Cleaner Engineering and Technology 29 (2025) 101109

Contents lists available at ScienceDirect

Cleaner Engineering and Technology

journal homepage: www.sciencedirect.com/journal/cleaner-engineering-and-technology

Utilizing agricultural wastes for fired bricks: A machine learning approach to compressive strength and water absorption predictions

Zahraa Jwaida [a,b,*], Luigi Di Sarno [a,c]

[a] School of Engineering, Department of Civil and Environmental Engineering, University of Liverpool, Liverpool, L69 7WS, UK
[b] Industrial Preparatory School of Vocational Education Department, Educational Directorate Babylon, Ministry of Education, Babylon, 51001, Iraq
[c] Department of Structures for Engineering and Architecture, University of Naples Federico II, 80125 Naples, Italy

ARTICLE INFO

Keywords:
Machine learning
Agricultural waste
Fired bricks
Performance
Compressive strength
Water absorption

ABSTRACT

The use of agricultural wastes in brick production is increasing due to their potential for sustainable construction and efficient waste utilization. Predicting the physical and mechanical properties of such bricks remains challenging because of complex interactions among process variables and waste materials. This study addresses this by developing predictive models using four machine learning (ML) algorithms, namely random forest regressor (RFR), extreme gradient boosting (XGBoost), artificial neural network (ANN), and ridge regression (RR), based on a dataset of 110 data points including bricks with agricultural wastes such as rice husk ash (RHA) and wheat husk (WH), along with the physical and processing parameters. The results indicate that all models show strong potential for predicting brick properties with optimized hyperparameters. RFR achieved the highest predictive performance ($R^2 = 0.879$ for compressive strength, 0.901 for water absorption), followed by XGBoost and ANN, which showed moderate predictive ability but signs of overfitting; RR performed the least effectively. SHapley Additive exPlanations (SHAP), Partial Dependence Plots (PDP), and Individual Conditional Expectation (ICE) plots revealed that manufacturing parameters were the most influential features. Sensitivity analysis showed that soil content (RMSE↑ 7.90), firing temperature (RMSE↑ 5.40), and brick size (RMSE↑ 4.95) had the highest impact, whereas waste additives exhibited low sensitivity (RMSE↑ < 2.0), supporting their sustainable inclusion. This study introduces a holistic workflow integrating predictive modeling, interpretable ML tools, and sensitivity analysis, as a template for materials science, highlighting its potential to optimize waste-based fired bricks and provide a transferable methodology for sustainable construction applications.

1. Introduction

Fired bricks represent building components that have been used for thousands of years, dating back to around 7000 BCE. The manufacturing process includes several steps: firstly, raw materials are optimized in a mixed design to obtain the design properties; secondly, raw materials are mixed and moulded by shaping or pressing; thirdly, drying and then burning of the products to high temperatures (up to 1200 °C) for obtaining the required mechanical and durability properties (Anton et al., 2021). Due to their relatively adequate mechanical properties and low cost, bricks are continually used in construction practices. However, with increasing environmental awareness, the focus has shifted from merely using durable materials to identifying eco-friendly alternatives that align with sustainable construction practices. The manufacturing process has evolved, although the primary raw materials remain the same, such as clay (Murmu et al., 2018). However, the manufacture contributes to environmental degradation and is considered non-eco-friendly due to its waste generation and high energy requirements (Hasan et al., 2021). It is best to utilize environmentally friendly and sustainable materials to minimise the impact on the environment. This has led researchers to explore alternative raw materials, particularly those derived from waste, to enhance sustainability in brick production.

Waste generation challenges are becoming increasingly prevalent due to economic and industrial growth (Al-Rbaihat et al., 2023). Global agricultural wastes contribute significantly to greenhouse gas emissions, which in turn contribute to climate change. The breakdown of organic materials, such as food waste and crop residues, represents the primary cause of these emissions. In terms of numbers, it was estimated that agricultural operations are responsible for about 13 % of greenhouse gas

* School of Engineering, Department of Civil and Environmental Engineering, University of Liverpool, Liverpool, L69 7WS, UK.
E-mail address: Z.A.Jwaida@liverpool.ac.uk (Z. Jwaida).

https://doi.org/10.1016/j.clet.2025.101109
Received 18 May 2025; Received in revised form 27 October 2025; Accepted 28 October 2025
Available online 3 November 2025

emissions worldwide, of which 5 % are attributable to the waste sector alone. Thus, waste management is essential and can be conducted in several ways, including recycling and reusing materials in the construction of buildings. This aims to reduce the amount of generated waste and the requirement of landfills, optimize usage, and promote sustainable practices (Parra-López et al., 2024; Pence et al., 2024; Lakhouit et al., 2025). By combining various waste materials, several researchers have attempted to create more environmentally friendly and sustainable bricks (Gencel et al., 2021a, 2021b; Muñoz et al., 2023). The utilization of waste materials to produce bricks can benefit from the reduction in the quantity of raw materials as well as the required energy in the combustion process. The use of waste in brick production offers a sustainable alternative to conventional materials. Several waste materials have been investigated as a replacement for soil or aggregates in bricks, such as marble stone powder (Sufian et al., 2021), agricultural wastes (Hassan, 2021), sawdust (Cultrone et al., 2020), slag and ferrous dust (Er et al., 2022), polymer wastes (Erdogmus et al., 2023), and other industrial wastes (Wahab et al., 2021; Abdel Hamid, 2021; PN et al., 2018).

Agricultural residues such as rice husk ash, wheat straw, sawdust, sugarcane bagasse ash, and groundnut shells have been widely investigated as partial replacements for clay or as organic additives in both fired and unfired bricks. These materials, often rich in silica and organic carbon, can improve thermal insulation, reduce firing energy due to their inherent calorific value, and lower the overall density of the bricks. In parallel, industrial wastes, including fly ash, bottom ash, blast furnace slag, silica fume, and cement kiln dust, have been successfully incorporated into geopolymer and conventional clay-based brick systems, offering enhanced compressive strength, durability, and resistance to chemical attack (Saravanan et al., 2023; Gupta et al., 2020). However, agricultural wastes were chosen for their wide availability, renewability, and minimal processing requirements compared to other types of waste. Their high silica and organic content enhance the properties of bricks, while their use supports both environmental sustainability and waste management, especially in rural areas.

While material innovations have advanced the sustainability of bricks, the complexity of their behaviour under various compositions and environmental conditions necessitates the development of efficient predictive tools. On the other hand, the utilization of computer-aided modelling is advancing for predicting the characteristics of building materials. Machine learning (ML) represents a significant branch of artificial intelligence that aims to create an algorithm to recognize intricate patterns in experimental or historical datasets without the need for a predetermined equation as a model, and provide comprehensive solutions and decisions (Chaabene et al., 2020). It can rapidly solve and model complex systems (Khambra et al., 2023). The general aim of ML is to build computer systems that can learn from experience and adjust to their surroundings. ML-based models can produce predictions and descriptions of knowledge acquisition from data (Rathnayaka et al., 2024). Conventional algorithms are time-consuming and challenging to implement for conducting the required tasks. To anticipate the characteristics of concrete (Asteris et al., 2021) or the failure of structural parts based on historical data (Taffese et al., 2017), ML has been expanded to be utilized in the built environment in various domains (Rachele et al., 2021).

Additionally, there are several types of ML, including supervised, semi-supervised, unsupervised, and reinforcement learning (Rathnayaka et al., 2024; Ivanović, 2015). The choice of a type mainly depends on the training resources. The two most prevalent forms of ML in a variety of fields, including engineering, are supervised and unsupervised learning (Taffese et al., 2017; Alrbai et al., 2024). A collection of learning examples is used in supervised learning, and each input, output value, or function is also described. The learning system aims to generate a hypothesis that provides an informed estimate of the relationship between input and output. Unsupervised learning, on the other hand, uses a set of learning examples where the only information known is the quantity of inputs, and no data is available about the correct output. Grouping inputs or forecasting the next value based on the current circumstances are two applications of unsupervised learning (Taffese et al., 2015, 2017). In an ML system, the tasks are as follows (Taffese et al., 2022; Ni et al., 2025): i) Classification in which the input variables are categorized, ii) Regression: this is used to model the relationship between input and output variables, iii) Prediction: this is used to forecast future results and represents a branch of regression, and iv) Clustering: this is used to identify patterns between two or more data sets. The goals specified for the three preceding steps (classification, regression, and prediction) are carried out using supervised methods; in contrast, clustering is carried out using an unsupervised approach. The various ML tasks are shown in Fig. 1. According to (Chaabene et al., 2020), ML modes can be divided into four primary models: artificial neural networks (ANN), support vector machines (SVM), decision trees (DT), and linear regressions (LR). These models are frequently used to forecast the mechanical characteristics of concrete. Despite the growing number of studies applying ML to concrete and other building materials, its application to waste-based fired bricks remains underexplored, especially in terms of model performance, generalizability, and interpretability. Thus, machine learning presents a promising yet underexplored approach to optimizing and predicting the performance of sustainable bricks made from agricultural waste.

1.1. The use of machine learning to predict the properties of waste bricks

A comprehensive review of the studies that have used ML to predict the properties of the bricks has been conducted. Several research engines were used, including Scopus, Web of Science, and Google Scholar. The review aims to explore studies that have been undertaken to predict the properties of various waste-based bricks, regardless of their type (e.g., fired brick, unfired brick, adobe, stabilized block, concrete brick, etc.). The goal is to identify and examine the broader applications of ML for bricks response analysis, providing a comprehensive overview while analysing existing gaps in the research. This approach was based on the assessment of the reported model performances and details of the methodology in the development of ML models, providing insights into the suitability of ML techniques for predicting the properties of bricks. Table 1 summarises the methods of the ML adopted by various researchers, while Table 2 shows performance indicators.

Several researchers have used the ANN model algorithm. Nakkeeran, Krishnaraj (Nakkeeran et al., 2024) and Obianyo, Auta (Obianyo et al., 2024) both reported excellent performance when using ANN. Nakkeeran, Krishnaraj (Nakkeeran et al., 2024) trained an ANN model using experimental data to predict the properties of bricks made from fly ash, cement, lime, and groundnut as fine aggregates (20–60 %). The results demonstrate the model's ability to predict compressive strength, dry density, and water absorption with a correlation coefficient of more than 0.99. In comparison, Ganasen, Krishnaraj (Ganasen et al., 2023) reported that ANN outperformed multiple linear regression (MLR) in predicting the properties of bricks made from rice husk and fly ash, emphasizing the advantage of ANNs in complex prediction tasks. Meanwhile, Lamba, Kaur (Lamba et al., 2024) extended this by utilizing Deep Neural Networks (DNN) for predicting water absorption and compressive strength in fly ash bricks, offering more profound insights into parameter impact. Bouzidi, Bouzidi and Quesada (Bouzidi et al., 2024) and Pamu and Svsndl (Pamu et al., 2024) also relied on ANN to predict various brick properties, with both studies noting excellent model performance to predict the water absorption and compressive properties of fired bricks and glass wool composite bricks, respectively. Khokhar, Khan (Khokhar et al., 2023) aimed at using various models to predict the properties of eco-efficient bricks incorporating waste glass. The authors reported that among the tested models, the ANN provided the highest accuracy, with an R^2 of 0.995 for shrinkage and 0.81 for compressive strength. Kifli, Absa and Musyafa (Kifli et al., 2017) reported less than 1 % error when using ANN for predicting lightweight

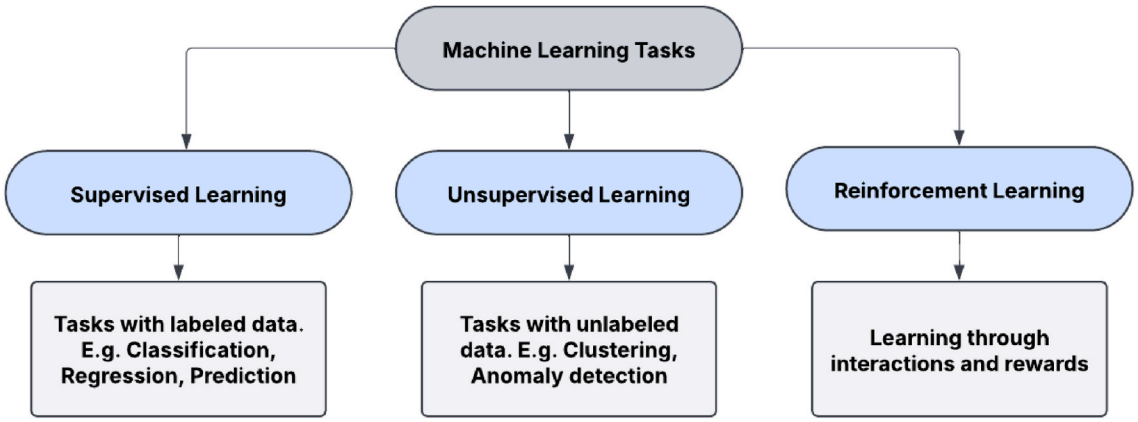

Fig. 1. Various machine learning tasks.

Table 1
ML summary of the method followed by various researchers.

Ref	Type of brick	Material used	Dataset number	Breakdown of the dataset	Model used	Input variables	Outputs	Statistics Index
Nakkeeran et al. (2024)	Bio bricks	groundnut shell (30 %–60 %), Cement-Fly Ash, and Hydrated Lime	–	70-15-15	ANN	Binder: fine aggregate, fly ash, hydrated lime, groundnut	compressive strength, water absorption, density	R^2
Obianyo et al. (2024)	Stabilized bricks	Lateritic soil, bone ash (0–6 %), palm bunch ash (0–6 %)	2268	80–20	14 model, best XGBoost	Soil and ashes percentages, curing age (days), Curing temperature (^oC), % Water	Compressive strength	R^2, RMSE
Praburanganathan et al. (2022)	Ash-based bricks	Fly ash, Granite tailings, and Glass waste (0–60 %), lime, gypsum, and stone dust	64	70–30	LR, lasso and ridge, DT, RF	curing time, glass waste, granite tailings, and fly ash	Compressive strength	R^2, MSE, RMSE
Vasić et al. (2023)	Ash-fired clay brick	Coal ashes, fly ash	303	Cross-validation	RFR, ANN	Na_2O, K_2O, CaO, Coal ash class, the share of the ashes, soaking time, firing temperature, furnace type.	Compressive strength, water absorption, and open porosity	R^2, RMSE
Ganasen et al. (2023)	Concrete bricks	Rice husk, hydrated lime, fly ash, cement, sand	14 for each test	–	ML, ANN, MLR	Cement, sand, rice husk, fly ash, hydrated lime	Compressive strength, water absorption, and Density	R^2, RMSE, MAE
Lamba et al. (2024)	Fly ash bricks	Waste plastic, fly ash, cement, sand			DNN	Water, fly ash, plastic, cement, and their content %	Compressive strength, water absorption	R^2, RMSE, MAE, MSE
Limami et al. (2023)	Unfired clay bricks	Plastic wastes	30	80–20	GPR, SVM, RFR	Additive percentages, brick size	Brick properties	MSE
Bouzidi et al. (2024)	Fired bricks	spent bleaching earth (0–50 %)	17 for each test	–	ANN	Water, spent bleaching earth (%), temperature	Compressive strength, water absorption, porosity, thermal conductivity	R^2, RMSE, MSE
Pamu et al. (2024)	Fired brick	Glass wool (4 %)	136	128-8 data points	ANN	glass wool replacement, temperature, load, and area	Compressive strength	R^2, MSE
Khokhar et al. (2023)	Fired brick	Waste glass	140	70–30	ANN, GPR, CART, SVM	glass (%), temperature (∘C), particle size (mm), and firing duration (hours)	Compressive strength, shrinkage	R^2, RMSE, MAE
Vasić et al. (2020)	Adobe clay bricks	heavy clays	139	60-20-20	ANN	Quantities of macro-oxides, mineralogical content, alevrolite- and clay-sized particles, residue on 0.063 mm sieve, and total carbonate content.	Shaping the moisture, plasticity coefficient, the shrinkage at Bigot's curve critical point, drying shrinkage, and compressive strength	R^2, RMSE
Sathiparan et al. (2023)	Cement-stabilized earth blocks	Cement (8–205), three soil types	180	–	LR, ANN, BTR, RFR, KNN, SVM	Cement%, electric resistivity, Ultrasonic pulse velocity	Compressive strength	R^2, RMSE

*ANN: Artificial Neural Network, XGBoost: Extreme Gradient Boosting, LR: Linear Regression, Lasso: Lasso Regression, Ridge: Ridge Regression, DT: Decision Tree, RF: Random Forest, RFR: Random Forest Regression, MLR: Multiple Linear Regression, DNN: Deep Neural Network, GPR: Gaussian Process Regression, SVM: Support Vector Machines, CART: Classification and Regression Tree, BTR: Boosted Tree Regression, KNN: K-Nearest Neighbours, R2: Coefficient of Determination, RMSE: Root Mean Squared Error, MAE: Mean Absolute Error, MSE: Mean Squared Error.

Table 2
Performance indexes of the models investigated by various researchers.

Ref	Type of brick	Model used	Outputs	R^2	MSE	MAE	RMSE
Nakkeeran et al. (2024)	Bio brick	ANN	Compressive strength	0.998	–	–	–
		ANN	Absorption	1.000	–	–	–
		ANN	Density	1.000	–	–	–
Obianyo et al. (2024)	Stabilized bricks	XGBoost	Compressive strength	0.995	–	–	0.060
Praburanganathan et al. (2022)	Ash-based bricks	LR	Compressive strength	0.690	0.830	–	0.910
		lasso and ridge regression	Compressive strength	0.620	1.030	–	0.912
		DT	Compressive strength	0.440	1.510	–	1.230
		RF	Compressive strength	0.900	0.440	–	0.660
Vasić et al. (2023)	Ash-fired clay brick	ANN	Compressive strength	0.968	–	–	3.506
		ANN	Water absorption	0.965	–	–	1.195
		ANN	Open porosity	0.980	–	–	1.420
		RFR	Compressive strength	0.756	–	–	10.655
		RFR	Water absorption	0.664	–	–	4.106
		RFR	Open porosity	0.668	–	–	6.475
Ganasen et al. (2023)	Concrete brick	ANN	Compressive strength	0.999	–	0.031	0.361
		ANN	Density	0.999	–	0.039	0.054
		ANN	Water absorption	0.999	–	5.526	7.033
		MLR	Compressive strength	0.986	–	0.038	0.392
		MLR	Density	0.977	–	0.026	1.283
		MLR	Water absorption	0.916	–	0.934	9.753
Lamba et al. (2024)	Fly ash brick	DNN	Compressive strength	0.854	–	0.036	0.175
		DNN	Water absorption	0.860	0.147	0.102	–
Bouzidi et al. (2024)	Fired bricks	ANN	Compressive strength	0.999	0.177	–	0.421
		ANN	Water absorption	0.999	0.032	–	0.180
		ANN	Thermal conductivity	0.999	0.000	–	0.001
		ANN	Porosity	0.994	0.220	–	0.469
Pamu et al. (2024)	Fired brick	ANN	Compressive strength	0.990	0.100	–	–
Khokhar et al. (2023)	Fired brick	SVM	Compressive strength	0.800	–	3.867	4.651
		SVM	Shrinkage	0.750	–	0.637	1.105
		CART	Compressive strength	0.730	–	4.271	4.980
		CART	Shrinkage	0.741	–	0.875	1.482
		ANN	Compressive strength	0.813	–	3.457	4.212
		ANN	Shrinkage	0.815	–	0.720	0.952
		GPR	Compressive strength	0.815	–	3.689	4.386
		GPR	Shrinkage	0.747	–	0.641	1.119
Vasić et al. (2020)	Adobe clay bricks	ANN	Shaping moist	0.506	–	–	1.702
		ANN	Plasticity coefficient	0.692	–	–	2.137
		ANN	Shrinkage – critical	0.812	–	–	0.607
		ANN	Loss of weight -critical	0.656	–	–	1.285
		ANN	Drying shrinkage	0.876	–	–	0.708
		ANN	Compressive strength	0.757	–	–	1.217
		ANN	Compressive strength	0.731	–	–	2.298
Sathiparan et al. (2023)	Cement-stabilized earth blocks	LR	Compressive strength	0.803	–	–	0.511
		ANN	Compressive strength	0.952	–	–	0.251
		RFR	Compressive strength	0.956	–	–	0.240
		BTR	Compressive strength	0.960	–	–	0.225
		KNN	Compressive strength	0.928	–	–	0.313
		SVM	Compressive strength	0.943	–	–	0.275

brick properties, demonstrating the high accuracy of ANN models in specific applications.

Additional ML algorithms have also been investigated. Limami, Guettioui (Limami et al., 2023) aimed to develop a model to predict the properties of unfired brick with plastic wastes through regression and reported an error percentage of 1 %. Obianyo, Auta (Obianyo et al., 2024) predicted the compressive strength of stabilized brick with bone ash and palm ash using various models. The Extreme Gradient Boosting Regressor achieved optimum performance, reporting a coefficient of correlation of 0.9945 and a root mean square error of 0.06. Praburanganathan, Chithra and Simha reddy (Praburanganathan et al., 2022) Random Forest was found to be the most effective model for predicting compressive strength among the models tested, including linear regression and decision trees, highlighting the importance of choosing the right model based on data characteristics. Similarly, Vasić, Jantunen (Vasić et al., 2023) used Random Forest Regressor and ANN to predict the compressive strength, water absorption, and porosity of ash-bricks (coal, fly ash, pond ash, and bottom ash) on a dataset collected from the literature of 20 years of research. The study analysed the impact of various parameters on the investigated properties and concluded that ANN models provided the best fit of the data. Similarly, Sathiparan and Jeyananthan (Sathiparan et al., 2023) predicted the compressive strength of cement-stabilized earth blocks with various contents of cement, ultrasonic pulse velocity, and electrical resistivity. Boosted tree regression and ANN were reported to perform better than other tested models.

Other researchers have investigated the use of ML for developing models for mortar-brick composites with the prediction of shear strength (Morsali et al., 2020) or improved elastic deformation resistance (Demir et al., 2017). In addition, Pezo, Arsenović and Radojević (Pezo et al., 2014) calculated the mineral content of raw materials and used ANN to analyse the properties of brick and reported an R^2 ranging between 0.869 and 0.0.984. In another study, Vasić, Pezo and Radojević (Vasić et al., 2020) studied the granulometry and composition of various types of clays on adobe bricks through ANN models and reported R^2 ranging from 0.58 to 0.907, indicating that model performance can vary significantly depending on the type of brick and dataset used. Marín-García, Bienvenido-Huertas (Marín-García et al., 2023) performed DNN using images of brick facades to identify efflorescence and reported the significance of the model in creating a map of exposed brick facades.

Despite significant advancements in using ML to predict brick properties, several gaps remain in the research. One key issue is the

limited exploration of non-ANN models, such as Random Forest and XGBoost, which may offer competitive performance for predicting various brick types. Moreover, the lack of standardized evaluation metrics across studies complicates comparisons between models. Model interpretability is another critical feature, as many ML models, particularly ANNs and Deep Neural Networks, are often used as “black boxes," making it difficult to understand their decision-making process. This is a major concern in applications such as construction, where transparency is crucial. Furthermore, the real-world applicability of these models remains underexplored, particularly in terms of cost, scalability, and adaptability in industrial settings. Addressing these gaps in future research will lead to more robust, transparent, and practical ML models that can be effectively applied in the construction and materials science industries.

1.2. Significance of the research

This study introduces a novel application of four ML algorithms to predict compressive strength and water absorption of fired bricks made from agricultural waste materials. The models used are ANN (Artificial Neural Network), XGBoost (Extreme Gradient Boosting), RFR (Random Forest Regressor), and RR (Ridge Regression). The use of ML to investigate the impact of several waste materials on the prediction of mechanical or physical properties has been studied in concrete (Sifan et al., 2023; Anand et al., 2025; Golafshani et al., 2024a), though the approach has not been explored in bricks. The integration of agricultural by-products such as rice husk, wheat straw, and bagasse offers a promising solution to reduce the environmental impacts of brick production. This study not only addresses the growing demand for eco-friendly construction materials but also offers a structured and transparent ML framework that combines predictive modelling with tools for understanding and interpreting model behaviour, creating a methodological reference for materials research. The core contribution lies in evaluating the predictive performance of commonly used ML models for key properties of agricultural waste-based bricks using experimental data from previous studies. SHAP analysis, sensitivity analysis, partial dependence plots (PDP), and individual conditional expectation (ICE) were used to interpret the effects of individual variables on the prediction outcomes. Performance metrics, including Mean Squared Error (MSE), Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE), were used to evaluate the models. This research provides a practical tool for predicting the performance of agricultural waste-based fired bricks, thereby contributing to the development of more sustainable and cost-effective building materials and promoting resource-efficient, environmentally responsible construction practices in fired brick production. By focusing on agricultural waste-based bricks, this study addresses some of the analysed gaps while also pointing to areas that require further research to enhance the practical application of ML models in the construction industry.

2. Methodology

The modelling processes use the Anaconda-based Python programming (version 3.13) installed on an *HP* containing a built-in memory of 8 GB. It is a potent tool for data research (Al-jabery et al., 2020). The following sections provide a summary of the steps used for developing the models. The methodology of the study is shown in Fig. 2.

2.1. Dataset

A reliable dataset is crucial for successfully implementing models that predict the compressive strength and water absorption of fired bricks. Thus, a comprehensive literature review was carried out to extract the necessary data for the manufacturing process of bricks, along with their performance results in terms of compressive strength and water absorption. Only reliable research articles were used for this purpose. Table 3 shows a summary of the resources for the dataset. A total of 110 data points were used. Several inputs were evaluated during the initial testing phase. However, given the significance of these features (refer to Fig. 3), only the primary factors about water absorption and compressive strength of fired bricks were included in the analysis. A spreadsheet was employed to systematically document the input variables, encompassing the type, content, and particle sizes of waste materials, moisture content, soil content, brick dimensions, and firing temperatures. A possible challenge in the formulation of predictive machine learning models for compressive strength and water absorption of fired bricks utilizing agricultural wastes may arise from the variability in data collected from diverse studies. Disparities in material properties, waste types, processing techniques, and testing standards result in inconsistencies that can substantially impact model performance and generalizability. To mitigate these concerns, a comprehensive data preprocessing included the normalization of units and the imputation of missing values. Moreover, issues related to imbalanced data distributions were addressed by implementing advanced validation methodologies, such as k-fold cross-validation, to enhance model robustness and reduce bias. The study assumes that the dataset reflects real-world brick production using agricultural waste, with independent and sufficient input features, free from measurement errors. Standardized testing procedures and homogeneous material composition were maintained, while extreme outliers were excluded. External factors like long-term durability and environmental exposure were beyond the model scope. These controlled assumptions and boundaries ensure the models are both reliable and applicable to practical, sustainable brick manufacturing.

2.2. Data visualization

The dataset’s summary statistics, including key metrics and visualizations, are presented in Table 4 and Figs. 4–7. The dataset can be

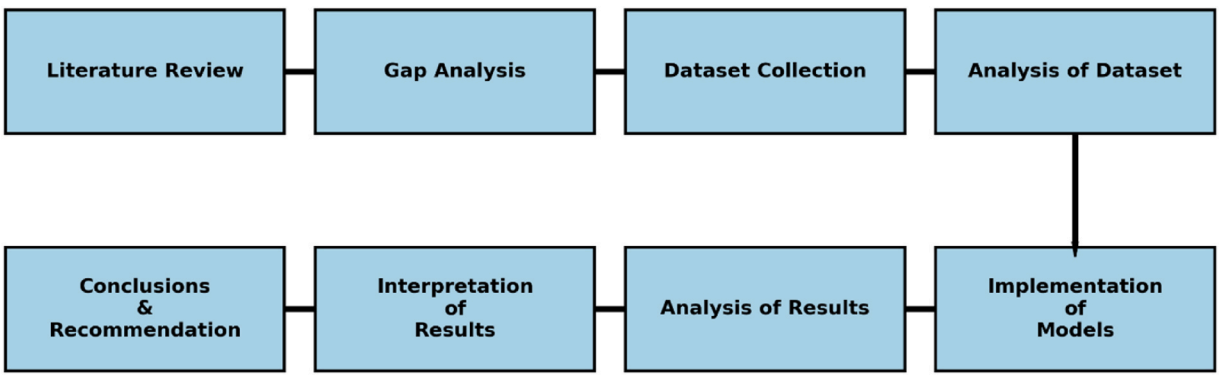

Fig. 2. Methodology of the research.

Table 3
Summary of the resources for the dataset used in the research.

Ref.	Type of waste	NO. of data points	Water absorption (%)	Compressive strength (MPa)
Eliche-Quesada et al. (2017)	Rice husk ash	7	16.9–32.7	17.5–53.4
De Silva et al. (2018)	Rice husk ash	7	22.0–26.7	2.6-1.9
Adazabra et al. (2017)	Spent shea waste	20	3.0–33.0	4.0–13.0
Ahmad et al. (2017)	Wheat Husk	8	15.0–35.0	9.8–25.0
Kazmi et al. (2018)	Sugarcane bagasse ash	4	15.3–23.9	5.0–9.4
Kazmi et al. (2018)	Rice husk ash	3	17.7–20.5	5.5–7.2
Ahmed et al. (2023)	Pomegranate peel waste	5	13.8–20.2	4.6–18.5
Kakkar et al. (2023)	Rice straw ash	6	3.5–20.0	3.1–13.9
Agbede et al. (2011)	Rice husk ash	5	14.8–19.0	6.9–18.6
Djamaluddin et al. (2021)	Palm oil fuel ash	4	23.0–32.0	2.5–7.3
Demir (2006)	waste tea	4	22.5–35.0	19.5–34.3
Kizinievič et al. (2018)	oat husk	6	14.0–43.0	0.8-7.5
Kizinievič et al. (2018)	barley husk	6	14.0–49.0	0.3–9.5
Heniegal et al. (2020)	Rice straw ash	2	14.0–17.0	12.0–15.0
Heniegal et al. (2020)	Sugarcane bagasse ash	2	15.0–19.0	15.0–19.0
Heniegal et al. (2020)	wheat straw ash	2	15.0–21.0	7.0–12.0
Srisuwan et al. (2018)	Coconut husk	3	12.0–19.0	4.0–10.0
Moujoud et al. (2023)	coconut shell	9	16.0–46.0	4.5–16.0
Sarani et al. (2023)	Palm Kernel Shell	3	8.0–12.0	11.0–19.8
Ghorbani et al. (2021)	Potato peels	3	25.0–31.0	2.0–7.0

visualized to gain a better understanding of its nature. For this, a correlation matrix was used, which shows the relationship between all variables. The matrix for the collected dataset is shown in Fig. 4. This step is crucial for understanding the impact of each parameter on the others. The values of the correlation range between −1 and 1. A negative value indicates a negative correlation, i.e., when a variable increases, the other decreases. A positive value indicates a positive correlation, i.e., the two variables change in the same manner. If the correlation value is 0, it means that there is no relationship between the two variables. The matrix uses contour graphs to show the relationship between the variables. The highest concentration of each variable, as determined by the data samples, is displayed in the darkest area. The relative maximum distribution of the complete data, which strengthens the connection, is shown in the graph's rectangular inside portion. For example, the coefficient of correlation of 0.52 between the size of brick and moisture content indicates a good correlation between them, i.e., the increase in the size of bricks will impact the moisture content. The low values (ex, −0.02 to −0.09) suggest that there is no minor link between their respective features.

The statistical parameters of the dataset, including minimum, maximum, mean, and standard deviation values, are presented in Table 4. Additionally, Figs. 5 and 6 display the standard deviation of all variables and the distribution of mean values by type of waste, respectively. A closer examination of the mean and standard deviation values reveals notable differences, indicating that the selection of material has a significant influence on brick characteristics. The average amount of waste that is mixed with the bricks is 10.05 %, with a comparatively large standard deviation of 9.17 %. This shows a wide variety of waste content across various materials, with mean values ranging from 3.75 % (for PP and CH) to 21.25 % (for WH). The average particle size is 244.24 μm, with a standard deviation of 112.27 μm. This suggests that the texture of the materials employed varies greatly. Some waste materials may have finer particles that improve the uniformity of the mix, while others may have coarser particles. The soil content remains a dominant component in brick production, with a mean of 87.09 wt% and a standard deviation of 13.21 wt%. Moderate variability in soil content suggests different mixing ratios may be required depending on the waste material used. Similarly, the moisture content has a mean value of 12.98 wt%, with a high standard deviation of 10.11 wt%. This significant variation indicates that some waste materials contribute more moisture than others, which may affect the drying process and overall workability of the bricks. The size of the bricks varies considerably, with a mean of 0.54 m^3 and a standard deviation of 0.71 m^3. The high standard deviation suggests inconsistent brick sizes, potentially due to differences in production methods or the influence of waste material properties on shaping and compaction. Additionally, the firing temperatures required for brick production are relatively consistent, with a mean of 958.05 °C and a standard deviation of 100.84 °C.

There are variations in the characteristics, such as water absorption and compressive strength. With a standard deviation of 8.18 MPa and a mean compressive strength of 10.94 MPa, some waste materials may increase the strength of bricks, while others may decrease it. Similarly, the mean water absorption is 20.31 % with a standard deviation of 8.85 %, indicating that brick porosity varies according to the kind and quantity of waste material employed. While lower values indicate denser bricks, higher water absorption values signal more porosity, which may affect durability.

The analysis of the dataset (Fig. 7) highlights the varied use of different waste materials in brick production. RHA and SShW appear most frequently, with 19 and 16 data points, respectively, suggesting their strong relevance in sustainable construction practices. Other materials such as CSh, WH, and RSA also show consistent representation, reflecting a diverse approach to utilizing agricultural and industrial by-products. Materials like PKSh, PP, and WSA, while appearing less frequently, still contribute to the dataset and demonstrate the broad potential of alternative resources in brickmaking. This distribution reflects ongoing efforts to explore a wide range of eco-friendly materials, with opportunities to expand research across all categories further.

2.3. Processing of data

Before training the models, the data should be normalized into a uniform scale. For ML algorithms, this is crucial, particularly for ones that are sensitive to variations in feature sizes. It can prevent difficulties and supremacy due to the number or variability of the data, as well as improve the compatibility of the data. Additionally, the overfitting issue can be avoided by using data normalization, as various variables have distinct dimensions that could affect the analysis of the data. Several normalization techniques were considered during the data preprocessing phase to address scale disparities among features and ensure compatibility with machine learning algorithms sensitive to feature magnitude. Techniques such as min-max scaling, robust scaling, and quantile transformation were evaluated for their ability to mitigate the impact of skewed distributions and outliers. However, following initial data analysis, the input features did not show significant skewness or extreme outliers that would necessitate more complicated transformations. Consequently, the data were normalized using StandardScaler from the Sklearn module, ensuring that data points have a balanced scale. The StandardScaler normalizes the data by adjusting it to have a mean of 0 and a standard deviation of 1. It processes each feature (column in the dataset) by subtracting the mean of that feature and dividing by its standard deviation. Without scaling, models may be influenced disproportionately by features with broader ranges, leading

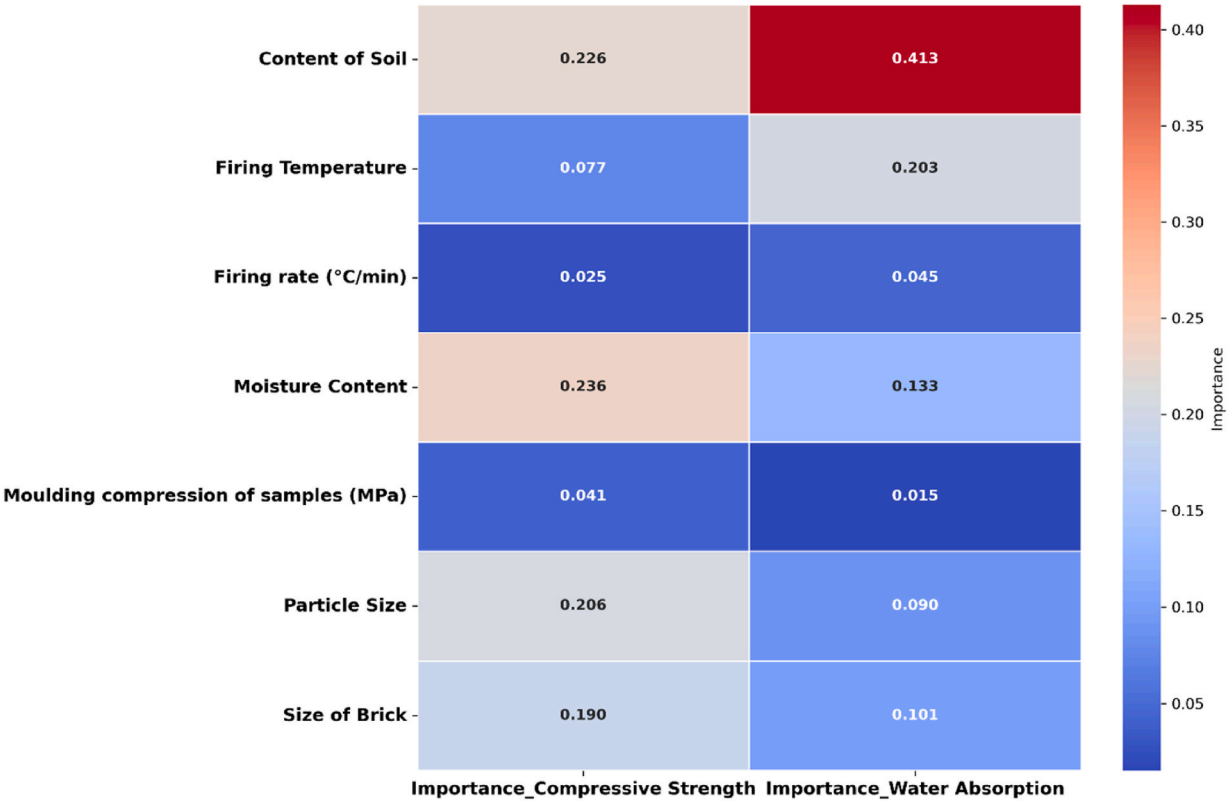

Fig. 3. Significance of various input features.

Table 4
Summary of the statistics of the dataset.

Item	Notation	Unit	Type of data	Min	Max	Mean	Std
Barely Husk	BH	wt.%	Input	0	20	11.67	6.83
Coconut Husk	CH	wt.%	Input	0	7.5	3.75	3.23
Coconut Shell	CSh	wt.%	Input	0	30	16.36	11.20
Oat Husk	OH	wt.%	Input	0	20	8.75	7.91
Palm Kernel Shell	PKSh	wt.%	Input	0	10	4.00	4.55
Palm Oil Fuel Ash	POFA	wt.%	Input	0	20	10.00	7.91
Potato Peels	PP	wt.%	Input	0	7	3.75	2.99
Pomegranate Peel Waste	PPW	wt.%	Input	0	15	7.50	5.59
Rice Husk Ash	RHA	wt.%	Input	0	30	10.50	8.54
Rice Straw Ash	RSA	wt.%	Input	0	15	6.25	4.74
Sugarcane Bagasse ash	SBA	wt.%	Input	0	15	8.33	6.06
Spent Shea Waste	SShW	wt.%	Input	0	20	10.00	7.25
Wheat Husk	WH	wt.%	Input	0	50	21.25	17.47
Waste Straw Ash	WSA	wt.%	Input	0	15	10.00	7.07
Waste Tea	WT	wt.%	Input	0	12.5	4.64	5.67
Particle size	–	μm	Input	63	500	244.24	112.27
Content of soil	–	wt.%	Input	50	100	87.09	13.21
Moisture content	–	wt.%	Input	0.135	40	12.98	10.11
Size of brick	–	m^3	Input	0.003	1.99	0.54	0.71
Firing temperature	–	^oC	Input	800	1200	958.05	100.84
Water absorption	–	%	output	3	49	20.31	8.85
Compressive strength	–	MPa	output	0.3	35.9	10.94	8.18

to inaccurate predictions and inferior outcomes. This is especially important for algorithms that depend on distance calculations and optimization methods, such as the models chosen for the analysis. This methodology was deemed adequate for achieving numerical stability and consistent model efficiency while preserving the interpretability and integrity of the original data distribution.

2.4. Implementation of the models

The principles of the algorithms for the selected models are presented in Appendix A (section A.1 and Figure A.1). The methods for implementing the models are shown in Figure A.2 (Appendix A, section A.2) while the principle of cross-validation is shown in Figure A.3 (Appendix A, section A.2). The number 10 for K-fold was assessed by various researchers and was found to be more effective than any other number of folds (Khan et al., 2021; Lyu et al., 2022; Ling et al., 2019; Nguyen et al., 2024). The distinction between the sole K-fold cross-validation and a train-test split with K-fold cross-validation lies primarily in the way data is partitioned and evaluated. For the K-fold cross-validation, the complete dataset is divided into ten equal

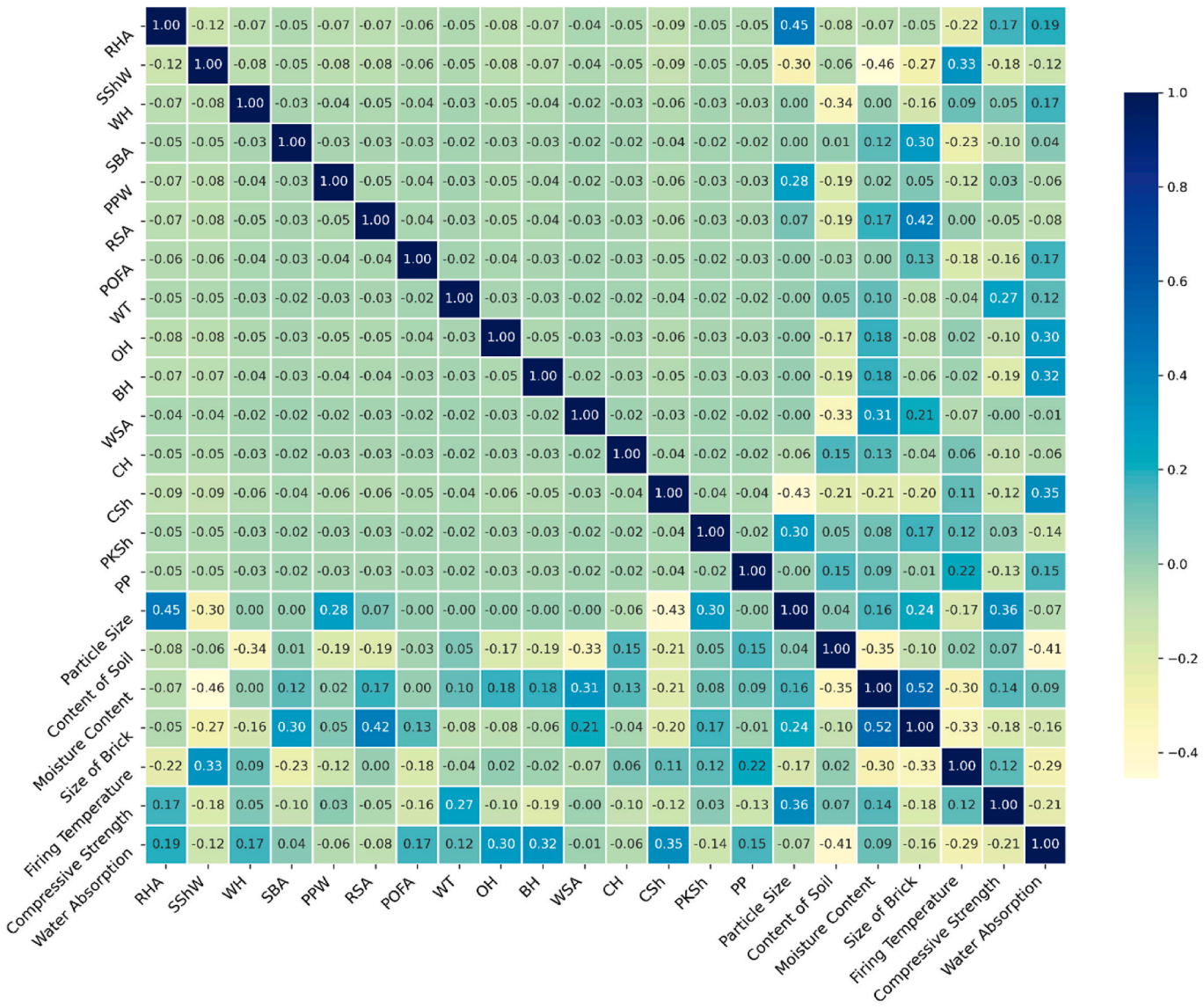

Fig. 4. Correlation matrix between inputs and outputs.

subgroups, or “folds." Each fold is utilized as the test set once, and the remaining nine folds are used for training, before the model is trained and assessed ten times. A train-test split, on the other hand, proceeds by dividing the dataset into separate training and test sets using K-fold cross-validation. The dataset was randomly divided into fixed ratios of training and testing sets, with 80 % allocated to the training set and 20 % to the testing set. Only the training set is subjected to the 10-fold cross-validation after the training data has been further split into 10 folds for model evaluation. The test set, which remains distinct throughout the procedure, is only used for the final assessment of the model’s performance after cross-validation is complete. This approach ensures that the model is trained on a substantial portion of the data. In contrast, the remaining data is reserved for validating the model’s performance during training and evaluating its final accuracy and generalization on unseen data. The approach restricts cross-validation to the training subset, reserving the test set for a final assessment. Other researchers have widely explored the same method (Chuluunsaikhan et al., 2021; Wong et al., 2022; Tipu et al., 2022).

Python’s scikit-learn library was used to implement RFR and RR, while XGBoost and Keras Tuner Libraries were used to implement XGBoost and ANN, respectively. The tuning of hyperparameters is essential since the performance of the algorithm depends on the setting of its hyperparameters (Probst et al., 2019). Initially, a set of values for one hyperparameter was pre-established, and the ML model was evaluated for various combinations of hyperparameter values to determine the best combination. Until the ideal combination of the best performance was discovered, this process was repeated. To find the perfect set of hyperparameters for a specific task, methods such as grid search, random search, and Bayesian optimization are frequently employed. For the study, GridSearchCV was used to optimize models, including the Random Forest Regressor and Ridge Regression. XGBoost integrates with GridSearchCV (from scikit-learn) and offers its optimization methods. For Artificial Neural Network, Keras uses Keras Tuner for hyperparameter tuning. To balance model complexity and avoid overfitting, the Random Forest Regressor (RFR) requires optimization of crucial parameters: n_estimators, max_depth, min_samples_leaf, and min_samples_split. Furthermore, bootstrapping regulates the diversity and randomness of trees, enhancing model performance. In XGBoost, subsample and colsample_bytree enhance generalization, while learning rate, n_estimators, and max_depth are tuned to control model complexity. While hidden layers regulate model complexity and prevent overfitting, learning rate, batch size, maximum iterations, activation function, and solver affect convergence in Artificial Neural Networks (ANNs). Alpha (for regularization strength) and solver are the primary

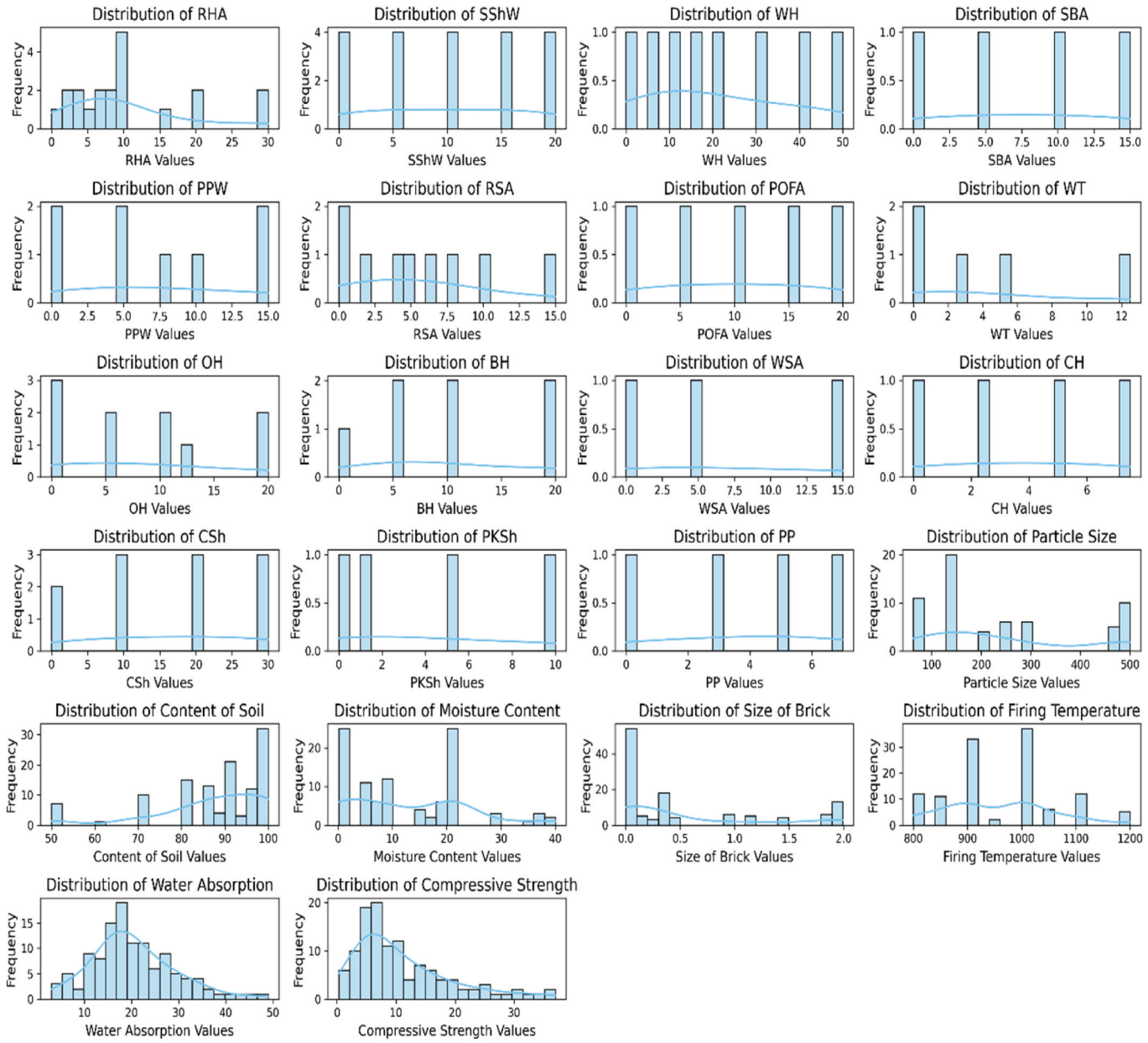

Fig. 5. Standard deviation of all input and output variables.

hyperparameters in Ridge Regression (RR) that need to be optimized. The range considered for the hyperparameter tuning is shown in Table 5. Training times varied notably based on model complexity. ANN and XGBoost were the most computationally intensive, requiring approximately 20–40 min for a hyperparameter optimization run due to iterative learning and tree boosting processes, respectively. In contrast, Ridge Regression and Random Forest models completed training significantly faster, typically under 15 min, making them more efficient for rapid deployment in practical applications.

2.5. Statistics index

Several performance metrics are used to measure the predictive efficacy of ML models, enabling accurate assessment. These metrics help evaluate the model's accuracy in forecasting the target variable and assess the degree to which its forecasts align with actual values (Sathiparan et al., 2025). R^2 (Coefficient of Determination) indicates the amount of the variance in the dependent variable that is explained by the model. It ranges from 0 to 1, with a value closer to 1 indicating a better fit (Sathiparan et al., 2023). The Mean Squared Error (MSE) measures the average squared difference between the actual and predicted values, indicating the magnitude of prediction errors (Kalabarige et al., 2024). When the MSE is squared, the resultant is the Root Mean Squared Error (RMSE), which.

RMSE is the square root of MSE and provides a measure of error in the same units as the target variable (Golafshani et al., 2024b). The average of the absolute discrepancies between the expected and actual values is determined by the Mean Absolute Error (MAE). With lower values indicating higher performance, it provides a clear indicator of how closely predictions align with actual results (Golafshani et al., 2024b). The equations for R^2, MSE, MAE, and RMSE are:

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (yi - \hat{yi})2}{\sum_{i=1}^{n} (yi - \bar{y})2} \quad (1)$$

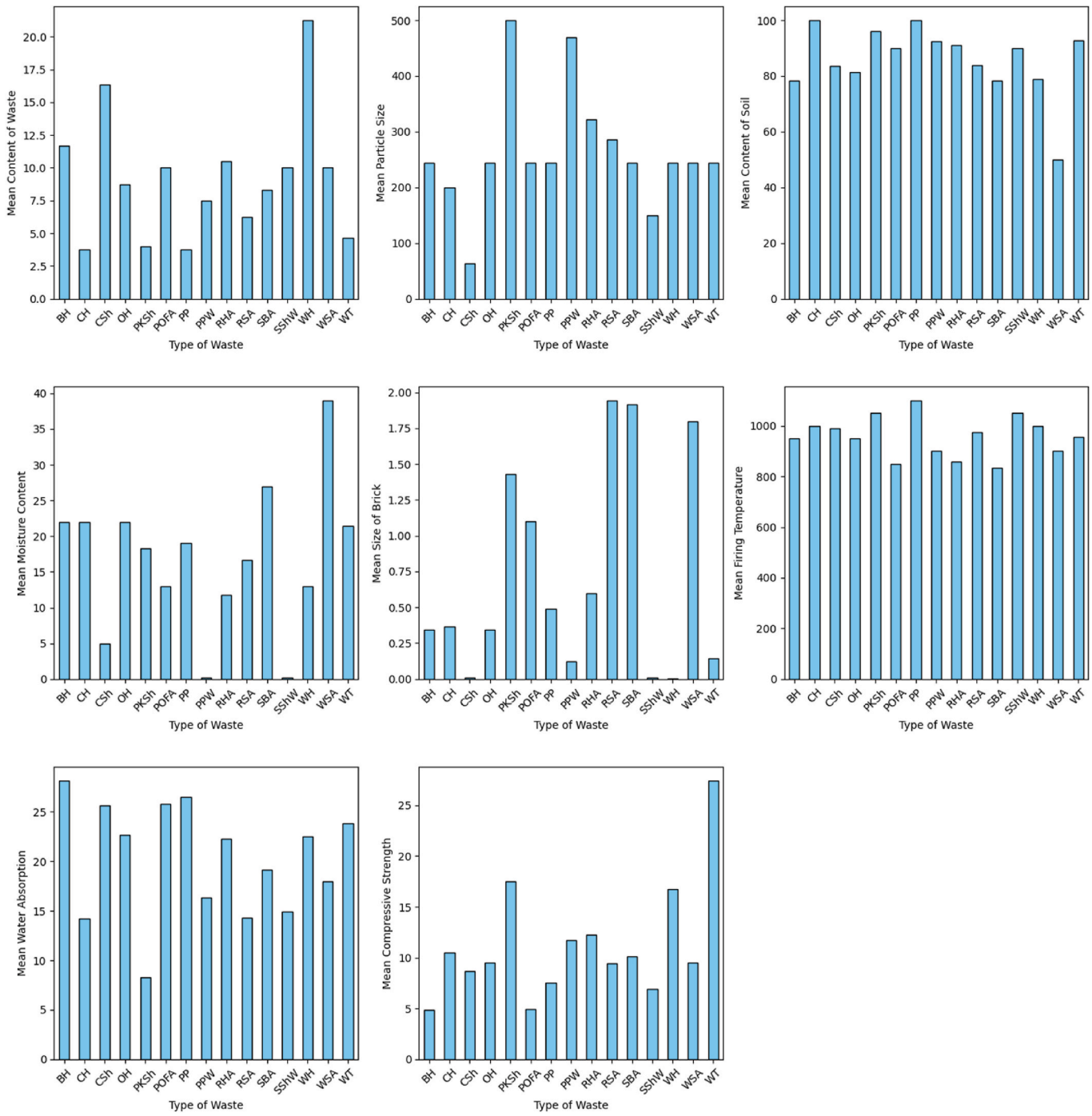

Fig. 6. Distribution of the mean values of variables by the type of waste.

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(yi - \hat{yi})2 \tag{2}$$

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|yi - \hat{yi}\right| \tag{3}$$

$$\mathrm{RMSE} = \sqrt{MSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(yi - \hat{yi})2} \tag{4}$$

Where:

yi is the actual value for the i-th data point, $\hat{yi}$ is the predicted value for the i-th data point, $\bar{y}$ is the mean of the actual values, and n is the number of data points.

The prediction accuracy of a model can be thoroughly understood using performance metrics such as R^2, MSE, RMSE, and MAE. A higher R^2 denotes a better fit, as it indicates that the model accounts for a larger percentage of the variation. The model is more accurate when the MSE, RMSE, and MAE values are smaller. RMSE provides a more comprehensible error metric in the same unit as the target variable. By showing the size of prediction mistakes as well as the model's capacity to provide accurate predictions within reasonable bounds, these metrics collectively enable a comprehensive assessment of the model's efficacy.

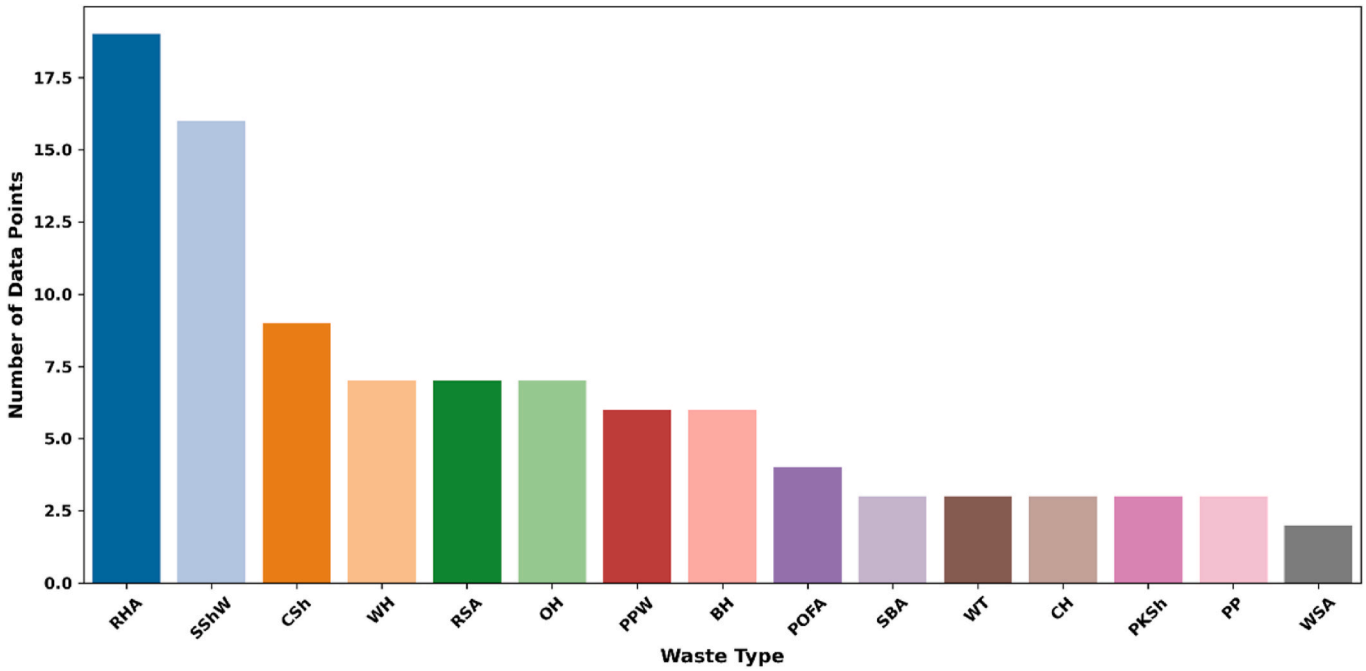

Fig. 7. Distribution of the data points by the type of waste.

Table 5
Search ranges for hyperparameters.

Model	Ranges
RFR	n_estimators: 50, 100, 150, 200; Max depth: 10, 15, 20, None; Min samples split: 2, 5, 10; Min samples leaf: 1, 2, 4; Bootstrap: True, False
XGBoost	Learning rate: 0.01, 0.1, 0.2, 0.3; Max depth: 3, 5, 7; n_estimators: 100, 150, 200; Subsample: 0.6, 0.7, 0.8, 1.0; Colsample_bytree: 0.6, 0.7, 0.8, 1.0
ANN	Hidden layer sizes: (50), (100), (150), (100,50); Activation: 'Relu', 'Tanh', 'Logistic'; Solver: 'Adam', 'Sgd'; Alpha: 0.0001, 0.001, 0.01; Learning rate: 'Constant', 'Adaptive'; Max iterations: 2000, 5000, 10000; Batch size: 'Auto', 16, 32
RR	Alpha: 0.1, 1, 10, 100; Solver: 'Auto', 'Svd', 'Cholesky', 'Lsqr'

2.6. Interpretation of models' performance

SHAP (SHapley Additive exPlanations) is a method rooted in game theory designed to explain the predictions made by ML models. SHAP analysis was conducted to obtain the contribution of all input parameters to the prediction of water absorption and compressive strength values in an ML model. To illustrate how each input variable affects the model's output, SHAP assigns a specific contribution to each feature in a model through the computation of Shapley values, which have their roots in cooperative game theory. By comparing the model's output when a feature is present against when it is absent, considering all possible feature permutations, the Shapley value quantifies the importance of a feature. This procedure ensures that each feature's contribution is equitably allocated and considered in relation to the baseline prediction, providing a clear and intelligible means of understanding model choices (Huang et al., 2025; Lundberg et al., 2020). Meanwhile, A sensitivity analysis was performed on the optimized model to assess the impact of individual input variables on compressive strength prediction. Each feature in the test set was systematically perturbed using five methods: mean, zero, shuffle, minimum, and maximum substitution. Model predictions after perturbation were evaluated using R^2, MSE, RMSE, and MAE. This analysis, conducted only on the test data, provided insights into model robustness and variable importance without retraining (Sarkar et al., 2024; Nguyen et al., 2023).

Other techniques are available to assist in interpreting the results, such as individual conditional expectation (ICE) plots and partial dependence plots (PDP). Based on a given dataset, the PDP provides the average effect of an input on the output, illustrating its partial dependence while accounting for the impacts of other inputs. While the ICE separates the averaged results in the PDP to indicate the response of a particular input value for each observation (Setvati et al., 2022; Goldstein et al., 2015).

3. Results and discussion

3.1. Performance of models

The best combinations of hyperparameters of each ML model to predict the compressive strength and water absorption of fired bricks are listed in Table 6. Every ML model contains distinct hyperparameters that are sensitive to determining how well the model performs. To guarantee reliable and accurate forecasts, it is crucial to achieve a balance between model complexity and performance. Consequently, the best hyperparameter combinations for each model were found.

The performance of the models with the optimal hyperparameters

Table 6
Best hyperparameter combinations of all models.

Models	Best hyperparameters
Compressive strength	
ANN	Hidden-layer-sizes: (100), max-iterations: 10000, activation function: tanh, Solver: Adam, Alpha: 0.0001, batch-size: auto, learning rate: Constant
RFR	bootstrap: True, max-depth: 20, min-samples-leaf: 1, min-samples-split: 2, n-estimators: 100
XGBoost	colsample-by tree: 1, learning-rate: 0.2, max-depth: 3, n-estimators: 150, subsample: 0.7
RR	Best Ridge Alpha: 10, Solver: Lsqr
Water absorption	
ANN	Hidden-layer-sizes: (150), max-iterations: 10000, activation function: Relu, Solver: Adam, Alpha: 0.001, batch-size: Auto
RFR	bootstrap: True, max-depth: 20, min-samples-leaf: 1, min-samples-split: 2, n-estimators: 150
XGBoost	colsample-bytree: 0.7, learning-rate: 0.2, max-depth: 3, n-estimators: 150, subsample: 0.8
RR	Best Ridge Alpha: 10, Solver: Auto, learning rate: Constant

Table 7
Performance results of various models.

Compressive Strength						
Models		R^2	MSE	MAE	RMSE	10-K folds
RFR	Training	0.933	4.079	1.381	2.020	Std. = 0.053, 95 %
	Testing	0.824	15.073	2.991	3.882	CI = [0.822, 0.902]
	All	0.879	9.576	2.186	2.951	
XGBoost	Training	0.991	0.559	0.567	0.748	Std. = 0.067, 95 %
	Testing	0.797	17.404	2.767	4.172	CI = [0.834, 0.936]
	All	0.894	8.982	1.667	2.460	
ANN	Training	0.902	6.028	1.549	2.455	Std. = 0.088, 95 %
	Testing	0.762	20.417	3.007	4.519	CI = [0.714, 0.856]
	All	0.832	13.223	2.278	3.487	
RR	Training	0.560	26.944	3.628	5.191	Std. = 0.074, 95 %
	Testing	0.638	31.060	3.654	5.573	CI = [0.506, 0.549]
	All	0.599	29.002	3.641	5.382	
Water Absorption						
Models		R^2	MSE	MAE	RMSE	10-K folds
RFR	Training	0.949	4.214	1.446	2.053	Std. = 0.044, 95 %
	Testing	0.853	8.841	2.545	2.973	CI = [0.840, 0.920]
	All	0.901	6.527	1.996	2.513	
XGBoost	Training	0.991	0.719	0.656	0.848	Std. = 0.057, 95 %
	Testing	0.782	13.126	2.821	3.623	CI = [0.811, 0.909]
	All	0.886	6.923	1.739	2.236	
ANN	Training	0.919	6.643	1.943	2.578	Std. = 0.074, 95 %
	Testing	0.721	16.774	3.116	4.096	CI = [0.714, 0.853]
	All	0.820	11.709	2.529	3.337	
RR	Training	0.818	14.897	2.995	3.860	Std. = 0.069, 95 %
	Testing	0.741	15.541	3.020	3.942	CI = [0.690, 0.810]
	All	0.780	15.219	3.007	3.901	

was analysed in terms of R^2, MSE, MAE, and RMSE. The results are presented in Table 7 and Figs. 8–11. The evaluation of compressive strength across the overall performance indicated that RFR and XGBoost were the highest-performing models, followed by ANN and finally RR. ANN also exhibits good performance, yielding an R^2 of 0.832, an MSE of 13.223, an MAE of 2.278, and an RMSE of 3.487. XGBoost achieves an overall R^2 of 0.8940, with relatively low errors, as indicated by an MSE of 8.982, MAE of 1.667, and RMSE of 2.460. RFR also showed superior performance, with an R^2 of 0.879, an MSE of 9.576, an MAE of 2.186, and an RMSE of 2.951. RR remains the weakest, which struggles to capture the complex relationships within the dataset, with an R^2 of 0.599, an MSE of 29.002, an MAE of 3.641, and an RMSE of 5.382. For water absorption, RFR exhibits the best performance, consistently providing accurate predictions across various datasets, with an overall R^2 of 0.901, an MSE of 6.527, an MAE of 1.996, and an RMSE of 2.513. XGBoost follows, with less stable overall performance than the RFR, achieving an R^2 of 0.886, an MSE of 6.923, an MAE of 1.739, and an RMSE of 2.236. Meanwhile, ANN also exhibits a good overall performance, yielding an R^2 of 0.820, an MSE of 11.709, an MAE of 2.529, and an RMSE of 3.337. RR, on the other hand, produces the lowest results, although its overall performance is weaker than that of compressive strength, resulting in an overall R^2 of 0.780, an MSE of 15.219, an MAE of 3.007, and an RMSE of 3.901. The performance of RR is less weak than compressive strength. RFR demonstrated the high CV R^2 values (0.86 and 0.88) and low standard deviations (0.053 and 0.044) for compressive strength and water absorption, respectively.

Its narrow 95 % confidence intervals ([0.822–0.902] and [0.840–0.920]) reflect low prediction variability across folds, confirming its robustness and generalization capability.

Since the models have relatively variable but, to some extent, similar overall behaviours, it is essential to analyse their performance in terms of training and testing set results. To address the issue of overfitting during model development, various regularization strategies were applied. Hyperparameter tuning was conducted using grid search, modifying essential parameters such as maximum depth, learning rate, and subsample ratios to manage model complexity. To avoid excessive training and improve generalization, early stopping criteria based on validation loss were also utilized. A detailed assessment of the trade-offs between model interpretability and predictive accuracy reveals clear distinctions among the four algorithms evaluated. Fig. 8 compares the difference between training and testing sets across the models for both compressive and water absorption. Fig. 9 shows the learning curves for R^2 between the training and testing sets. By also referring to Table 7, the figures show that although XGBoost achieved superior overall performance for the compressive strength, the variances were the highest, indicating some overfitting. On the other hand, RR shows the least variance between the training and testing predictions, though its overall

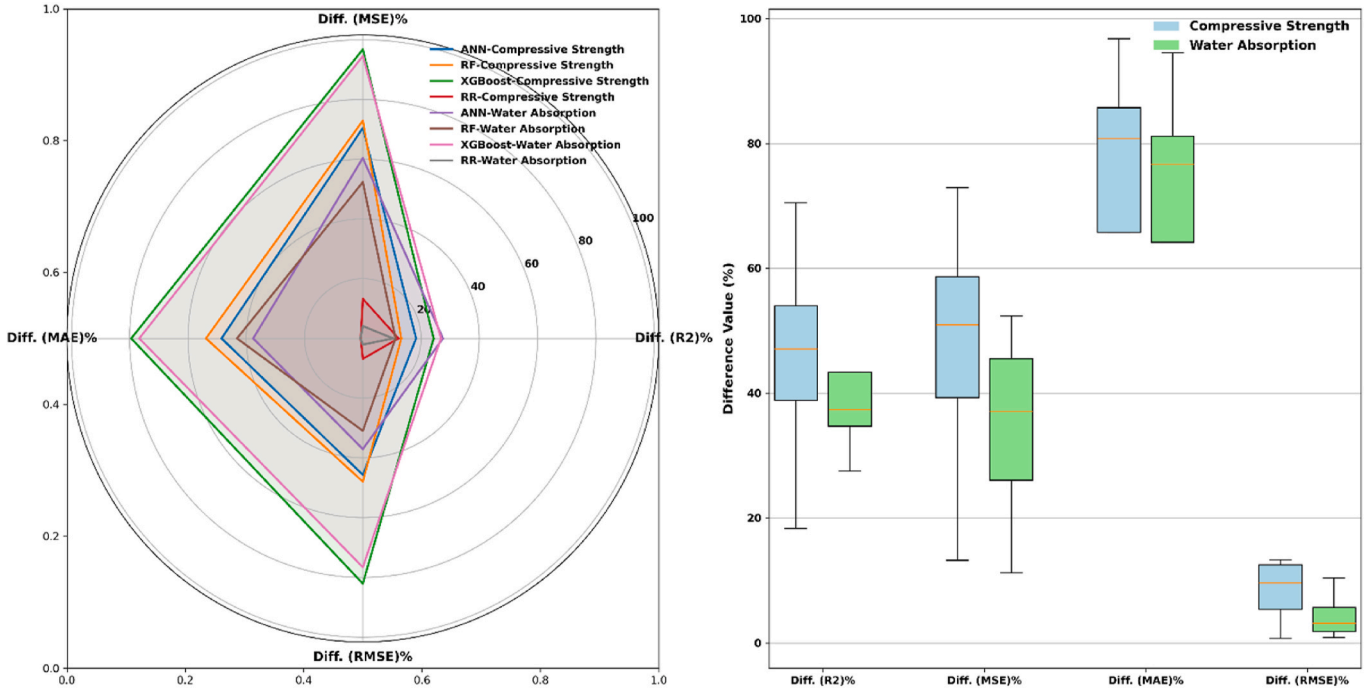

Fig. 8. Difference between training and testing of the performance model results for compressive strength and water absorption.

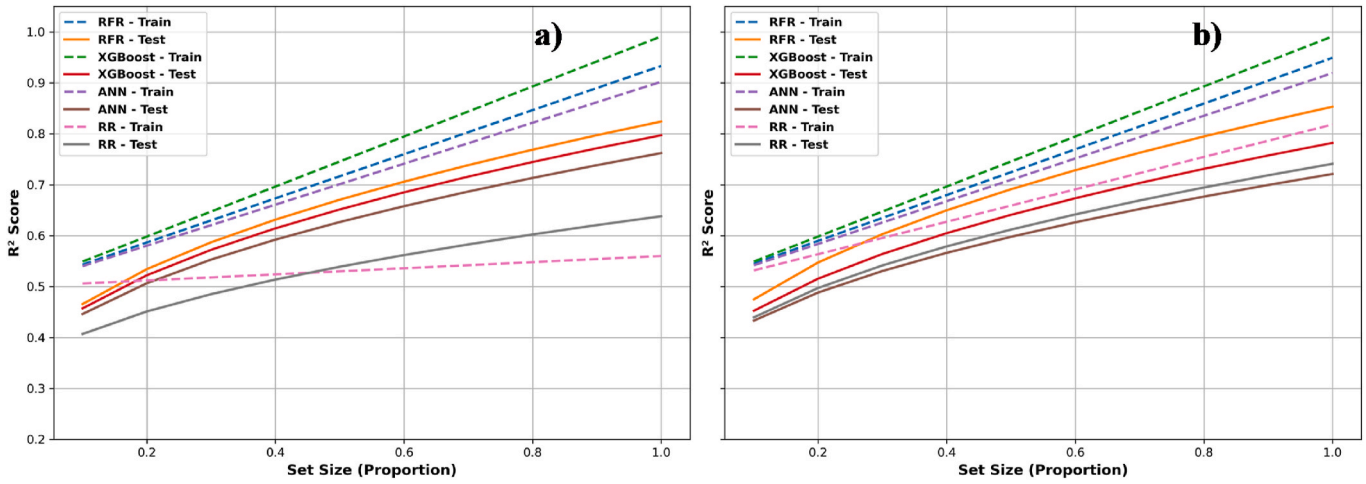

Fig. 9. Learning curves for R^2 between the training and testing sets a) compressive strength; b) water absorption.

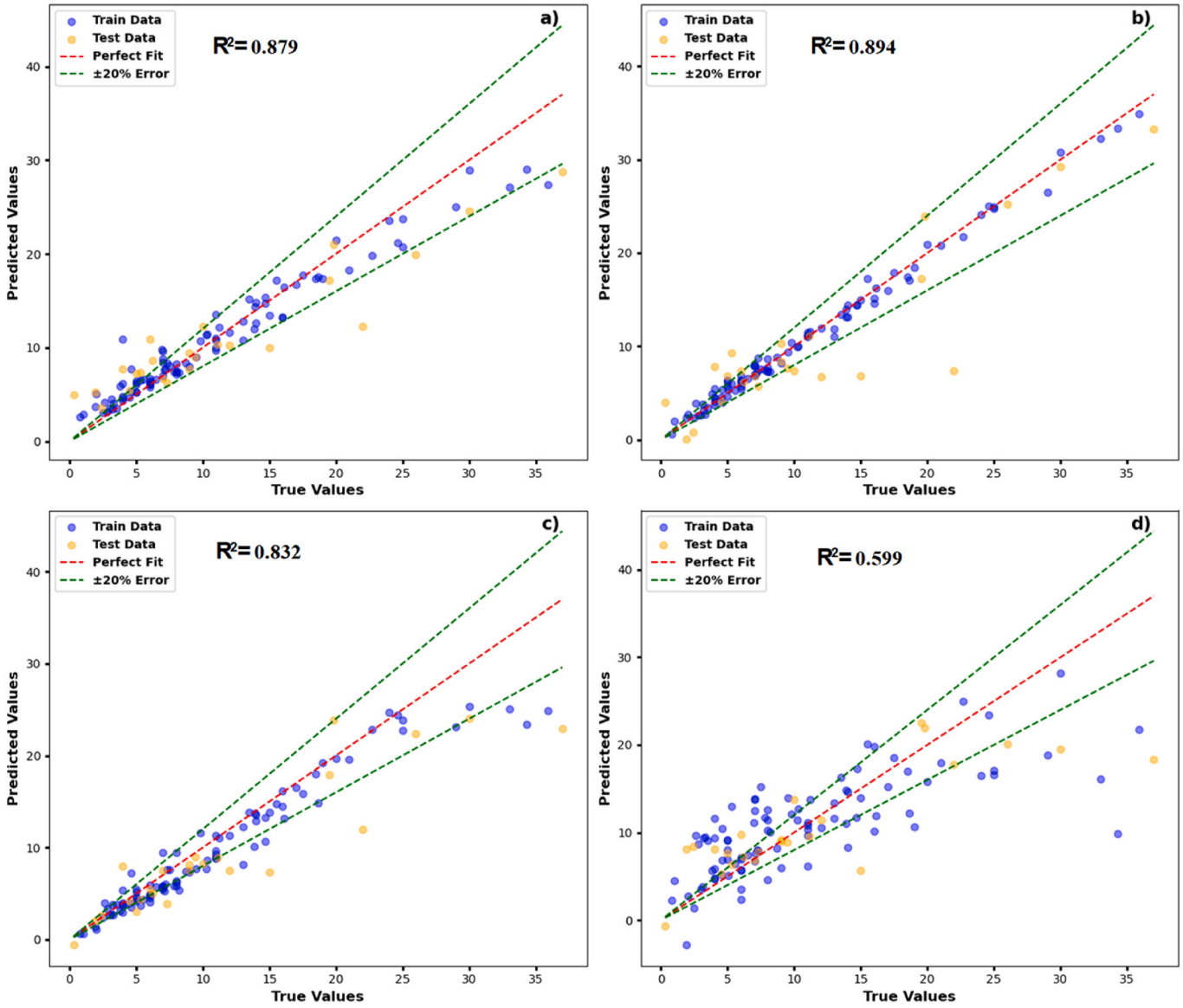

Fig. 10. Actual vs predicted compressive strength by: a) RFR, B) XGboost, c) ANN, and d) RR.

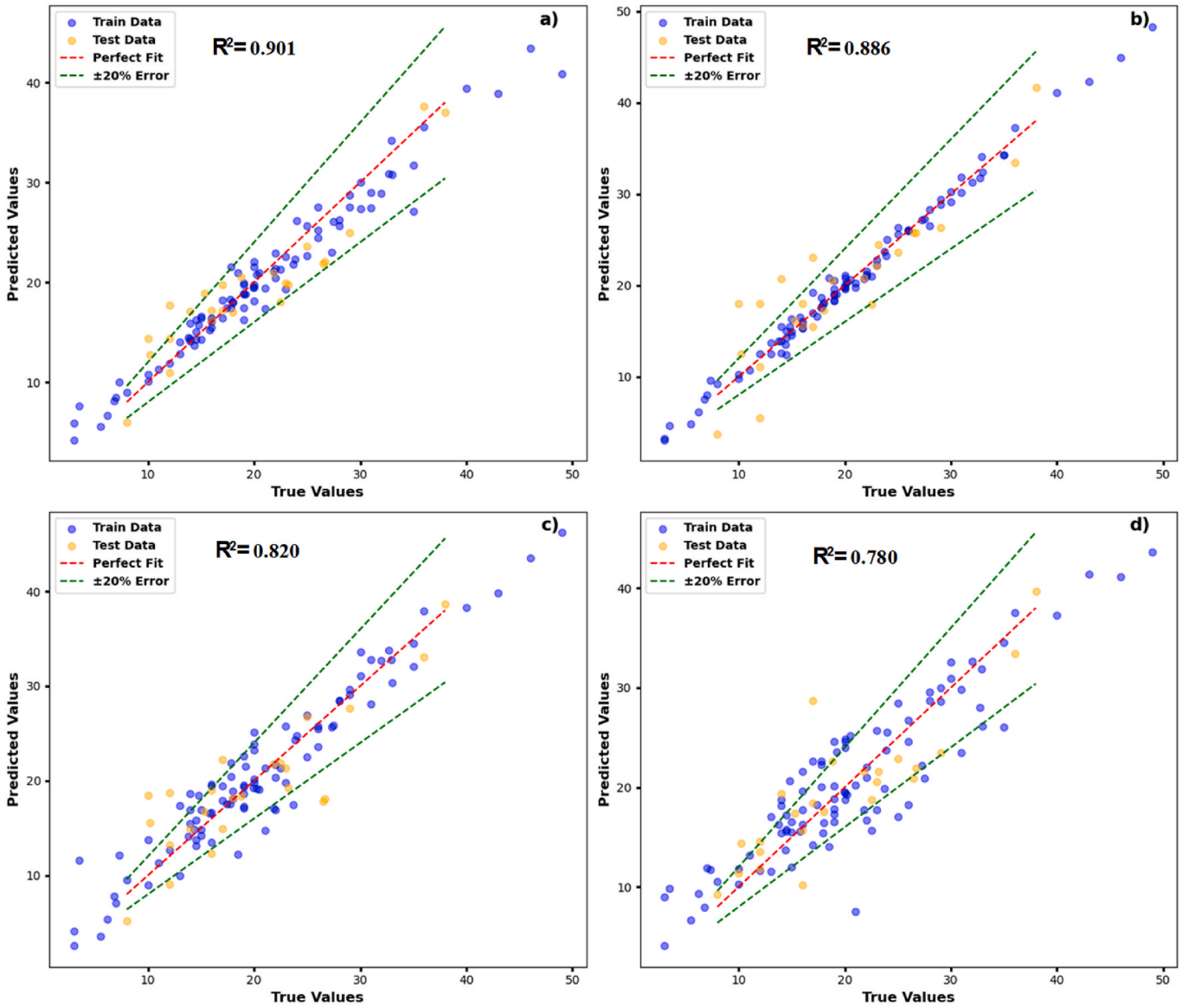

Fig. 11. Actual vs predicted water absorption by: a) RFR, B) XGboost, c) ANN, and d) RR.

performance is the lowest. RR achieves the lowest accuracy for both compressive strength and water absorption; however, it demonstrates the highest consistency in generalization, as evidenced by minimal differences between training and testing results. Its linear nature and regularization improve interpretability, which is advantageous for understanding feature impacts. However, this straightforwardness restricts its ability to capture complex, nonlinear relationships in the data. ANN has approximately similar variances to XGBoost, while RFR maintains a relatively high accuracy on training and testing data, indicating its strong generalization capability and effectiveness. In contrast, the variances for all models were relatively lower for the training and testing predictions of water absorption. However, XGBoost and ANN exhibit higher variance in terms of R2, indicating the prevalence of some overfitting. Both ANN and XGBoost show reduced interpretability and a higher likelihood of overfitting, as evidenced by notable discrepancies between training and testing performance. Overall, XGBoost, ANN, and RFR can be effectively used for compressive strength, while all four models can be used for water absorption; however, RFR provided superior performance for predicting both properties. The RFR offers a favourable compromise, delivering robust and precise predictions while maintaining moderate interpretability through a combination of decision trees.

RFR employs a bagging strategy by independently training multiple decision trees on randomly bootstrapped samples of data (also refer to Appendix A, A.1.1). This sampling procedure introduces diversity across the ensemble, as each tree learns from a distinct subset and develops a unique decision structure. When these predictions are then aggregated through simple averaging, the ensemble reduces variance, which is a key contributor to overfitting in individual trees. Given that individual trees are high-variance learners prone to overfitting their respective samples, their predictions can be unstable. However, aggregating these outputs through unweighted averaging effectively reduces overall variance, as the independent fluctuations of individual trees tend to cancel out. Since the trees are trained independently, the model is also less likely to capture spurious patterns in the data. Fundamentally, RFR improves generalisation by lowering variance through averaging, without increasing the bias of the individual trees. This approach inherently stabilises predictions by mitigating the influence of individual trees that may have overfit noise in their respective samples(Allen Akselrud, 2024; Belyadi et al., 2021). In contrast, XGBoost, for example, builds trees sequentially to iteratively reduce bias, which can increase the risk of overfitting, particularly in noisy or variable datasets. Its weighted

aggregation and focus on correcting residuals may cause it to capture spurious patterns, often leading to a larger gap between training and testing performance (Belyadi et al., 2021). Consequently, RFR exhibits strong generalisation and low overfitting, making it more reliable than ANN and XGBoost while also outperforming RR in terms of predictive power. Although the dataset used in this study was relatively small and heterogeneous, the machine learning models still demonstrated strong predictive performance, serving as a foundational step of data-driven approaches in optimizing waste-based brick properties. Nonetheless, subsequent validation using n a larger, more controlled experimental dataset is crucial to confirm model reliability, enhance accuracy, and extend applicability across broader contexts.

Figs. 10 and 11 display the scatter plots of actual versus predicted compressive strength and water absorption values for all developed models, respectively. The dotted green lines indicate the ±20 % prediction error. For compressive strength, a scatter of the data is more apparent for XGBoost and ANN compared to RFR, although most of the points lie within a ±20 % error, particularly on the training set, where the blue scatter points closely match the diagonal; the test set is somewhat more dispersed. In comparison, the data do not fit well in the RR, and most of the points lie beyond the ±20 % error. The RR model has the weakest predictive ability among these algorithms; line graphs created from actual and tested data reveal significant differences in correlation and a wide dispersion of scatter points. For water absorption, the performance is relatively similar to the compressive strength. However, the data are highly scattered for ANN and less scattered for RR.

3.2. Comparison with the state of the art review

Despite the extensive literature on utilizing machine learning to predict the properties of bricks, effective benchmarking is limited due to variations in datasets, assessment metrics, and reporting standards. Many studies report high R^2 values—often approaching 0.999 using models such as Artificial Neural Networks (ANN), as in Bouzidi, Bouzidi and Quesada (Bouzidi et al., 2024) and Pamu and Svsndl (Pamu et al., 2024), but omit essential error measures such as RMSE and MAE. This limits transparency and makes it challenging to assess model robustness and generalization. The ANN model developed in this research achieved an R^2 of 0.832, an RMSE of 3.487, and an MAE of 2.278 for compressive strength—metrics that, although lower in R^2 compared to some previous studies, were attained using a broader and more comprehensive dataset that included various waste materials. Additionally, the absence of standardized datasets across various studies complicates direct comparison, highlighting the necessity for open-access benchmarks and consistent evaluation protocols. Conversely, the current study employs a thorough and transparent evaluation framework that incorporates multiple metrics, including R^2, MSE, MAE, and RMSE, along with a detailed variance analysis between the training and testing phases to assess generalization. This method facilitates more reliable performance comparisons and model selection based on both accuracy and stability. The models created in this study, particularly the Random Forest Regressor (RFR). For predicting compressive strength, the RFR model attained an R^2 of 0.879 with an RMSE of 2.951, while XGBoost yielded an R^2 of 0.894 and a lower RMSE of 2.460. These outcomes are on par with or exceed those of previous models tested. For example, Praburanganathan, Chithra and Simha reddy (Praburanganathan et al., 2022) report an R^2 of 0.900 and RMSE of 0.660 for Random Forest applied to ash-based bricks. Although the error values in this study are higher, they reflect the enhanced complexity and diversity of the dataset, yielding more realistic and broadly applicable findings. Compared to models in Khokhar, Khan (Khokhar et al., 2023), where support vector machines (SVM), Gaussian process regression (GPR), and other techniques yielded R^2 values ranging from 0.730 to 0.815 with RMSEs exceeding 4.0, the models in this study consistently deliver superior performance. Deep neural networks (DNN) used in Lamba, Kaur (Lamba et al., 2024) also achieved an R^2 of 0.854, but were still outperformed by all models here except Ridge Regression (RR), which, while less accurate, provides consistent and interpretable results. Notably, the application of Ridge Regression demonstrates minimal variation between the training and testing phases, despite exhibiting a lower predictive performance ($R^2 = 0.599$). These characteristic underscores its potential utility in scenarios where consistent generalization, regulatory transparency, or model interpretability are of paramount importance.

The forecasting of water absorption strengthens the relative efficacy of the suggested models. Among them, the RFR model exhibited the best performance ($R^2 = 0.901$, RMSE = 2.513), surpassing earlier results such as those in Vasić, Jantunen (Vasić et al., 2023), where ANN and RFR achieved R^2 values of 0.965 and 0.664, respectively, but with significantly higher RMSEs of 1.195 and 4.106. The XGBoost and ANN models in this study also delivered strong performance, with R^2 values of 0.886 and 0.820, respectively, exceeding results from Khokhar, Khan (Khokhar et al., 2023) and Lamba, Kaur (Lamba et al., 2024), where many models failed to reach the 0.80 R^2 threshold. The results of this study indicate that, although some prior publications have reported seemingly greater accuracy, such performance is often a consequence of utilizing simpler or more constrained datasets. Conversely, the current research prioritizes model generalization and diverse input variability, which are crucial for real-world applicability, compensating for the slightly elevated error margins observed. A significant number of earlier studies have primarily concentrated on predictive performance, often favouring complex black-box models without adequately considering the trade-offs between explainability, complexity, and accuracy. The comprehensive comparative methodology employed in this study facilitates a critical evaluation of these trade-offs, thereby enhancing the methodological understanding within the field. For industry adoption, implementation would require digitized production records, standardization of input variables (e.g., waste composition, processing parameters), and training for quality control personnel in data-driven decision-making.

3.3. SHAP analysis

The SHAP analysis was conducted for the RFR model since it provided the optimum predictions. The bar chart in SHAP analysis represents a practical and straightforward approach to understanding the overall significance of the various features in a model. It shows the mean absolute SHAP values for each feature, which indicates the average contribution of that feature to the model's predictions across all instances in the dataset. Higher-impact features on the model's outputs have longer bars according to their average absolute value of the SHAP scores. Although the bar chart is vital in understanding the overall importance of features, it lacks detailed insights into the interaction between feature values and model predictions (Huang et al., 2025). Thus, a SHAP summary plot is generated, which serves as a vital visualization tool. It provides a more detailed view of the relationship between feature values and their contributions to model predictions. It arranges input features in order of their significance to the model's prediction, with each data sample represented by a dot. The colour of each dot (red or blue) indicates the value of the feature (high or low), while the horizontal position reflects the impact of that feature on the model's prediction for that sample. Features with high positive impact appear on the positive side of the plot, while those with negative influence are on the left side. By analysing the summary plot, it is clear which features contribute positively or negatively to the prediction. The model's accuracy improves when features on the positive side are incorporated, showcasing how each feature's contribution affects the model's performance (Lundberg et al., 2020). The SHAP bar charts and summary plots of the RFR models in this study are shown in Figs. 12 and 13, respectively, for the prediction of compressive strength and water absorption.

For compressive strength predictions (Fig. 12a), the content of the soil is the most important factor, followed by the size of a brick, moisture

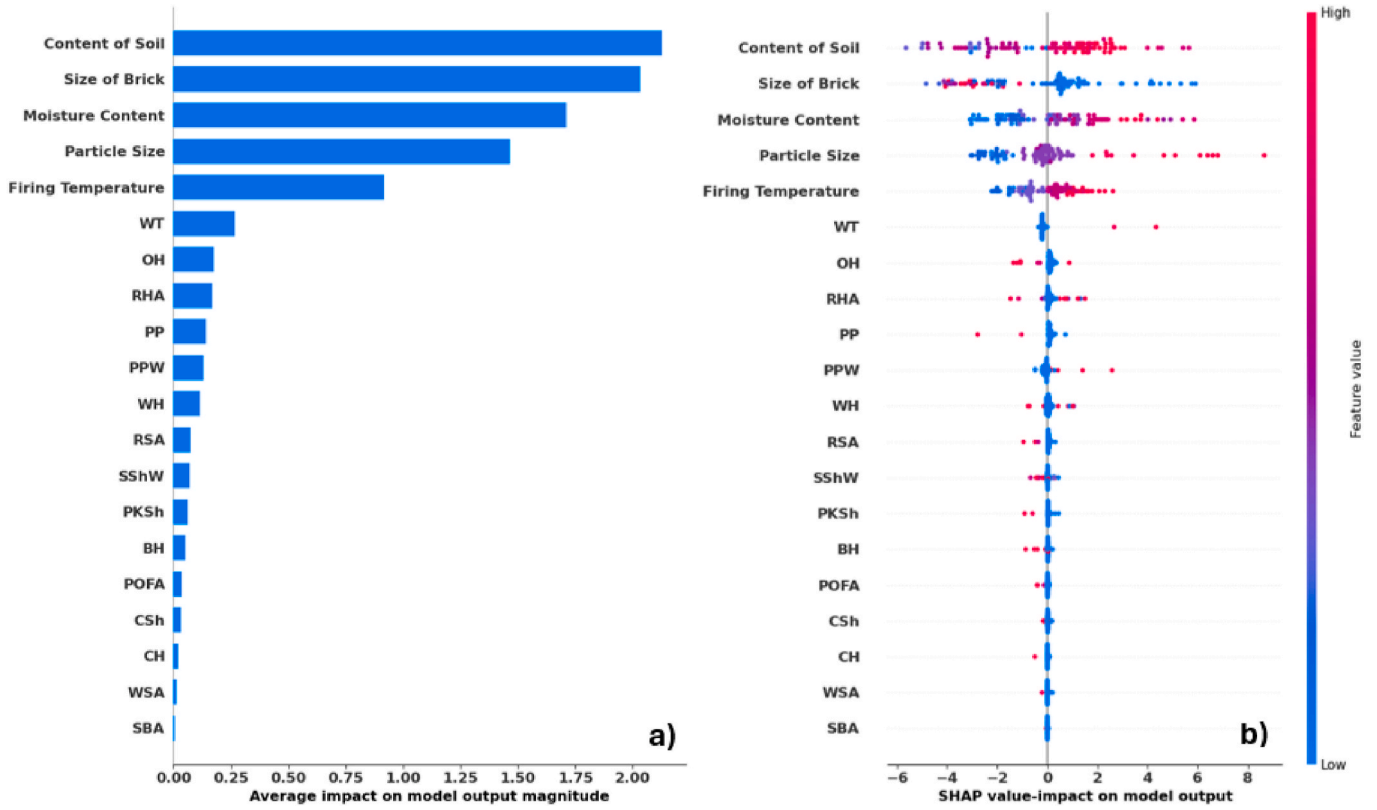

Fig. 12. SHAP analysis of RFR predictions for compressive strength: a) SHAP value, and b) SHAP summary plot.

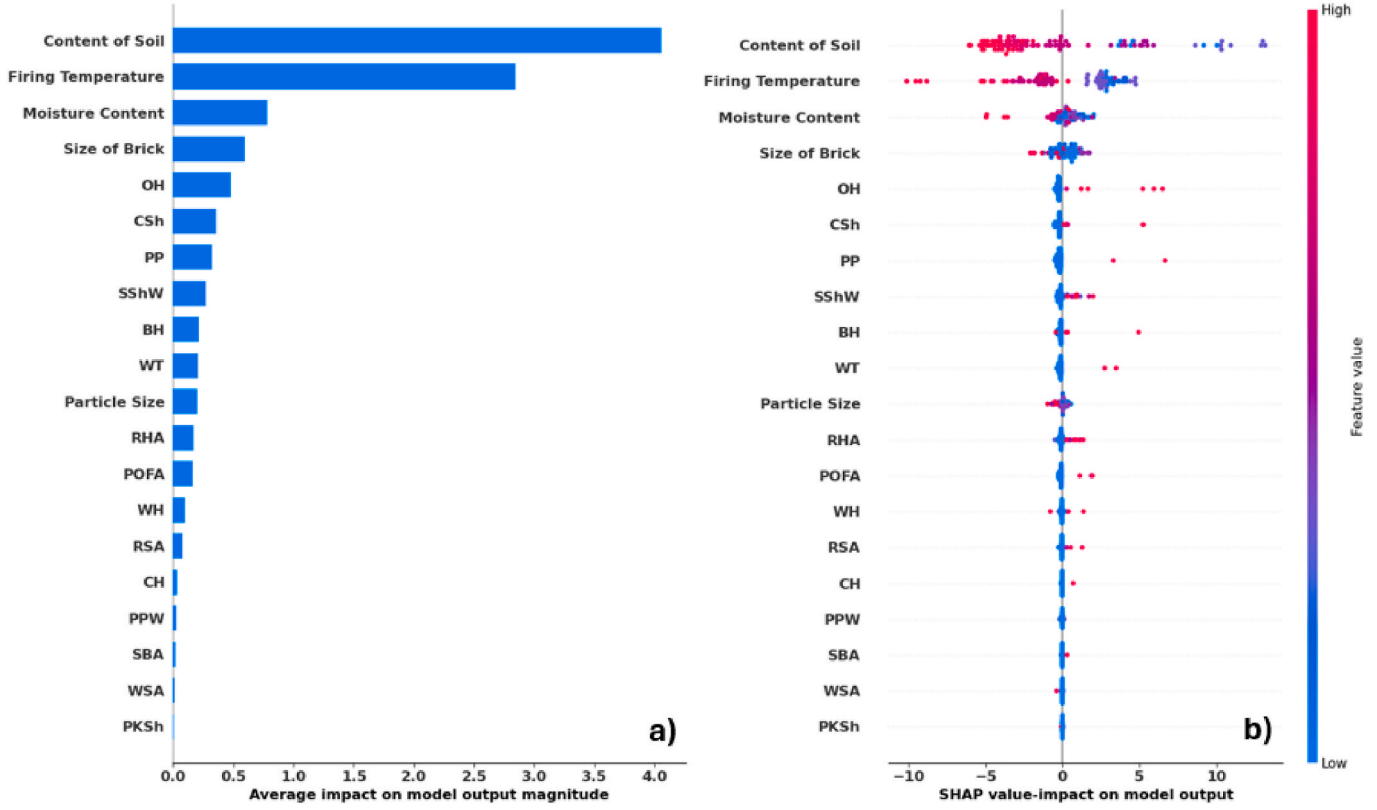

Fig. 13. SHAP analysis of RFR predictions for water absorption: a) SHAP value, and b) SHAP summary plot.

content, particle size, and firing temperature, with magnitude values of about 2.25, 2, 1.75, 1.5, and 1, respectively. Soil content and brick size are dominant features due to their role in structural integrity and packing density. Their interaction with moisture content and firing temperature enhances bonding and densification, mainly when coarser particles are used. For example, high soil content combined with moderate moisture levels improves compressive strength through better compaction. Interestingly, WT waste shows the highest impact among other types, followed by OH, RHA, and PP. The summary plot (Fig. 12b) also signifies the same features. High values of features, such as the content of soil and the size of bricks (red dots), generally drive predictions upward, while low values (blue) decrease them. SHAP values range from −6 to +8 in extreme cases. Design strategies should prioritize soil content within practical limits and adjust brick size to enhance mechanical stability. For example, incorporating RHA at levels up to approximately 15 %, particularly under firing temperatures around 950 °C, may significantly improve compressive strength without compromising structural performance. Additional enhancements can be achieved through the control of moisture content and the use of coarser raw materials, both of which support stronger inter-particle bonding and densification during firing. While firing temperature was of comparatively lower importance, maintaining it within an optimal range remains essential to ensure sufficient vitrification.

The use of RFR for water absorption predictions (Fig. 13a) reveals higher impacts from the content of soil and firing temperature, at values of approximately 4 and 3, respectively. However, the effects of the brick size and moisture content are less pronounced, at 1 and 0.5, respectively. Additionally, the impacts of several types of waste, such as OH, CSh, and PP, are more pronounced than those of other types of waste that affect compressive strength predictions and the particle size of the waste. Firing temperature and soil content are crucial in reducing porosity and sealing the microstructure, thereby lowering water uptake. Their interaction with waste type (e.g., OH and CSh) reveals that some wastes are more sensitive to thermal conditions. Bricks with higher soil content fired at ≥950 °C showed the lowest water absorption, highlighting the synergy between composition and process parameters. The summary plot (Fig. 13b) confirms the results, showing lower scattering and higher values that span from −10 to +12 in extreme cases. The content of soil and firing temperature exhibits the highest spread. In summary, the moisture content and size of brick contribute moderately. While the content of soil and firing temperature have the most significant impact on predicting water absorption. These findings imply that bricks intended for use in water-sensitive or outdoor environments should prioritize formulations with higher proportions of soil and firing regimes that reach or exceed 950 °C to ensure microstructural closure. It is suggested that to design bricks for water-exposed environments, OH and CSh content should be minimized relative to other wastes, and firing temperatures should be maximized within the practical range of the process (e.g., ≥950 °C), as these settings are directly associated with reduced SHAP contributions to water absorption.

3.4. Sensitivity analysis

The sensitivity analysis identified the features that most significantly

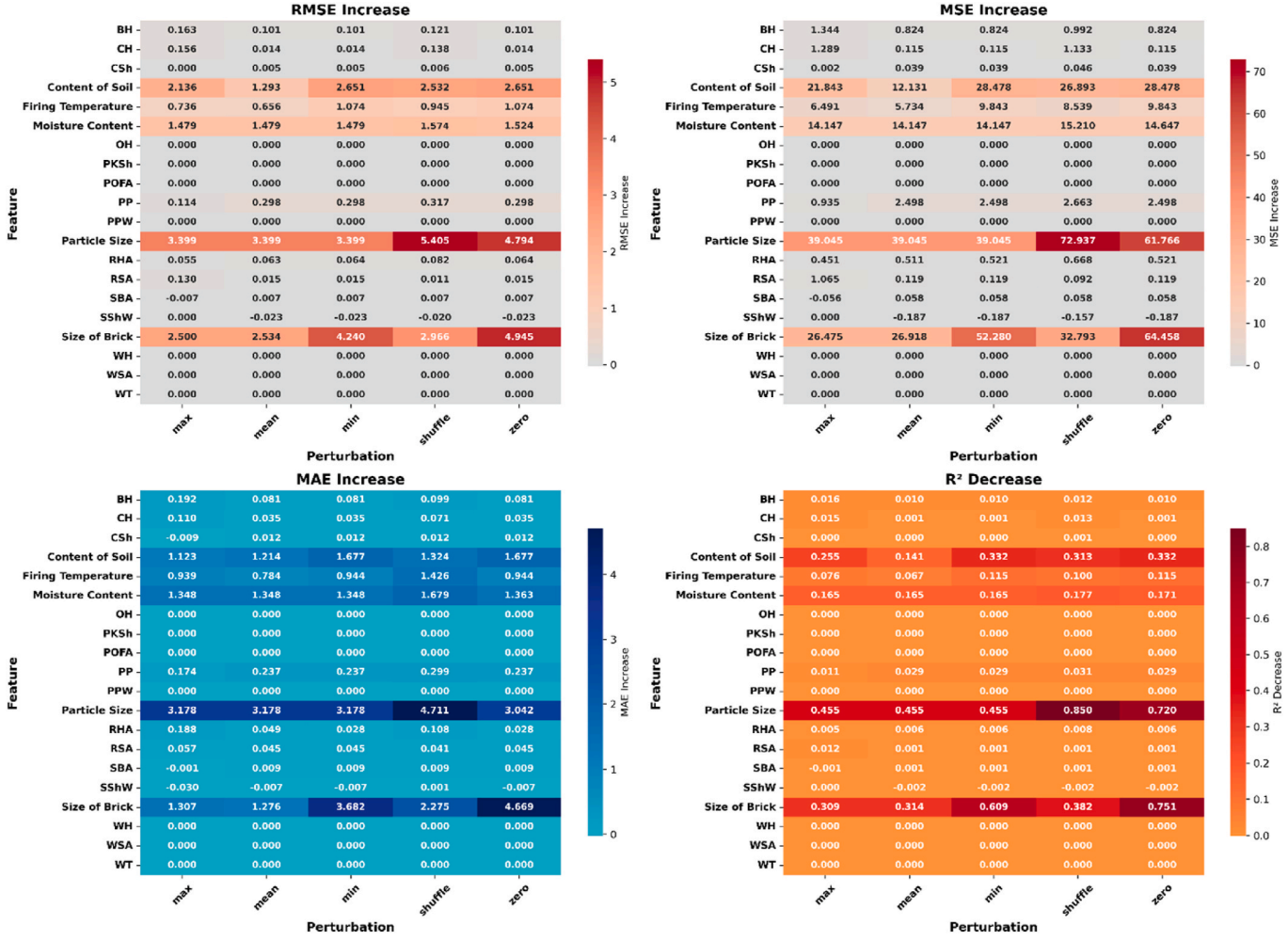

Fig. 14. Sensitivity analysis plots for compressive strength predictions.

influence the model's predictive performance when altered, emphasizing the primary waste sources that impact compressive strength predictions. It was conducted on the optimized models, i.e., RFR. Features that displayed larger increases in Root Mean Square Error (RMSE) and Mean Squared Error (MSE), or greater declines in R^2, were deemed critical inputs, highlighting the essential need for accurate measurement and control of these variables to achieve reliable model outcomes (Razavi et al., 2021; Demay et al., 2025). The combined heatmaps in Figs. 14 and 15 facilitated the effective visual identification of perturbation types and features that significantly impaired model performance.

For compressive strength (Fig. 14), the analysis confirmed that variations in input features substantially affect the prediction errors of machine learning models concerning fired brick properties. Among all variables, the size of the brick exhibited the highest sensitivity, with an RMSE increase of 4.945, MSE increase of 64.458, MAE increase of 4.669, and a decrease in R^2 of 0.751. This suggests that even minor changes in brick size significantly impact model accuracy. Particle size followed closely, showing an RMSE increase of 4.794, MSE increase of 61.766, MAE increase of 3.178, and a reduction in R^2 of 0.720, further confirming its critical influence on predictive performance. Notably, soil content, while ranked highest in overall feature importance, showed moderate sensitivity in terms of error variation, with increases in RMSE, MSE, and MAE of 2.799, 30.480, and 1.850, respectively, alongside an R^2 decrease of 0.355. Moisture content and firing temperature also demonstrated moderate sensitivity, with moisture content causing an RMSE increase of 1.524, MSE of 14.647, MAE of 1.736, and R^2 drop of 0.171, while firing temperature led to increases in RMSE (1.074), MSE (9.843), MAE (0.944), and a decrease in R^2 of 0.115. In contrast, the model showed low sensitivity to agricultural waste additives such as PP, BH, CH, and RHA. For instance, PP led to minimal changes: RMSE increase of 0.298, MSE of 2.498, MAE of 0.237, and R^2 decrease of 0.029. Similar negligible impacts were observed for other additives. These results suggest that while feature importance rankings highlight soil content, brick size, moisture content, particle size, and firing temperature as key contributors to model predictions, the model's error sensitivity is most impacted by brick size and particle size. Therefore, strict control over these physical and process parameters is crucial. Simultaneously, the low sensitivity to agricultural waste additives reinforces their viability for sustainable integration without significantly compromising model reliability.

For water absorption (Fig. 15), the content of soil demonstrates the highest sensitivity, with a root mean square error (RMSE) increase of 7.900, a mean squared error (MSE) increase of 109.029, a mean absolute error (MAE) increase of 6.635, and a notable decrease in R^2 of 1.815. Firing temperature ranks as the second most sensitive parameter, exhibiting increases in RMSE, MSE, and MAE of 5.396, 60.957, and 4.253, respectively, along with a reduction in R^2 of 1.015. This highlights its significant impact on microstructural transitions and vitrification processes, which influence material strength and porosity. CSh and BH demonstrate notable sensitivity among waste additives, with RMSE increases of 2.975 and 2.841, respectively, indicating substantial

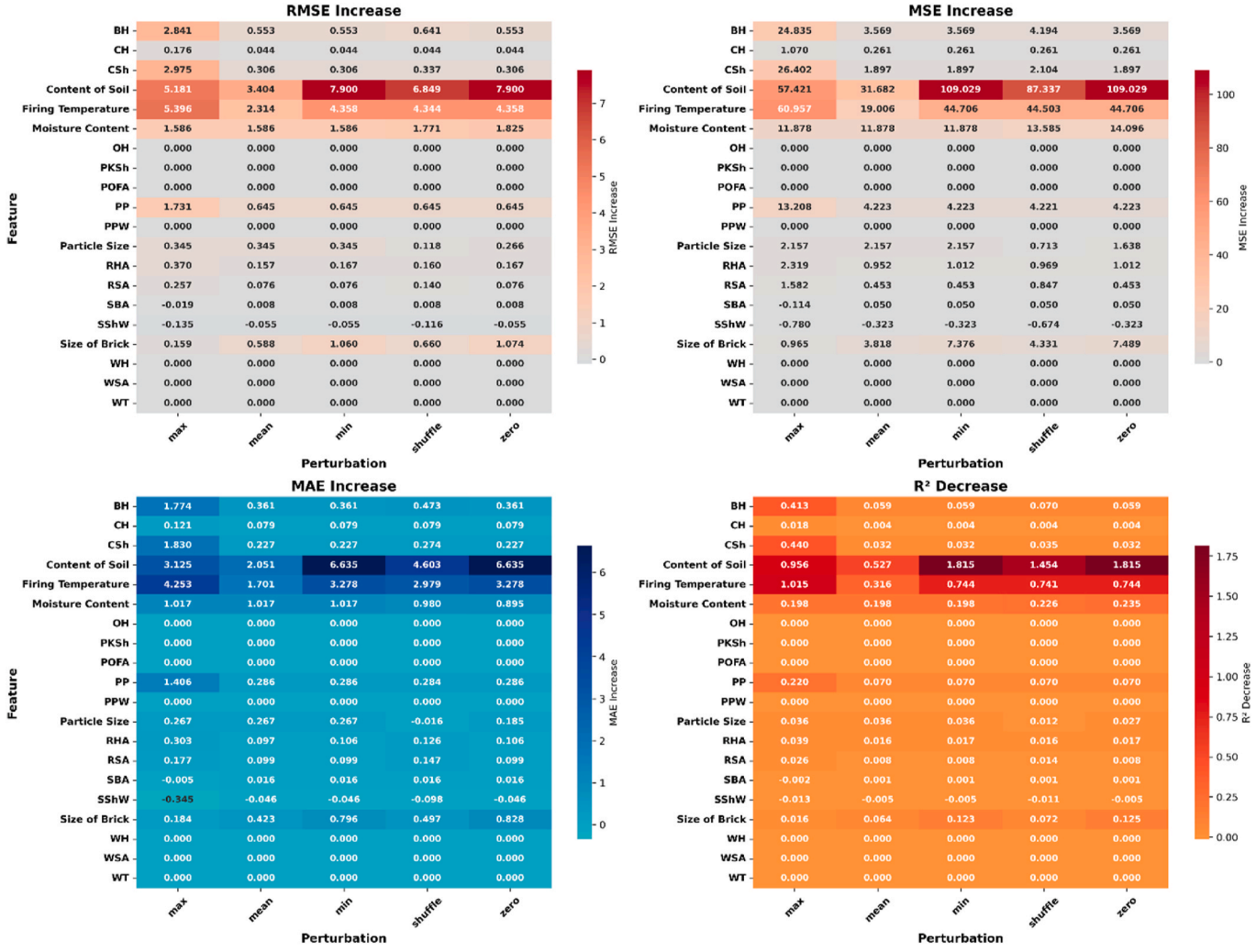

Fig. 15. Sensitivity analysis plots for water absorption predictions.

effects on model predictions. Moisture content and PP also contribute significantly to model sensitivity, with RMSE increases of 1.972 and 1.731, respectively, highlighting their importance in modifying mechanical properties and water resistance. Conversely, physical characteristics such as brick size and particle size show lower sensitivity, with RMSE increases of 1.074 and 0.345, respectively. This suggests that compositional and processing parameters are more influential than dimensional factors in this analysis. Overall, these results highlight the importance of optimizing soil content and firing temperature to enhance model reliability and brick quality, while underscoring the significant impact of specific waste additives on predictive accuracy and physical performance.

3.5. PDP and ICE analysis

The analysis was conducted for the RFR model since it provides the optimum predictions. Figs. 16 and 17 display the plots for compressive strength and water absorption, respectively. In these figures, the PDPs are shown by the dark blue lines while ICEs are indicated by the light blue colours for all input variables. PDPs illustrate the global average impact of an input on the output, making it easier to understand the effect of a particular input on the output. Meanwhile, ICEs offer a more granular, individual-level perspective, showing the impact of varying inputs on the output and capturing the heterogeneity of the model's behaviour across different instances. Both assist in providing a more

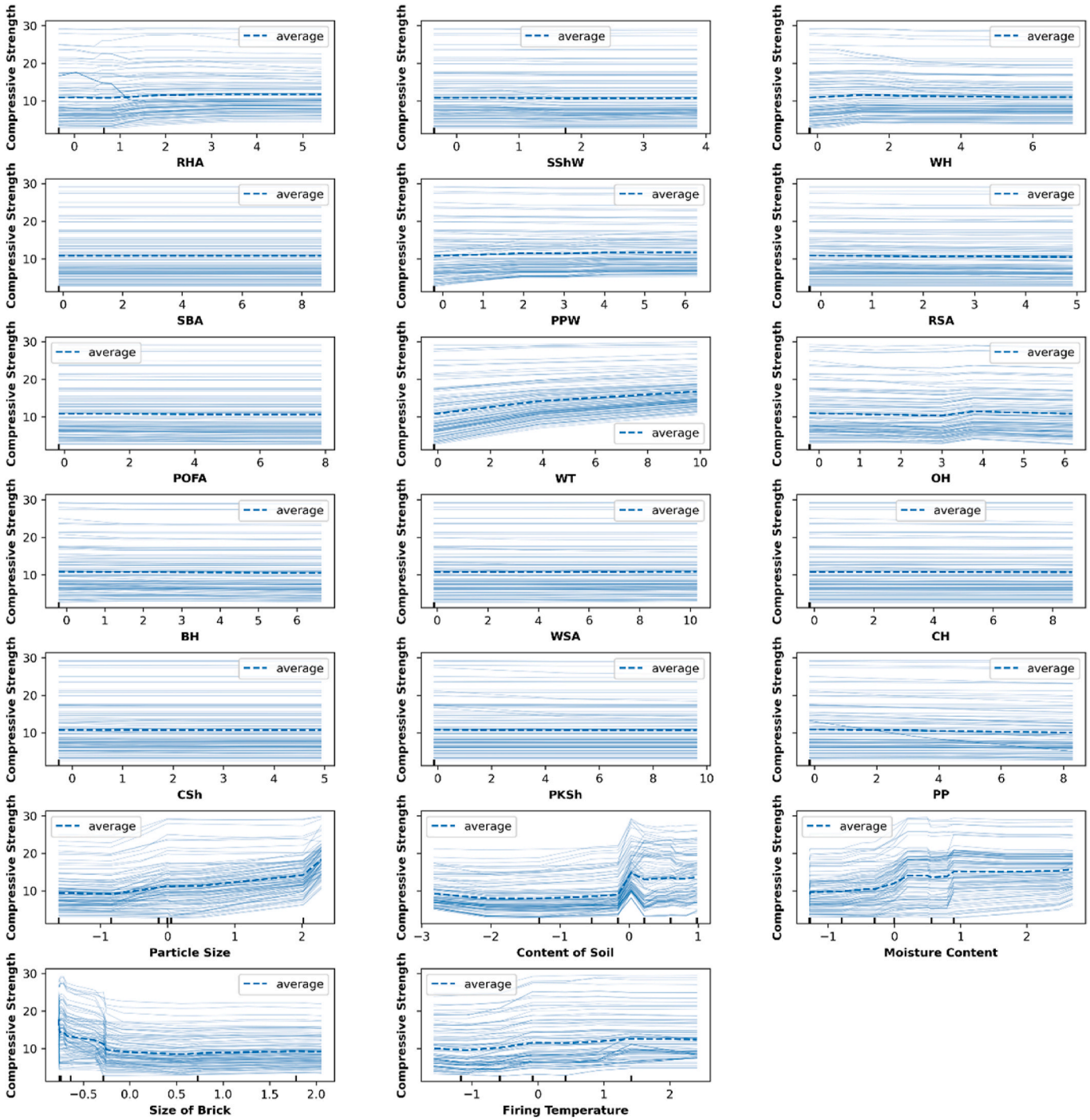

Fig. 16. PDP and ICE plots for the RFR model for compressive strength predictions.

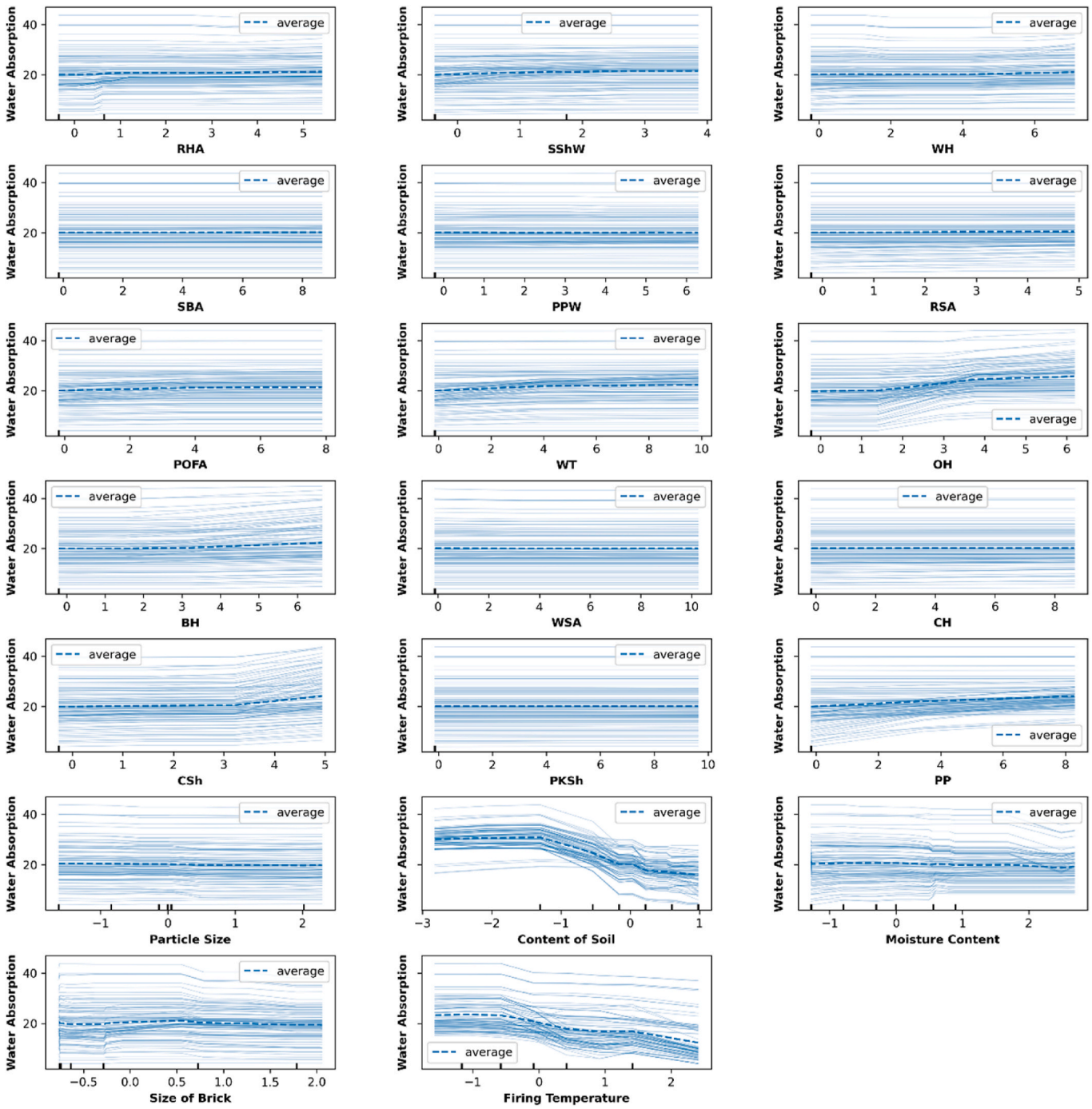

Fig. 17. PDP and ICE plots for the RFR model for water absorption predictions.

comprehensive picture of the impact of an input on the required output. By comparing the trends across different variables, it is easier to determine which features have a strong, weak, or negligible influence on the target property (Sifan et al., 2023).

Fig. 16 illustrates a slight positive impact of varying the content of RHA and WT on the compressive strength of fired bricks, among other types of waste. The variation in the content of different kinds of wastes, such as SBA, POFA, WSA, CH, and PKSh, shows minimal or flat influence. Meanwhile, the variations in the content of PPW and OH show mild upward trends. On the other hand, more significant impacts are evident for physical and process-related variables. The particle sizes and moisture content of the waste show a noticeable positive influence, as indicated by the increases in compressive strength. However, the impacts of both variables are downward in some observations, suggesting their possible adverse effect on the strength of bricks. The firing temperature exhibits a consistent upward trend, indicating an improvement in strength with increasing temperatures. In addition, the impact of the soil content is the strongest, though somewhat variable and fluctuating, which can reflect the content of waste since most waste materials are used as a replacement for soil. The size of bricks exhibits an overall declining trend, indicating a possible negative impact on their compressive strength.

For water absorption (Fig. 17), the change in the contents of most waste materials, such as RHA, SShW, POFA, BH, and WT, shows slight increasing trends, indicating that they may have a subtle impact on porosity and thereby increase water absorption. The change in the

content of other wastes, on the other hand, such as CSh, OH, and PP, shows noticeable upward trends, suggesting that their use may noticeably increase water absorption. Meanwhile, wastes such as RSA, SBA, PPW, CH, WSA, and PKSh provide flat lines with limited impacts on the water absorption of fired bricks. More substantial impacts are observed for physical and manufacturing variables. The content of soil and firing temperature show strong negative trends, suggesting that their increases reduce water absorption, possibly because of better sintering and bonding. Conversely, the particle size of the waste and the sizes of the bricks show relatively upward trends, indicating that the increases in both variables tend to absorb more water, likely due to weaker compaction or a larger pore volume. Moisture content exhibits a mild downward trend, which may benefit the reduction in water absorption in some cases. The findings suggest that while the type of waste has some impact, optimizing physical parameters plays a critical role in enhancing the compressive strength and water absorption of fired bricks.

4. Conclusions and recommendations

The study involved the application of four machine learning algorithms: Random Forest Regressor (RFR), Extreme Gradient Boosting (XGBoost), Artificial Neural Network (ANN), and Ridge Regression (RR), to predict the compressive strength and water absorption of fired bricks made with various agricultural waste materials. Unlike prior works, this research couples robust predictive modelling with interpretability techniques such as SHAP and PDP/ICE plots, offering deep insights into the influence of both physical and process parameters on compressive strength and water absorption. Based on the comprehensive analysis of the study, the following conclusions are drawn:

1. All four models show strong potential for predicting the compressive strength and water absorption of fired bricks, with varying degrees of predictive performance obtained through hyperparameter optimization.
2. XGBoost and RFR achieved highest overall predictive performance for compressive strength ($R^2 = 0.894$ and 0.879, respectively), while RFR showed the best performance for water absorption ($R^2 = 0.901$). In contrast, RR consistently yielded the lowest predictive performance across both targets due to its limited ability to capture nonlinear relationships in the dataset.
3. The variances between the performance results of training and testing sets for the XGBoost model were the highest, which indicates the occurrence of some overfitting. ANN exhibits similar behaviour to XGBoost, while RR has minimal variance but poor prediction accuracy. RFR provided a balanced performance for both predictions, demonstrating strong potential as an effective predictive tool.
4. RFR was identified as the most promising model for predicting both compressive strength and water absorption. Overall, it achieves an R^2 of 0.879, an MSE of 9.576, an MAE of 2.186, and an RMSE of 2.951 for compressive strength and an R^2 of 0.901, with an MSE of 6.527, an MAE of 1.996, and an RMSE of 2.513 for water absorption.
5. The model interpretation tools, such as SHAP and PDP/ICE plots, demonstrated that the target characteristics are significantly influenced by physical and process-related variables, including soil content, firing temperature, moisture content, particle size, and brick size. Waste products and their contents, on the other hand, also show some effects.
6. The sensitivity analysis demonstrated that soil content and firing temperature are the most influential features for water absorption prediction, with RMSE increases of 7.90 and 5.40, respectively. At the same time, brick size and particle size had the most significant impact on compressive strength, with RMSE increases of 4.95 and 4.79. In contrast, agricultural waste additives such as banana peels, bagasse husk, and rice husk ash showed relatively low sensitivity, with RMSE changes typically below 2.0, supporting their sustainable inclusion without significantly compromising model reliability.
7. Among all input features, soil content (reflecting a waste content), moisture content, brick size, and firing temperature are the most actionable for manufacturers due to their strong and consistent influence on both strength and water absorption. These parameters can be directly controlled during material selection and processing. For burned bricks to perform better, it is therefore crucial to optimize critical physical characteristics, such as soil content and firing temperature.

Therefore, it can be concluded that machine learning models, specifically RFR, show strong potential for application in the predictions of the properties of waste-based bricks. Extensive testing is required by conventional experimental methods, which can be resource-intensive and time-consuming. Conversely, machine learning models offer a robust alternative for interpreting the impact of individual mix components on key properties of fired brick, such as compressive strength and water absorption. By enabling simulation of a wide range of mix designs, these models allow researchers and practitioners to focus experimental validation on the most promising formulations, optimizing the use of agricultural by-products without compromising quality.

Despite promising results, the study is subject to certain limitations. Overfitting was observed in some algorithms, notably XGBoost and ANN, suggesting the need for further data augmentation or regularization techniques. Additionally, the applicability of the models to other brick types, such as unfired or concrete bricks, remains to be validated. Building on these findings, it is recommended to further explore the subject by using other types of waste, focusing on the chemical composition of waste materials, and examining different or specific types of bricks, such as unfired bricks and concrete bricks. Additionally, the inclusion of variables like curing and extended firing conditions should be considered. Investigating environmental impacts through lifecycle assessments combined with predictive modelling also represents a valuable avenue for holistic sustainability evaluation. Additionally, future work should investigate other key brick performance indicators, such as thermal conductivity, density, porosity, and durability under varying environmental conditions, to develop more holistic predictive frameworks. Despite the limitations of using secondary data, this study establishes a framework for developing high-performing, interpretable models to guide future experimental investigations. Future studies with expanded and well-controlled datasets will be crucial to confirm model robustness, enhance predictive accuracy, and extend applicability to sustainable construction materials, thereby bridging the gap between predictive modelling and practical industrial implementation, ultimately supporting the development of high-performance, eco-friendly bricks. The concept of machine learning in the brick development process can lead to more efficient, sustainable, and cost-effective practices, thereby improving performance while optimizing waste utilization. By leveraging these models, manufacturers can significantly reduce the cost and time associated with experimental testing, optimize mix designs, and confidently integrate agricultural waste into their production processes.

Declaration of competing interest

There is no conflict of interest, whAT DO you mean by SIT, i dont find any link at end of the manuscript PDF, please upload the screenshot in EW, so that I can send to the eidtor for reference.

Acknowledgment

This study was funded by the Higher Committee for the Development of Education (Iraq) and the University of Liverpool (United Kingdom) as a part of a PhD research project. The financial, scientific, and other resources provided by these institutions were instrumental to the successful completion of this research.

Nomenclature

Abbreviation	Description
ANN	Artificial Neural Network
XGBoost	Extreme Gradient Boosting
LR	Linear Regression
Lasso	Lasso Regression
Ridge	Ridge Regression
DT	Decision Tree
RF	Random Forest
RFR	Random Forest Regression
MLR	Multiple Linear Regression
DNN	Deep Neural Network
GPR	Gaussian Process Regression
SVM	Support Vector Machines
CART	Classification and Regression Tree
BTR	Boosted Tree Regression
KNN	K-Nearest Neighbours
SHAP	SHapley Additive exPlanations
PDP	Partial Dependence Plot
ICE	Individual Conditional Expectation
ML	Machine Learning
MSE	Mean Squared Error
RMSE	Root Mean Squared Error
MAE	Mean Absolute Error
R^2	Coefficient of Determination

Symbol	Meaning
X	Input feature vector, i.e., x=(x1,x2, …,xn)x = (x_1, x_2, …, x_n)
x_i	The i-th individual feature in the feature vector
h(x)	Prediction function of a single decision tree
$g_j(x)$	Output of the j-th decision tree in an ensemble
J	Total number of decision trees in the ensemble
L	General loss function measuring the error between prediction and target
y/Y	Actual or target output value
$\hat{y}_i$	Predicted value of the i-th data point
$\bar{y}$	Mean of actual values
n/nn	Number of features or data points
ψ[y, f(x)]	Loss function in boosting (error between yy and f(x)f(x))
f(x)	Model's predicted output for input xx
f′(x)	Updated model prediction after adding a new tree
argmin	Value that minimizes a given function
β (beta)	Regression coefficient in ridge/lasso regression
α (alpha)	Regularization parameter controlling penalty strength
∑ (sigma)	Summation symbol
Min/Max	Minimum/Maximum value of dataset variable
Mean	Average value of dataset variable
Std	Standard deviation of dataset variable
CI	Confidence Interval

Symbol/Term	Description
BH	Barley Husk
CH	Coconut Husk
CSh	Coconut Shell
OH	Oat Husk
PKSh	Palm Kernel Shell
POFA	Palm Oil Fuel Ash
PP	Potato Peels
PPW	Pomegranate Peel Waste
RHA	Rice Husk Ash
RSA	Rice Straw Ash
SBA	Sugarcane Bagasse Ash
SShW	Spent Shea Waste
WH	Wheat Husk
WSA	Waste Straw Ash
WT	Waste Tea
Particle size	Average particle size
Soil content	Proportion of soil in mix
Moisture content	Moisture presents in the mix
Brick size	Volume of brick
Firing temperature	Temperature during firing
Water absorption	Percentage of water absorbed by the brick
Compressive strength	Mechanical strength of brick under load

Appendix A

A.1 Fundamentals of the models' algorithms

A.1.1 Random Forest Regressor (RFR)

A random forest Regressor (RFR) is a regression algorithm that represents robust regression algorithm in ML and has been broadly used (Han et al., 2019). The principle of the model is shown in Figure A.1a. The main difference between this model and Decision Tree models (DT) lies in the number of trees. While DT constructs a single tree, the RFR builds several trees, collectively referred to as a forest. Each of the trees is built from a new training set, and dissimilar data is allocated and selected arbitrarily for the trees using the bagging method (Chehreh Chelgani et al., 2016). The bagging method, also known as bootstrapping aggregation, represents a technique with two steps: bootstrapping and aggregation. Initially, the original set of data is randomly resampled to produce distinct datasets that are identically distributed. According to Amin, Iqbal (Amin et al., 2022), bootstrapping is essential for ensuring that every decision tree in the RF model is distinct, which lowers the variance and noise of the anticipated outcomes. Each decision tree in the forest is trained on different random subsets of the feature vector x = (x1, x2, …, xn), reflecting the model's ability to learn various aspects of the data. The function h(x) represents the prediction function of a single tree, and to refine this prediction, the model minimizes the loss function L, which measures the difference between the predicted value (from h(x)) and the actual output Y. Secondly, the base predictors are trained separately using the new datasets. The aggregation method is used for obtaining the final prediction by averaging the predictions from all the individual trees. This is mathematically expressed as:

$$h(x) = 1/J \sum J\, j = 1\ gj(x) \ldots \ldots \quad (A.1)$$

Where each gj(x) is the output of the j-th decision tree.

Because it eliminates potential biases and unpredictability from individual trees, the aggregation stage is essential for producing a more reliable and consistent forecast. (Vasić et al., 2023). Multiple trees may have overfitted or underfitted the data in different ways, but by integrating their results, the ensemble model becomes less susceptible to noise and outliers, increasing its overall accuracy and generalization.

A.1.2 Extreme Gradient Boosting (XGBoost)

Gradient Boosting (GB) is a tree-based boosting machine learning algorithm that builds an ensemble of decision trees to create a "strong" learner from a series of "weak" learners (individual decision trees) (Sifan et al., 2023). The principle of the model is shown in Figure A.1b. The fundamental principle of GB is to iteratively construct trees that fix the mistakes of earlier trees, producing a more accurate model. At every iteration, a new weak tree is added to the ensemble, and each tree is trained to concentrate on the model's residual errors. This iterative procedure continues until the model's overall error is either sufficiently reduced or hits a predetermined acceptable limit. Reducing the total error, commonly known as the loss function, is a primary goal of the GB training process. The approach modifies the model by incorporating a new tree to gradually decrease the loss function, which assesses how well its forecasts match the actual target values (Golafshani et al., 2024b). The loss function, denoted as:

$$\psi[y, f(x)] \ldots \ldots \quad (A.2)$$

Where y is the target variable and f(x) is the model's prediction for input x.

The estimation function f'(x) is the model's prediction after the addition of a new tree, which is determined by the formula:

$$f'(x) = \text{argmin}\ f(x)\ \psi[y, f(x)] \ldots \ldots \quad (A.3)$$

This equation expresses the goal of finding the best function f(x) that minimizes the difference between the predicted and actual values.

A-1-3 Artificial Neural Networks (ANNs)

Artificial Neural Networks (ANNs) exhibit remarkable competence in areas such as data analysis, pattern recognition, and solving complex problem-solving tasks (Nakkeeran et al., 2024). The principle of the model is shown in Figure A.1c. ANN represents an ML algorithm that simulates the human brain in terms of its functionality and neural network structure (Nazemi et al., 2019). It consists of interconnected nodes, known as neurons, arranged in a structure comprising one input layer, one or more hidden layers, and one output layer. The information is received and passed from one neuron to the next, and it is affected by a weight that is set according to the importance of the inputs to the outputs (DeRousseau et al., 2018). A combined function is used to emerge the information from various neurons. After that, the merged data is sent to the subsequent nodes. Until the maximum number of iterations has been reached or the algorithm accurately matches the data, as shown by the convergence of the error rate, this iterative procedure is repeated (Bourdeau et al., 2019). For training and evaluating the model, the input layer transmits input parameters. Linking the input layer to the output layer, which provides the model's outcome, is the responsibility of the hidden layer or layers. To generate the neurons' output and guarantee data transfer between the output and hidden layers, activation functions are necessary (Mohandes et al., 2019). In an ANN, each connection between neurons has a weight that determines the strength of the connection, and each neuron has a bias that helps adjust the output. To learn, an ANN iteratively adjusts the weights of its connections in response to input data and the intended output.

A-1-4 Ridge Regression (RR)

A common regression method for analysing multicollinear regression data that gets beyond the drawbacks of ordinary least squares (OLS) is ridge regression (Liu et al., 2023). It is a regularization technique that reduces overfitting by appending a penalty term to the cost function. When the OLS model picks up on the noise in the data, it becomes overfit. The model consequently performs poorly on the test set but exceptionally well on the training set. To prevent overfitting, ridge regression employs L2 regularization (Liang et al., 2022). Adding a penalty term to the OLS's cost function lowers the coefficients in the direction of zero. The equation for this term is given by:

$$\alpha * \sum(i = 1 \text{ to } n)\ \beta^{2,} \ldots \ldots \quad (A.3)$$

Where α represents the regulation parameter, while β and n refer to the coefficient and their numbers, respectively (Alyami et al., 2024).

The value of α determines the amount of regularization. A lower α results in less regularization, while a higher α results in more regularization. The principle of the algorithm is shown in Figure A.1d. Ridge regression addresses the bias-variance trade-off by shrinking coefficients towards zero, reducing variance while introducing some bias. The key challenge is selecting the optimal α to balance this trade-off. It helps handle multicollinearity between predictors, prevents overfitting, and enhances model accuracy, though it may introduce bias by not allowing coefficients to fully reach zero (Sharma et al., 2023).

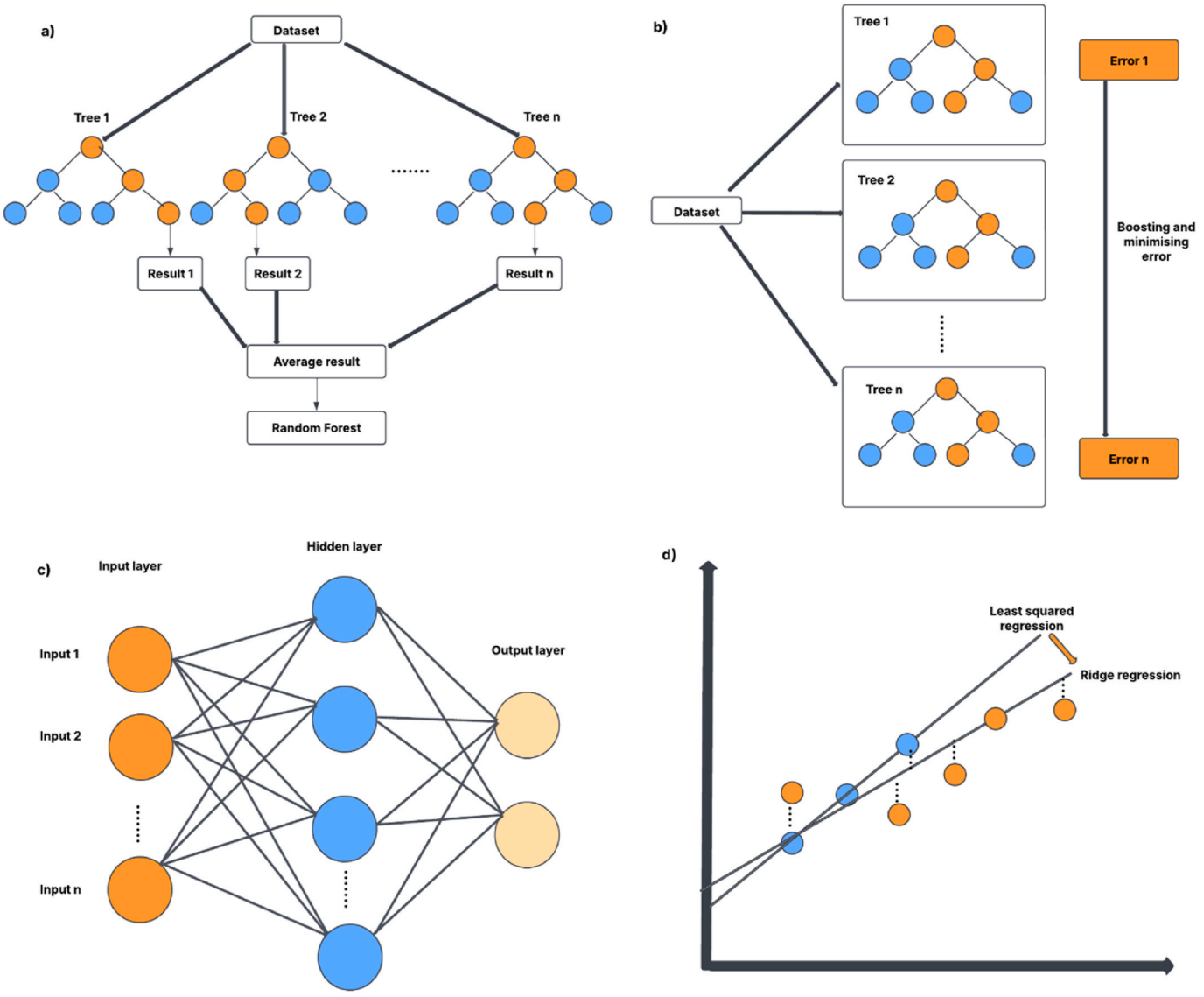

Fig. A.1. principles of the model used, a) RFR; b) XGBoost; c) ANN, and d) RR.

A.2 Models implementation methods figures

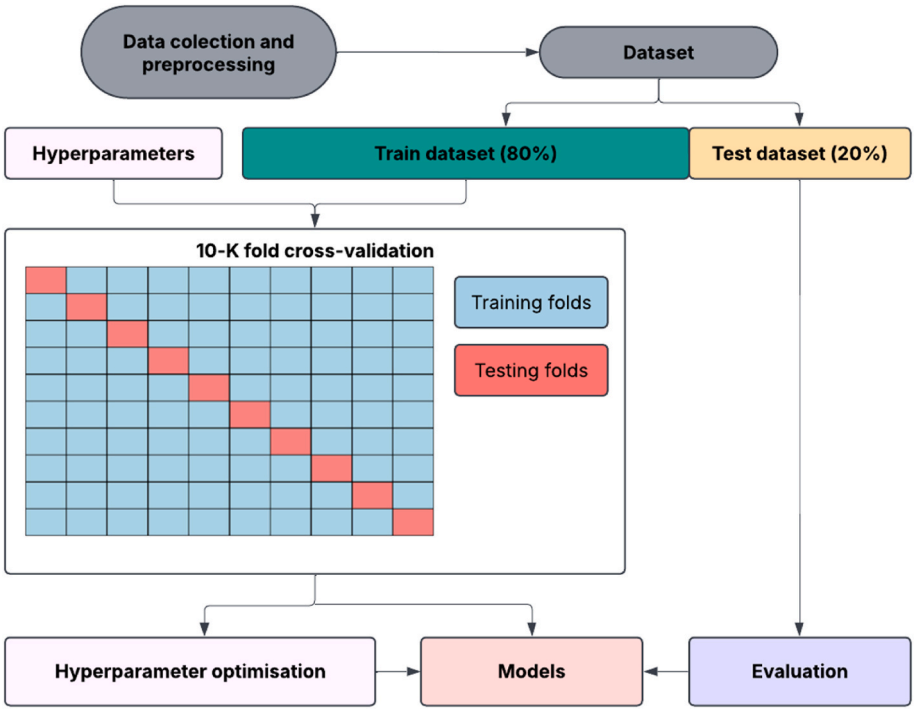

Fig. A.2. Method of the research.

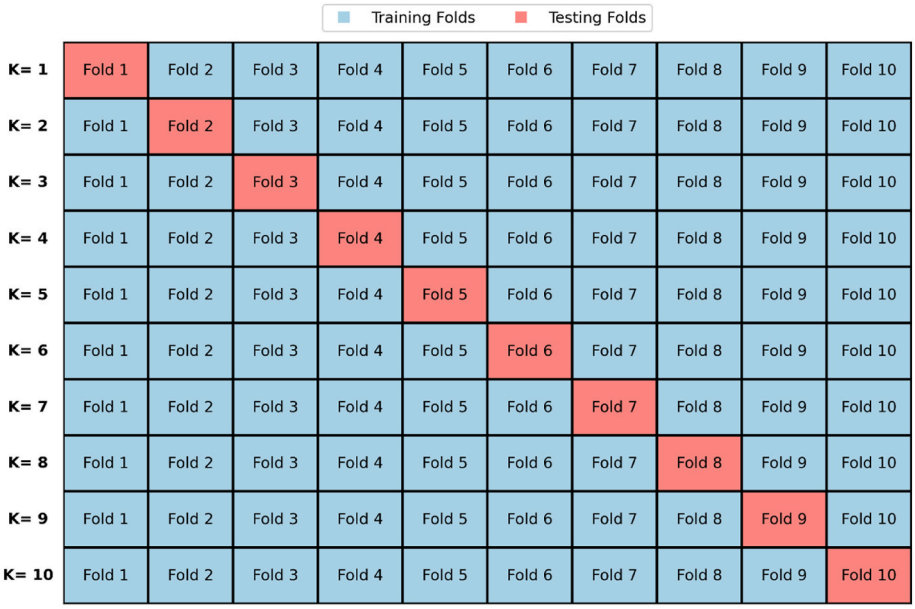

Fig. A.3. The principle of k-fold cross-validation.

Data availability

Data will be made available on request.

References

Abdel Hamid, E., 2021. Investigation of using granite sludge waste and silica fume in clay bricks at different firing temperatures. HBRC Journal 17 (1), 123–136.
Adazabra, A.N., Viruthagiri, G., Kannan, P., 2017. Influence of spent shea waste addition on the technological properties of fired clay bricks. J. Build. Eng. 11, 166–177.
Agbede, O., Joel, M., 2011. Effect of rice husk ash (RHA) on the properties of Ibaji burnt clay bricks. Am. J. Sci. Ind. Res. 2, 674–677.
Ahmad, S., Iqbal, Y., Muhammad, R., 2017. Effects of coal and wheat husk additives on the physical, thermal and mechanical properties of clay bricks. Bol. Soc. Espanola Ceram. Vidr. 56 (3), 131–138.
Ahmed, S., et al., 2023. Improving the thermal performance and energy efficiency of buildings by incorporating biomass waste into clay bricks. Materials 16 (7), 2893.
Al-jabery, K., et al., 2020. Data Analysis and Machine Learning Tools in MATLAB and Python, pp. 231–290.
Al-Rbaihat, R., Al-Marafi, M.N.I., 2023. Combined effect of silicon dioxide and titanium dioxide nanoparticles on concrete properties. Journal of Ecological Engineering 24 (12), 319–335.
Allen Akselrud, C.I., 2024. Random forest regression models in ecology: accounting for messy biological data and producing predictions with uncertainty. Fish. Res. 280, 107161.

Alrbai, M., et al., 2024. Integration and optimization of a waste heat driven organic rankine cycle for power generation in wastewater treatment plants. Energy 308, 132829.
Alyami, M., et al., 2024. Estimating compressive strength of concrete containing rice husk ash using interpretable machine learning-based models. Case Stud. Constr. Mater. 20, e02901.
Amin, M.N., et al., 2022. Prediction of strength and CBR characteristics of chemically stabilized coal gangue: ANN and random forest tree approach. Materials 15 (12), 4330.
Anand, P., et al., 2025. Optimizing agricultural waste by-products: a machine learning approach for sustainable construction practices. Circ. Econ. Sustain.
Anton, C., et al., 2021. Machine learning methods applied for modeling the process of obtaining bricks using silicon-based materials. Materials 14 (23), 7232.
Asteris, P.G., et al., 2021. Predicting concrete compressive strength using hybrid ensembling of surrogate machine learning models. Cement Concr. Res. 145, 106449.
Belyadi, H., Haghighat, A., 2021. Chapter 5 - supervised learning. In: Belyadi, H., Haghighat, A. (Eds.), Machine Learning Guide for Oil and Gas Using Python. Gulf Professional Publishing, pp. 169–295.
Bourdeau, M., et al., 2019. Modeling and forecasting building energy consumption: a review of data-driven techniques. Sustain. Cities Soc. 48, 101533.
Bouzidi, M.A., Bouzidi, N., Quesada, D.E., 2024. Prediction of mechanical and physical properties of spent bleaching earth based fired bricks: an experimental study using RSM and ANN. Asian Journal of Civil Engineering 25 (8), 5811–5833.
Chaabene, W.B., Flah, M., Nehdi, M.L., 2020. Machine learning prediction of mechanical properties of concrete: critical review. Constr. Build. Mater. 260, 119889.
Chehreh Chelgani, S., Matin, S.S., Makaremi, S., 2016. Modeling of free swelling index based on variable importance measurements of parent coal properties by random forest method. Measurement 94, 416–422.
Chuluunsaikhan, T., et al., 2021. Comparative analysis of predictive models for fine particulate matter in Daejeon, South Korea. Atmosphere 12, 1295.
Cultrone, G., et al., 2020. Sawdust recycling in the production of lightweight bricks: how the amount of additive and the firing temperature influence the physical properties of the bricks. Constr. Build. Mater. 235, 117436.
De Silva, G.H.M.J.S., Perera, B.V.A., 2018. Effect of waste rice husk ash (RHA) on structural, thermal and acoustic properties of fired clay bricks. J. Build. Eng. 18, 252–259.
Demay, M.B., et al., 2025. On the meaning of sensitivity analysis. Measurement: Sensors 38, 101511.
Demir, I., 2006. An investigation on the production of construction brick with processed waste tea. Build. Environ. 41 (9), 1274–1278.
Demir, A., Kumanlioglu, A.A., 2017. The prediction of brick wall strengths with artificial neural networks model. In: AIP Conference Proceedings. AIP Publishing.
DeRousseau, M.A., Kasprzyk, J.R., Srubar, W.V., 2018. Computational design optimization of concrete mixtures: a review. Cement Concr. Res. 109, 42–53.
Djamaluddin, A.R., et al., 2021. Fired clay bricks incorporating palm oil fuel ash as a sustainable building material: an industrial-scale experiment. Int. J. Sustain. Eng. 14 (4), 852–864.
Eliche-Quesada, D., et al., 2017. Characterization and evaluation of rice husk ash and wood ash in sustainable clay matrix bricks. Ceram. Int. 43 (1, Part A), 463–475.
Er, Y., et al., 2022. Recycling of metallurgical wastes in ceramics: a sustainable approach. Constr. Build. Mater. 349, 128713.
Erdogmus, E., et al., 2023. Enhancing thermal efficiency and durability of sintered clay bricks through incorporation of polymeric waste materials. J. Clean. Prod. 420, 138456.
Ganasen, N., et al., 2023. Soft computing techniques for predicting the properties of raw rice husk concrete bricks using regression-based machine learning approaches. Sci. Rep. 13 (1), 14503.
Gencel, O., et al., 2021a. Feasibility of using clay-free bricks manufactured from water treatment sludge, glass, and marble wastes: an exploratory study. Constr. Build. Mater. 298, 123843.
Gencel, O., et al., 2021b. Recycling industrial slags in production of fired clay bricks for sustainable manufacturing. Ceram. Int. 47 (21), 30425–30438.
Ghorbani, M., et al., 2021. A comparative study on physicochemical properties of environmentally-friendly lightweight bricks having potato peel powder and sour orange leaf. Constr. Build. Mater. 276, 121937.
Golafshani, E.M., et al., 2024a. A framework for low-carbon mix design of recycled aggregate concrete with supplementary cementitious materials using machine learning and optimization algorithms. Structures 61, 106143.
Golafshani, E., et al., 2024b. Modelling the compressive strength of geopolymer recycled aggregate concrete using ensemble machine learning. Adv. Eng. Software 191, 103611.
Goldstein, A., et al., 2015. Peeking inside the black box: visualizing statistical learning with plots of individual conditional expectation. J. Comput. Graph Stat. 24 (1), 44–65.
Gupta, V., et al., 2020. A state of the art review to enhance the industrial scale waste utilization in sustainable unfired bricks. Constr. Build. Mater. 254, 119220.
Han, Q., et al., 2019. A generalized method to predict the compressive strength of high-performance concrete by improved random forest algorithm. Constr. Build. Mater. 226, 734–742.
Hasan, M.R., et al., 2021. Effects of waste glass addition on the physical and mechanical properties of brick. Innovative infrastructure solutions 6, 1–13.
Hassan, M., 2021. Recycling of Cullet, Waste Clay Bricks and Wastes Resulted from Wheat and Sugarcane Cultivations in the Manufacture of Fired Clay Bricks, vol. 17. WSEAS Transactions on Environment and Development, pp. 192–200.
Heniegal, A.M., et al., 2020. Study on properties of clay brick incorporating sludge of water treatment plant and agriculture waste. Case Stud. Constr. Mater. 13, e00397.
Huang, X., Huang, J., Kaewunruen, S., 2025. An explainable machine learning system for efficient use of waste glasses in durable concrete to maximise carbon credits towards net zero emissions. Waste Manag. 193, 539–550.
Ivanović, M.M., 2015. Modern machine learning techniques and their applications. Electronics, Communications and Networks IV.
Kakkar, Y., Singh, R., Patel, M., 2023. Utilization of agricultural waste for the sustainable production of clay bricks. In: Proceedings of Indian Geotechnical and Geoenvironmental Engineering Conference (IGGEC) 2021, vol. 2. Springer Nature Singapore, Singapore.
Kalabarige, L.R., et al., 2024. Machine learning modeling integrating experimental analysis for predicting compressive strength of concrete containing different industrial byproducts. Adv. Civ. Eng. 2024 (1), 7844854.
Kazmi, S.M.S., et al., 2018. Thermal performance enhancement of eco-friendly bricks incorporating agro-wastes. Energy Build. 158, 1117–1129.
Khambra, G., Shukla, P., 2023. Novel machine learning applications on fly ash based concrete: an overview. Mater. Today Proc. 80, 3411–3417.
Khan, M.A., et al., 2021. Geopolymer concrete compressive strength via artificial neural network, adaptive neuro fuzzy interface system, and gene expression programming with K-Fold cross validation. Front. Mater. 8.
Khokhar, S.A., et al., 2023. A predictive mimicker for mechanical properties of eco-efficient and sustainable bricks incorporating waste glass using machine learning. Case Stud. Constr. Mater. 19, e02424.
Kifli, Z., Absa, M., Musyafa, A., 2017. Prediction of mechanical properties of light weight brick composition using artificial neural network on autoclaved aerated concrete. Asian Journal of Applied Sciences 5 (3).
Kizinievič, O., et al., 2018. Eco-friendly fired clay brick manufactured with agricultural solid waste. Arch. Civ. Mech. Eng. 18 (4), 1156–1165.
Lakhouit, A., et al., 2025. Integrating machine learning for precision agriculture waste estimation and sustainability enhancement. Comput. Electron. Agric. 230, 109933.
Lamba, P., et al., 2024. Repurposing plastic waste: experimental study and predictive analysis using machine learning in bricks. J. Mol. Struct. 1317, 139158.
Liang, M., et al., 2022. Interpretable ensemble-machine-learning models for predicting creep behavior of concrete. Cement Concr. Compos. 125, 104295.
Limami, H., et al., 2023. Machine learning forecasting of thermal, mechanical and physicochemical properties of unfired clay bricks with plastic waste additives. Mater. Today Proc. 72, 3509–3513.
Ling, H., et al., 2019. Combination of support vector machine and K-Fold cross validation to predict compressive strength of concrete in marine environment. Constr. Build. Mater. 206, 355–363.
Liu, K., et al., 2023. Development of compressive strength prediction platform for concrete materials based on machine learning techniques. J. Build. Eng. 80, 107977.
Lundberg, S.M., et al., 2020. From local explanations to global understanding with explainable AI for trees. Nat. Mach. Intell. 2 (1), 56–67.
Lyu, Z., et al., 2022. Back-propagation neural network optimized by K-Fold cross-validation for prediction of torsional strength of reinforced concrete beam. Materials 15 (4), 1477.
Marín-García, D., et al., 2023. Deep learning model for automated detection of efflorescence and its possible treatment in images of brick facades. Autom. ConStruct. 145, 104658.
Mohandes, S.R., Zhang, X., Mahdiyar, A., 2019. A comprehensive review on the application of artificial neural networks in building energy analysis. Neurocomputing 340, 55–75.
Morsali, S., Qian, D., Minary-Jolandan, M., 2020. Designing bioinspired brick-and-mortar composites using machine learning and statistical learning. Communications Materials 1 (1), 12.
Moujoud, Z., et al., 2023. Study of fired clay bricks with coconut shell waste as a renewable pore-forming agent: technological, mechanical, and thermal properties. J. Build. Eng. 68, 106107.
Muñoz, P., et al., 2023. Assessing technological properties and environmental impact of fired bricks made by partially adding bottom ash from an industrial approach. Constr. Build. Mater. 396, 132338.
Murmu, A.L., Patel, A., 2018. Towards sustainable bricks production: an overview. Constr. Build. Mater. 165, 112–125.
Nakkeeran, G., et al., 2024. Mechanical properties optimization and cost analysis of agricultural waste as an alternative in brick production. Sci. Rep. 14 (1), 24075.
Nazemi, E., et al., 2019. Estimation of volumetric water content during imbibition in porous building material using real time neutron radiography and artificial neural network. Nucl. Instrum. Methods Phys. Res. Sect. A Accel. Spectrom. Detect. Assoc. Equip. 940, 344–350.
Nguyen, T.-D., et al., 2023. Artificial intelligence algorithms for prediction and sensitivity analysis of mechanical properties of recycled aggregate concrete: a review. J. Build. Eng. 66, 105929.
Nguyen, Q.D., et al., 2024. Fracture behaviour assessment of high-performance fibre-reinforced concrete at high strain rates using interpretable modelling approaches. Heliyon 10 (2), e24704.
Ni, B., et al., 2025. A review on properties and multi-objective performance predictions of concrete based on machine learning models. Mater. Today Commun., 112017
Obianyo, I.I., et al., 2024. Machine learning models for predicting the compressive strength of agro-waste stabilized bricks for sustainable buildings. Discover Civil Engineering 1 (1), 45.
Pamu, Y., Svsndl, P., 2024. ANN-based prediction of compressive strength in glass wool reinforced brick. In: E3S Web of Conferences, vol. 559.
Parra-López, C., et al., 2024. Integrating digital technologies in agriculture for climate change adaptation and mitigation: state of the art and future perspectives. Comput. Electron. Agric. 226, 109412.

Pence, I., et al., 2024. Animal-based CO2, CH4, and N2O emissions analysis: machine learning predictions by agricultural regions and climate dynamics in varied scenarios. Comput. Electron. Agric. 226, 109423.
Pezo, L., Arsenović, M., Radojević, Z., 2014. ANN model of brick properties using LPNORM calculation of minerals content. Ceram. Int. 40 (7, Part A), 9637–9645.
PN, M.L., et al., 2018. Energy efficient production of clay bricks using industrial waste. Heliyon 4 (10), e00891.
Praburanganathan, S., Chithra, S., Simha reddy, Y.B., 2022. Synergistic effect on the performance of ash-based bricks with glass wastes and granite tailings along with strength prediction by adopting machine learning approach. Environ. Sci. Pollut. Control Ser. 29 (36), 54193–54218.
Probst, P., Boulesteix, A.-L., Bischl, B., 2019. Tunability: importance of hyperparameters of machine learning algorithms. J. Mach. Learn. Res. 20 (53), 1–32.
Rachele, J.N., et al., 2021. Using machine learning to examine associations between the built environment and physical function: a feasibility study. Health Place 70, 102601.
Rathnayaka, M., et al., 2024. Machine learning approaches to predict compressive strength of fly ash-based geopolymer concrete: a comprehensive review. Constr. Build. Mater. 419, 135519.
Razavi, S., et al., 2021. The future of sensitivity analysis: an essential discipline for systems modeling and policy support. Environ. Model. Software 137, 104954.
Sarani, N.A., et al., 2023. Physical-mechanical properties and thermogravimetric analysis of fired clay brick incorporating palm kernel shell for alternative raw materials. Constr. Build. Mater. 376, 131032.
Saravanan, J., Rao, P.V., 2023. Past investigations on development of sustainable bricks – a comprehensive review. Sustainable Chemistry for the Environment 3, 100030.
Sarkar, N., Dasgupta, K., 2024. Machine learning-based sensitivity analysis of engineering demand parameters for a reinforced concrete wall-frame building. Structures 70, 107477.
Sathiparan, N., Jeyananthan, P., 2023. Predicting compressive strength of cement-stabilized earth blocks using machine learning models incorporating cement content, ultrasonic pulse velocity, and electrical resistivity. Nondestr. Test. Eval.
Sathiparan, N., et al., 2025. Response surface regression and machine learning models to predict the porosity and compressive strength of pervious concrete based on mix design parameters. Road Mater. Pavement Des. 26 (3), 618–657.
Setvati, M.R., Hicks, S.J., 2022. Machine learning models for predicting resistance of headed studs embedded in concrete. Eng. Struct. 254, 113803.
Sharma, U., Gupta, N., Verma, M., 2023. Prediction of compressive strength of GGBFS and Flyash-based geopolymer composite by linear regression, lasso regression, and ridge regression. Asian Journal of Civil Engineering 24 (8), 3399–3411.
Sifan, M., et al., 2023. Efficient mix design method for lightweight high strength concrete: a machine learning approach. Structures 55, 1805–1822.
Srisuwan, A., Phonphuak, N., Saengthong, C., 2018. Improvement of thermal insulating properties and porosity of fired clay bricks with addition of agricultural wastes. Suranaree Journal of Science and Technology 25, 49–58.
Sufian, M., et al., 2021. An experimental and empirical study on the use of waste marble powder in construction material. Materials 14 (14), 3829.
Taffese, W.Z., Sistonen, E., Puttonen, J., 2015. CaPrM: carbonation prediction model for reinforced concrete using machine learning methods. Constr. Build. Mater. 100, 70–82.
Taffese, W.Z., Sistonen, E., 2017. Machine learning for durability and service-life assessment of reinforced concrete structures: recent advances and future directions. Autom. ConStruct. 77, 1–14.
Taffese, W.Z., Espinosa-Leal, L., 2022. Prediction of chloride resistance level of concrete using machine learning for durability and service life assessment of building structures. J. Build. Eng. 60, 105146.
Tipu, R.K., Panchal, V.R., Pandya, K.S., 2022. Prediction of concrete properties using machine learning algorithm. J. Phys. Conf. 2273 (1), 012016.
Vasić, M.V., Pezo, L.L., Radojević, Z., 2020. Optimization of adobe clay bricks based on the raw material properties (mathematical analysis). Constr. Build. Mater. 244, 118342.
Vasić, M.V., et al., 2023. Influence of coal ashes on fired clay brick quality: random forest regression and artificial neural networks modeling. J. Clean. Prod. 407, 137153.
Wahab, R.A.A., et al., 2021. Physical properties of low energy consumption fired industrial waste-clay bricks from cockle shells and soda lime silica glass. In: AIP Conference Proceedings. AIP Publishing.
Wong, L., Mui, K.W., Tsang, T.W., 2022. Updating Indoor Air Quality (IAQ) assessment screening levels with machine learning models. Int. J. Environ. Res. Publ. Health 19, 5724.