moisture content. Red dots represent features with high positive SHAP values, indicating they tend to increase the predicted moisture content. The plot reveals that UPV has the strongest positive influence on predicted moisture content. This is evident by the highest SHAP value on the positive side. In simpler terms, higher UPV is associated with higher predicted moisture content within the range of values used in this study. Conversely, cement content has a negative influence. The negative SHAP value on the negative side suggests that lower cement content tends to be associated with higher predicted moisture content.

Sand content also has a negative influence, while the impact of the coefficient of curvature (Cc) is mixed. On the other hand, the uniform coefficient (Cu) and fine content exhibit positive contributions to predicted moisture content. It is important to note that

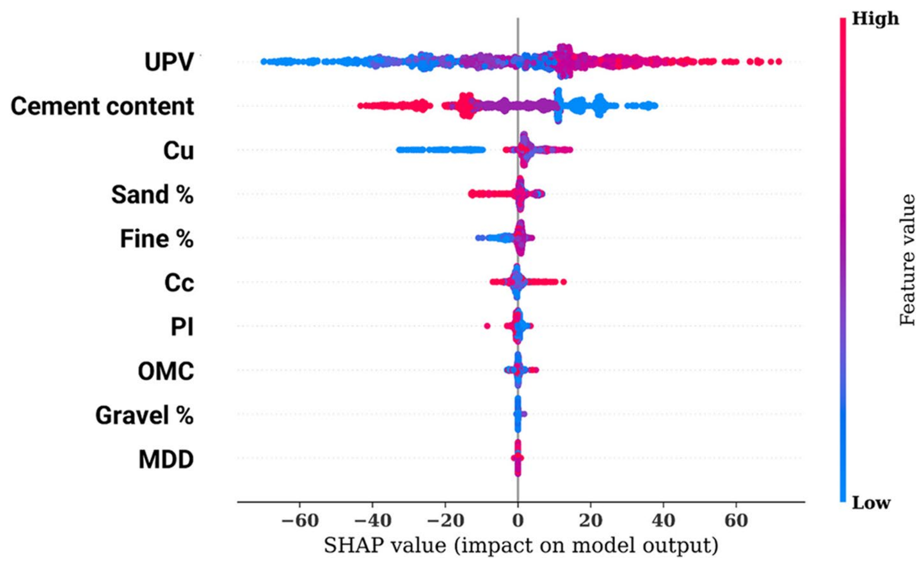


Fig. 12 SHAP summary plot for input variable based on ANN model

the changes in moisture content due to variations in features other than UPV seem to be relatively small. However, the SHAP summary plot offers valuable insights by revealing the relative importance of each feature for the XGB model's predictions. This understanding allows us to better interpret the model's behavior and identify critical factors that influence the moisture content of CSEBs.

Feature importance

After identifying key features using SHAP analysis, this study explored how well the XGB model could predict moisture content with fewer input variables. This is crucial for real-world applications where getting specific data might be difficult, or you must make predictions with limited information. Figure 13 shows the model's performance with different combinations of features. Features based on their importance (determined by SHAP) was gradually added, starting with the most impactful ones. Case 1 considers only the two most critical features: UPV and cement content. Cases 2 through 9 then progressively incorporate additional features with lower SHAP values.

As expected, using more features improves prediction accuracy, as shown by the lower RMSE. However, the accuracy improvement becomes less significant after Case 2, which uses only UPV, cement content, and uniformity coefficient. This combination achieves an RMSE of around 10% for both training and validation datasets. Adding sand content as an input variable further reduces the RMSE to around 9%. Notably, this performance is closer to using all input parameters. While using all available features can enhance model accuracy, focusing on a smaller set of critical features identified through SHAP analysis retains a significant portion of the predictive power. This reduction in required input parameters is valuable when acquiring all data points might be impractical.

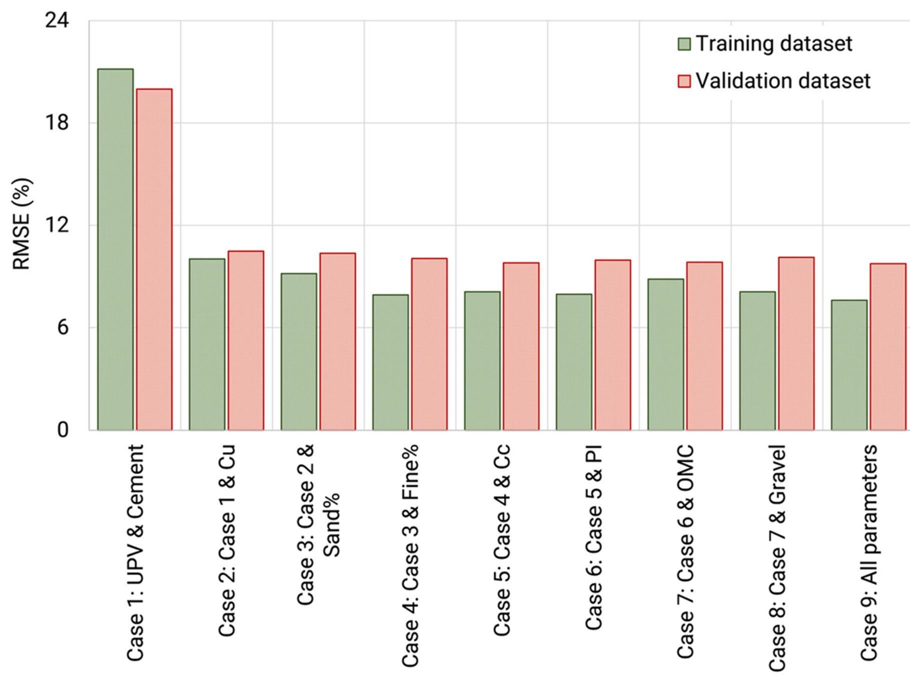


Fig. 13 RMSE variation for the various combinations of input parameters

Conclusions

This study has underscored the critical role of moisture content in CSEBs and explored the potential of machine learning for its prediction based on soil properties, cement content, and UPV. Based on the results, the following conclusions were drawn:

- By comparing various machine learning algorithms, XGBoost emerged as the most effective, demonstrating superior accuracy in predicting moisture content.
- The SHAP analysis revealed that UPV was identified as the most significant factor affecting predicted moisture content, followed by cement content and specific soil properties such as uniformity coefficient.
- The moisture content prediction with combination of UPV, cement content, and uniformity coefficient as an input variable can achieve good accuracy. This is particularly valuable for real-world applications where obtaining all data points might be impractical or time-consuming.

Construction sites often require rapid moisture content assessments to ensure optimal material performance. This simplified approach offers a practical solution for such scenarios. This study has significantly contributed to understanding and managing moisture content in CSEBs. By highlighting the detrimental impact of moisture and demonstrating the effectiveness of machine learning models in their prediction, the research provides valuable tools for optimizing CSEB production and construction practices. Furthermore, identifying critical factors influencing moisture content and exploring reduced-feature models pave the way for further advancements in CSEB technology.

Abbreviations

ANN	Artificial neural network
KNN	K-nearest neighbors
LR	Linear regression
RF	Random forest regression
SVR	Support vector regression
XGB	Extreme gradient boosting
R^2	Coefficient of determination
RMSE	Root-mean-squared error
MAE	Mean absolute error
SI	Scatter index
CSEBs	Cement-stabilized earth blocks
Cu	Uniformity coefficient
Cc	Coefficient of curvature
PI	Plasticity index
SHAP	SHapley Additive exPlanations
UPV	Ultrasonic pulse velocity
σ_c	Compressive strength

Acknowledgements
The author is grateful for the assistance of Perera A. A. S. A. and Abeynayake P. P., Department of Civil and Environmental Engineering, University of Jaffna, for their valuable support to carry out the experimental program. The author expresses his sincere gratitude for the support given by the Department of Civil and Environmental Engineering, Faculty of Engineering, University of Jaffna.

Authors' contributions
Study conception and design, NS and PJ; data acquisition, NS and TR; analysis and interpretation of results, NS, TR, and PJ; draft manuscript preparation, NS and PJ; and manuscript review and editing, NS.

Funding
Not applicable.

Availability of data and materials
Data can be made available on request by interested parties.

Declarations

Competing interests
The authors declare that they have no competing interests.

Received: 2 August 2024 Accepted: 14 September 2024
Published online: 19 September 2024

References

1. Sathiparan N (2023) Performance of sustainable cement mortar containing different types of masonry construction and demolition wastes. Clean Technol Environ Policy. https://doi.org/10.1007/s10098-023-02681-2
2. Sathiparan N, Subramaniam DN, Malsara KGN, Akmal MSM (2022) Thermal comfort analysis of fired-clay brick, cement-sand block and cement stabilized earth block masonry house models. Innov Infrastruct Solut 7(2):147. https://doi.org/10.1007/s41062-022-00744-9
3. Yogananth Y, Thanushan K, Sangeeth P, Coonghe JG, Sathiparan N (2019) Comparison of strength and durability properties between earth-cement blocks and cement–sand blocks. Innov Infrastruct Solut 4(1):50. https://doi.org/10.1007/s41062-019-0238-8
4. Assia Z, Fazia F, Abdelmadjid H (2019) Sustainability of the stabilized earth blocs under chemicals attack's effects and environmental conditions. Constr Build Mater 212:787–798. https://doi.org/10.1016/j.conbuildmat.2019.03.324
5. Chen X, Huang W, Zhou J (2012) Effect of moisture content on compressive and split tensile strength of concrete. Indian J Eng Mater Sci 19:427–435
6. Elahi TE, Shahriar AR, Islam MS (2021) Engineering characteristics of compressed earth blocks stabilized with cement and fly ash. Constr Build Mater 277:122367. https://doi.org/10.1016/j.conbuildmat.2021.122367
7. Popovics S (1998) John Wiley & Sons. Incorporated, New York
8. Foraboschi P (2017) Specific structural mechanics that underpinned the construction of Venice and dictated Venetian architecture. Eng Fail Anal 78:169–195. https://doi.org/10.1016/j.engfailanal.2017.03.004
9. Foraboschi P (2016) The central role played by structural design in enabling the construction of buildings that advanced and revolutionized architecture. Constr Build Mater 114:956–976. https://doi.org/10.1016/j.conbuildmat.2016.03.092
10. Foraboschi P, Vanin A (2014) Experimental investigation on bricks from historical Venetian buildings subjected to moisture and salt crystallization. Eng Fail Anal 45:185–203. https://doi.org/10.1016/j.engfailanal.2014.06.019
11. Witzany J, Cejka T, Zigler R (2010) The effect of moisture on significant mechanical characteristics of masonry. Statybinės Konstrukcijos ir Technologijos 2(3):79–85. https://doi.org/10.3846/skt.2010.11

12. AM Amde, J Martín, J Colville. The effects of moisture on compressive strength and modulus of brick masonry. In 13th International Brick and Block Masonry Conference. Amsterdam: 2004.
13. Sathiparan N, Anburuvel A, Maduwanthi KAPN (2022) SRACB Dasanayake, Effect of moisture condition on cement masonry blocks with different fine aggregates: river sand, lateritic soil and manufactured sand. Sādhanā 47(4):270. https://doi.org/10.1007/s12046-022-02054-3
14. Phillipson MC, Baker PH, Davies M, Ye Z, McNaughtan A, Galbraith GH, McLean RC (2007) Moisture measurement in building materials: an overview of current methods and new approaches. Build Serv Eng Res Technol 28(4):303–316. https://doi.org/10.1177/0143624407084184
15. Dietsch P, Franke S, Franke B, Gamper A, Winter S (2015) Methods to determine wood moisture content and their applicability in monitoring concepts. J Civ Struct Heal Monit 5(2):115–127. https://doi.org/10.1007/s13349-014-0082-7
16. Orr SA, Fusade L, Young M, Stelfox D, Leslie A, Curran J, Viles H (2020) Moisture monitoring of stone masonry: a comparison of microwave and radar on a granite wall and a sandstone tower. J Cult Herit 41:61–73. https://doi.org/10.1016/j.culher.2019.07.011
17. Rymarczyk T, Kłosowski G, Kozłowski E (2018) A non-destructive system based on electrical tomography and machine learning to analyze the moisture of buildings. Sensors 18(7):2285. https://doi.org/10.3390/s18072285
18. Almutawa N, Eid W (2023) Soil moisture content estimation using active infrared thermography technique: an exploratory laboratory study. Kuwait J Sci 50(3):399–404. https://doi.org/10.1016/j.kjs.2022.11.001
19. Ahmad A, Ahmad W, Aslam F, Joyklad P (2022) Compressive strength prediction of fly ash-based geopolymer concrete via advanced machine learning techniques. Case Studi Construct Mat 16:e00840https://doi.org/10.1016/j.cscm.2021.e00840
20. Zhang LV, Marani A, Nehdi ML (2022) Chemistry-informed machine learning prediction of compressive strength for alkali-activated materials. Constr Build Mater 316:126103. https://doi.org/10.1016/j.conbuildmat.2021.126103
21. Wu J, Jing H, Yin Q, Yu L, Meng B, Li S (2020) Strength prediction model considering material, ultrasonic and stress of cemented waste rock backfill for recycling gangue. J Clean Prod 276:123189https://doi.org/10.1016/j.jclepro.2020.123189
22. Bilgehan M (2011) A comparative study for the concrete compressive strength estimation using neural network and neuro-fuzzy modelling approaches. Nondestruct Test Eval 26(1):35–55. https://doi.org/10.1080/10589751003770100
23. Vidya Sagar R, Dutta M (2021) Combined usage of acoustic emission technique and ultrasonic pulse velocity test to study crack classification in reinforced concrete structures. Nondestruct Test Eval 36(1):62–96. https://doi.org/10.1080/10589759.2019.1692013
24. Asteris PG, Skentou AD, Bardhan A, Samui P, Lourenço PB (2021) Soft computing techniques for the prediction of concrete compressive strength using non-destructive tests. Constr Build Mater 303:124450. https://doi.org/10.1016/j.conbuildmat.2021.124450
25. Khashman A, Akpinar P (2017) Non-destructive prediction of concrete compressive strength using neural networks. Proc Comp Sci 108:2358–2362. https://doi.org/10.1016/j.procs.2017.05.039
26. El-Mir A, El-Zahab S, Sbartaï ZM, Homsi F, Saliba J, El-Hassan H (2023) Machine learning prediction of concrete compressive strength using rebound hammer test. J Build Eng 64:105538.https://doi.org/10.1016/j.jobe.2022.105538
27. Li D, Tang Z, Kang Q, Zhang X, Li Y (2023) Machine learning-based method for predicting compressive strength of concrete. Processes 11(2):390. https://doi.org/10.3390/pr11020390
28. Sathiparan N, Jeyananthan P, Subramaniam DN (2024) Surface response regression and machine learning techniques to predict the characteristics of pervious concrete using non-destructive measurement: ultrasonic pulse velocity and electrical resistivity. Measurement 225:114006. https://doi.org/10.1016/j.measurement.2023.114006
29. N Sathiparan, P Jeyananthan, Predicting compressive strength of quarry waste-based geopolymer mortar using machine learning algorithms incorporating mix design and ultrasonic pulse velocity, Nondestructive testing and evaluation 1-24.https://doi.org/10.1080/10589759.2024.2304257
30. N Sathiparan, P Jeyananthan, Predicting compressive strength of cement-stabilized earth blocks using machine learning models incorporating cement content, ultrasonic pulse velocity, and electrical resistivity, Nondestruct Testing Eval. 1–25. https://doi.org/10.1080/10589759.2023.2240940.
31. Giaccone D, Santamaria U, Corradi M (2020) An experimental study on the effect of water on historic brickwork masonry. Heritage 3(1):29. https://doi.org/10.3390/heritage3010003
32. ASTM-D422 (2007) Standard test method for particle-ssize analysis of soils. ASTM International, West Conshohocken, PA
33. ASTM-D4318 (2017) Standard test methods for liquid limit, plastic limit, and plasticity index of soils. ASTM International, West Conshohocken, PA
34. CSN-EN-16322 (2013) Conservation of cultural heritage - test methods - determination of drying properties. British Standards Institution (BSI), London
35. ASTM-C597 (2010) Standard test method for pulse velocity through concrete. ASTM International, West Conshohocken, PA
36. Sathiparan N, Jeyananthan P (2023) Prediction of masonry prism strength using machine learning technique: effect of dimension and strength parameters. Mat Today Commun 35:106282https://doi.org/10.1016/j.mtcomm.2023.106282
37. Zhang K, Ma C, He Q, Lin S, Chen Y, Zhang Y, Fang NX, Zhao X (2019) Metagel with broadband tunable acoustic properties over air–water–solid ranges. Adv Func Mater 29(38):1903699. https://doi.org/10.1002/adfm.201903699
38. Attal E, de l'Epine YB Dauchez N Dubus B (2021) Experimental investigation of the effect of moisture on the acoustic properties of lightweight substrates used in green envelopes. Appl Acoustics 180:108108. https://doi.org/10.1016/j.apacoust.2021.108108
39. Baldovino JD, de la Rosa YEN, Calabokis OP (2024) Effect of porosity/binder index on strength, stiffness and microstructure of cemented clay: the impact of sustainable development geomaterials. Materials 17(4):921. https://doi.org/10.3390/ma17040921

40. Panzera TH, Rubio JC, Bowen CR, Vasconcelos WL, Strecker K (2008) Correlation between structure and pulse velocity of cementitious composites. Adv Cem Res 20(3):101–108. https://doi.org/10.1680/adcr.2008.20.3.101
41. Sakshi S, Kumar R (2019) A neuro-genetic technique for pruning and optimization of ANN weights. Appl Artif Intell 33(1):1–26. https://doi.org/10.1080/08839514.2018.1525524
42. Cubillos M, Wøhlk S, Wulff JN (2022) A bi-objective k-nearest-neighbors-based imputation method for multilevel data. Expert Syst Appl 204:117298. https://doi.org/10.1016/j.eswa.2022.117298
43. Tarwidi D, Pudjaprasetya SR, Adytia D, Apri M (2023) An optimized XGBoost-based machine learning method for predicting wave run-up on a sloping beach. MethodsX 10:102119. https://doi.org/10.1016/j.mex.2023.102119
44. Sathiparan N, Jeyananthan P, Subramaniam DN (2023) Prediction of compressive strength of fly ash blended pervious concrete: a machine learning approach. Int J Pavement Eng 24(2):2287146. https://doi.org/10.1080/10298436.2023.2287146
45. Shah SFA, Chen B, Zahid M, Ahmad MR (2022) Compressive strength prediction of one-part alkali activated material enabled by interpretable machine learning. Constr Build Mater 360:129534. https://doi.org/10.1016/j.conbuildmat.2022.129534
46. Zhang J, Niu W, Yang Y, Hou D, Dong B (2022) Machine learning prediction models for compressive strength of calcined sludge-cement composites. Constr Build Mater 346:128442. https://doi.org/10.1016/j.conbuildmat.2022.128442
47. Quan Tran V, Quoc Dang V, Si Ho L (2022) Evaluating compressive strength of concrete made with recycled concrete aggregates using machine learning approach. Construct Build Mat 323:126578https://doi.org/10.1016/j.conbuildmat.2022.126578
48. Kashifi MT (2023) Investigating two-wheelers risk factors for severe crashes using an interpretable machine learning approach and SHAP analysis. IATSS Research 47(3):357–371. https://doi.org/10.1016/j.iatssr.2023.07.005
49. Ghafarian F, Wieland R, Lüttschwager D, Nendel C (2022) Application of extreme gradient boosting and SHapley Additive exPlanations to predict temperature regimes inside forests from standard open-field meteorological data. Environ Model Softw 156:105466. https://doi.org/10.1016/j.envsoft.2022.105466

Publisher's Note