On the Use of Machine Learning Technique to Appraise Thermal Properties of Novel Earthen Composite for Sustainable Housing in Sub-Saharan Africa

Assia Aboubakar Mahamat[1,2](✉) and Moussa Mahamat Boukar[1]

[1] Nile University of Nigeria, Abuja, Federal Capital Territory, Nigeria
aassia@aust.edu.ng

[2] African University of Science and Technology (AUST), Abuja, Federal Capital Territory, Nigeria

Abstract. Earthen based bio-composite reinforced with agricultural waste represent a very important alternative for eco-friendly sustainable building materials. In addition, to the environmental-friendly aspect the use of agro-waste plays a major role in waste management primary by reducing the price related to the waste proper disposal. A novel bio-composite was modeled and tested for its thermal properties to enable the comfort to its habitant. The experimental results were used as primary data to test, train and validate two different machine learning algorithms. The two machine learning models used to predict the thermal conductivity are decision tree regressor (DTR) and random forest (RF). Various inputs were used based on their importance/relationship with the predicted output. The machine learning models were compared based on their efficiency/performance via the evaluation metrics R^2, RMSE, MSE and MAE. Decision tree displayed $R^2 = -0.26$, RMSE $= 0.077$, MSE $= 0.006$ and MAE $= 0.05$ while random forest displayed values $R^2 = -17.7$, RMSE $= 0.197$, MSE $= 0.039$ and MAE $= 0.119$. The results corroborate that both RFR and DTR performed poorly during the predictions, thus they are not suitable for similar composite with the selected input variables.

Keywords: decision tree · random forest · earthen composite · thermal conductivity

1 Introduction

Access to standard housing is an essential human right. This fundamental right must be fulfilled without destroying the nature. Therefore, the idea of learning from and emulating certain aspects of nature to solve human problems. This sustainability concept derives from the fact that nature has sustained itself for several million years unlike human-designed systems that fail to operate within a long period. To achieve standard housing, ordinary Portland cement (OPC) is the most used material. However, production and utilization of the OPC releases greenhouses gas (GHG) and significantly destabilizes the

A. Seeam et al. (Eds.): InterSol 2023, LNICST 541, pp. 161–170, 2024.
https://doi.org/10.1007/978-3-031-51849-2_11

environment. Alkali activated binders have gained a lot of attention as ‘green’ alternative to ordinary Portland cement (OPC). This technology is based on the dissolution of aluminosilicate in the presence of alkaline activator. The alkali activation process involves alkali solutions (two-part alkali activated binders) or solid alkali source (one-part alkali activated binders) [1]. The use of two-part alkali activation process is corrosive, difficult, and not user-friendly [2]. Meanwhile, one-part alkali activation process requires solid alkali source with the aluminosilicate precursor and water. During this process, the dry ingredients are mixed before addition of water. This is similar to the preparation of OPC, and it enables better cast-in-situ features which are among the main requirement in the construction field. During the mixture chemical reactions take place, among which the dissolution of the aluminosilicate precursor activated by the alkaline source (K_2CO_3) [3], followed by polycondensation and formation of amorphous network [4]. Artificial intelligence (AI) approaches are considered as an alternative or improvement of traditional statistical methods. They are used by many researchers, and they were found to be more efficient in prediction as compared to the empirical/traditional methods. Machine Learning (ML) is defined as a set of algorithmic structures enabling computer systems to learn and train their performances through established patterns [5]. Machine Learning is a subset of Artificial Intelligence (AI) that has been substantially used in different fields, especially in civil engineering to resolve complex problems related to materials science [6], structural engineering, geotechnical engineering etc. [7]. Machine Learning is a subgroup of AI that is divided into supervised and unsupervised techniques. Machine Learning application have gained a lot of interest in the medical field, engineering, etc. due to their high accuracy/efficiency in data processing. Machine Learning techniques are used in the construction field mainly to predict mechanical properties such as compressive and flexural strength [8]. They’re used to save time of the experiments because many experiments require long curing period before getting the results and those experiment are costly due to their destructive nature. Thus, Machine Learning can be used because the algorithms require low/minimal human interference during the training and decision making. However, the selection of the appropriate features with the appropriate inputs is the bedrock for the predictions’ efficiency. However, some of the ML models that are commonly used in the construction field are: Artificial Neural Network (ANN) [9], support vector machine (SVM) [10], decision trees (DT) [11]. This study intends to develop decision tree regression (DTR) and random forest (RF) models to predict the thermal capacity of an alkali-activated earthen composite.

Decision tree regression (DTR) model uses training data to create a model in the shape of a tree in which each internal node represents a test, the branches represent the outcome while the leaves represent the decisions during the training [12]. This kind of modeling involves two steps: The first step is the building tree step which entails dividing the training dataset into well-defined fragments. The second step is the tree pruning where branches from the built tree are selected to be removed to lower the dimension of the decision trees fragments that are non-critical or irrelevant [13].

Random Forest Regression (RFR) is a supervised ML algorithm that uses classification and regression trees (CART) for prediction. In RFR, inputs are randomly selected at each node to grow a tree. The accuracy of the individual classifiers and the dependence between them. Random Forest regression can handle large features with small samples, this characteristic makes RF a suitable option for this study where the primary data set are not large. Random Forest regression model is used to predict the compressive, splitting tensile and flexural strength of concrete incorporated with metakaolin. During their study, the evaluation metrics was the Coefficient of determination (R^2) and found to be 0.99, 0.98 and 0.99 for compressive strength, splitting tensile strength and flexural strength respectively [14]. For all the predicted properties the value of R^2 is all close to 1, that shows the efficiency of those models for this kind of prediction. The compiled data set consists of the measurements for the thermal conductivity which are obtained from experimental results thus primary data set. Due to the scarcity of existing literature on similar earthen composite the authors generated the primary data set. The use of the primary data set is also governed by the desire to avoid errors that would be generated from the mapping when using secondary data set.

The aim of the investigation is to explore ecofriendly alternative materials for sustainable construction in the sub-Saharan African region to ease the access and break down the prices of standard housing in that region. Henceforth the techniques used are green techniques with renewable materials and low energy requirement during manufacturing. Thus, the scaling up to industrial level wouldn't be challenging in that region and worldwide.

The oncoming sections of this investigation are organized as follows: Sect. 1 provides a brief literature review of the different ML algorithms used to predict the behavior of the novel earthen composite plus the contribution of this investigation to the knowledge gaps. In Sect. 2, the materials and methods used are presented with a brief description of the quality assessment used during this investigation. In Sect. 3, the results obtained from the training and testing of various models used are presented and discussed. In Sect. 4, important conclusions from this investigation are presented.

2 Materials and Methods

2.1 Experimental Part

a) **Bio-composite manufacture**

The materials used to fabricate the composite are excavated soil, Borassus fruit fiber and synthetic potash (KCO_3). The soil was obtained from an excavation in a construction field in the FCT, Nigeria. The soil was obtained at no cost and its use is intended to reduce waste placement. The soil was sieved to remove coarse particles. After sieving the size of the soil's particles retained for this investigation are within the range of 2 mm–2 μm. The chemical stabilisation technique used during this investigation is the alkali activation of the matrix. Therefore, synthetic potash (KCO_3) was used as activator with 99.8% purity. The activator content was fixed at 3wt% [1]. The fiber was manually extracted from the Borassus fruit and didn't undergo any chemical treatment. Fiber content of 0%, 0.5wt%, 0.75wt% and 1wt% were used during the samples fabrication [3].

In the manufacturing process the dry materials were mixed in a laboratory mixer before the addition of distilled water at room temperature (27 °C). The water's quantity used during manufacture was obtain from the optimum moisture content (OMC) of the soil which was equal to 18wt%. The paste was left on to cool for 5 min because of the exothermic reaction before being poured into mould and pressed with a hydraulic press prior to thermal conductivity testing. The samples were oven dried for 7-, 14- and 90 days at 60 °C. Each sample was replicated 5 times for each test for the various composition at different curing days. The overall manufacturing procedure is shown in Fig. 1.

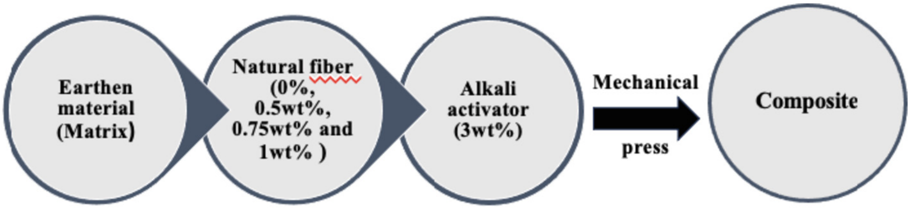


Fig. 1. Flowchart describing the production of the composite.

b) **Thermal capacity analysis**

For construction materials, thermal properties govern the thermal comfort of the building's habitant [15]. The thermal properties were assessed with the aid of the hot-disk technique (TPS hot disk equipment). Prior to the thermal properties experiments, the samples were enveloped in a polystyrene foil conserved in a closed environment to avoid contact with moisture. During the experiment the thermal conductivity (λ), and specific heat (Cp) were obtained in a single measurement meanwhile the density (ρ) was calculated, and the thermal diffusivity (α) was deduced from the Eq. (1):

$$\lambda = \alpha \, \rho \, Cp \tag{1}$$

To accurately predict the thermal conductivity (λ) or Y six independent variables were used as inputs X1, X2, X3, X4, X5 and X6 denoting the activator, fiber content, curing days, density, specific heat, and thermal diffusivity respectively (Table 1).

Table 1. Experimental data used as inputs for data training and validation of the ML models for the thermal conductivity's prediction.

Inputs	Output
X1 (activator), X2 (fiber content), X3 (curing days), X4 (density), X5 (specific heat) and X6 (thermal diffusivity)	Y (Thermal conductivity)

2.2 Machine Learning Models

The models used for the thermal conductivity's prediction are regression-based models namely the decision tree regressor and the random forest regression. They're used because of their efficiency to predict accurately outputs with limited dataset. Figure 2 shows the pseudocode of both models used during the prediction.

```
def random_forest_regression(data, n_trees, n_features):
  # Initialize the forest as a list of trees.
  forest = []

  # Iterate over the number of trees.
  for i in range(n_trees):
    # Bootstrap a sample from the data.
    bootstrap_sample = data.sample(frac=1, replace=True)

    # Select a random subset of features.
    features = bootstrap_sample.columns.sample(n=n_features)

    # Build a decision tree on the bootstrap sample.
    tree = decision_tree_regression(bootstrap_sample[features])

    # Add the tree to the forest.
    forest.append(tree)

  # Return the forest.
  return forest
```

a)

```
def decision_tree_regression(data):
  # Initialize the tree as a root node.
  tree = Node()
  # Recursively build the tree.
  for attribute in data.columns:
    split_value = find_best_split_value(data, attribute)
    tree.add_child(attribute, split_value)
  # Return the tree.
  return tree
def find_best_split_value(data, attribute):
  # Calculate the standard deviation of the target variable.
  target_std = data["target"].std()
  # Initialize the best split value as the mean of the target variable.
  best_split_value = data["target"].mean()
  # Iterate over all possible split values.
  for split_value in data[attribute].unique():
    # Calculate the standard deviation of the target variable in each
split.
    split_std = data[data[attribute] == split_value]["target"].std()
    # If the standard deviation of the split is less than the standard
deviation of the whole dataset, then update the best split value.
    if split_std < best_split_value:
      best_split_value = split_value
  return best_split_value
```

b)

Fig. 2. Pseudocode of the used models: a) Random Forest regression and b) Decision Tree regression

a) **Decision Tree Regression (DTR)**
   A decision tree regressor is a class of supervised machine learning technique that fulfills the task of both classification and regression. It's classification form is a tree-structure that displays the given dataset. The internal nodes of the tree represent the dataset [16].
b) **Random Forest Regressor (RFR)**
   A random forest regressor is also a type of supervised machine learning technique that uses averaging to increase the accuracy of the prediction and control over-fitting. Numerous decision trees are generated from the data set to enable ensemble learning within the decision tree framework. It leads to excellent predictions due to the average of the results that are used to get a new result [13].
c) **Quality assessment of results**
   To assess the efficiency of the developed machine learning models in this investigation, four performance indicators are used: Mean Square Error (MSE), Root Mean Square Error (RMSE), coefficient of determination (R^2) and Cross Validation Score (CV Score).

   - **Mean Square Error (MSE)**
     The mean square error (MSE) displays the quantity of error in the statistical models between the input variables and the predicted output. It calculates the average squared difference when the MSE is equal to 0 it means that the model is error-free. The MSE is obtained based on the following equation:

$$\mathrm{MSE} = \frac{1}{n}\sum\nolimits_{i=1}^{n}(\mathrm{Tm} - \mathrm{To})^2 \tag{2}$$

- **Root Mean Square Error (RMSE)**
The root mean square error (RMSE) is a metric indicating the dispersion of the prediction errors and the way they are, meaning the distance between the regression and the data points. When the value of the RMSE is low it indicates high accuracy of the prediction. The RMSE is expressed by the following equation:

$$\mathrm{RMSE} = \sqrt{MSE} \tag{3}$$

- **Mean Absolute Error (MAE)**
The mean absolute error (MAE) is the arithmetic mean of the deviation [17]. The MAE measures the errors between the input and predicted values, and it is defined as:

$$\mathrm{MAE} = \frac{1}{n}|\mathrm{Tm} - \mathrm{To}| \tag{4}$$

where n is the number of samples, T_o and T_m are the observed and modeled values.
- **Coefficient of determination (R^2)**
The coefficient of determination (R^2) is a metric that evaluates the statistical relationship between the experimental and predicted values. Its value is comprised between 0 and 1, the higher its value means higher correlation. The equation to determine the R^2 is given in Assia et al. investigation [10].
- **Experimental framework**
Intel Core (TM) i5-4790 CPU, 3.60 GHz and 4 GB RAM was the environment used for the execution and the experiments were carried out in Python 2.7.12 during this investigation. The summarized stages undertaken during this study is displayed in Fig. 3.

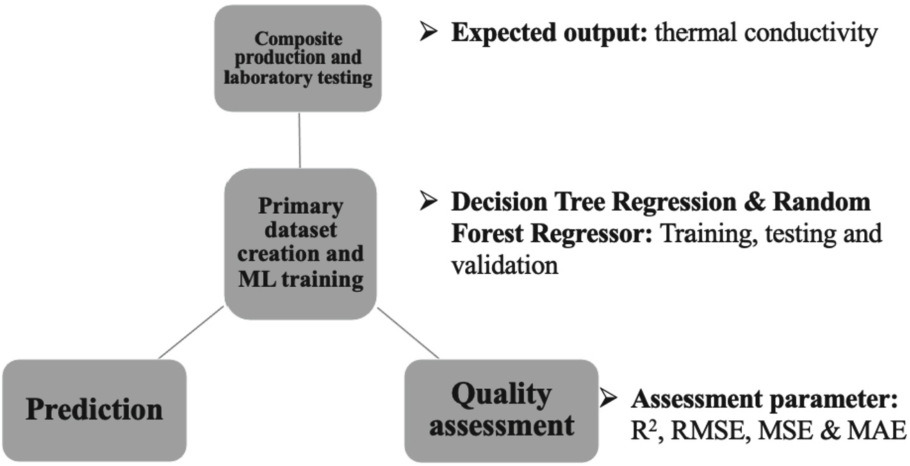
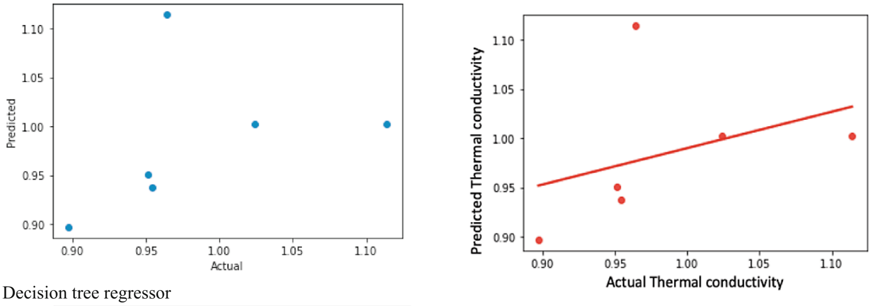
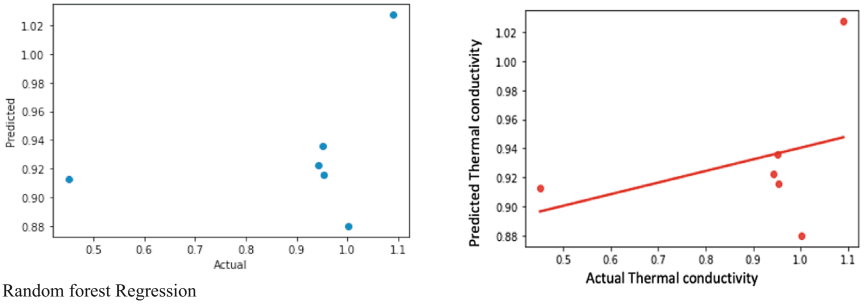


Fig. 3. Flow diagram representing the experimental scheme summarizing the various stages of this study.

3 Results and Discussions



a) Decision tree regressor



b) Random forest Regression

Fig. 4. Predicted vs experimental values of the thermal conductivity via: a) decision tree regressor and b) random forest regressor.

The results of the developed models to predict the thermal conductivity of the novel earthen composite is shown in Fig. 4. These results validate the successful training, testing and validation of the RF and DTR during the prediction of the thermal conductivity.

The models' quality assessment showcased in Fig. 4 corroborates that the error values to assess the models exhibited $R^2 = -0.26$, RMSE = 0.077, MSE = 0.006 and MAE = 0.05 for the decision tree regressor (DTR) model and $R^2 = -17.7$, RMSE = 0.197, MSE = 0.039 and MAE = 0.119 for random forest regressor (RFR) model. The negative value obtained for the coefficient of determination from both models indicate the weak correlation between the chosen inputs and output during this prediction. However, the value obtained for the DTR was closer to 0 than the value obtained for the RF highlighting the inaccuracy level of both models for this prediction. Because for a higher accuracy the model's coefficient of determination should be closer to 1.

The negative error obtained from the thermal conductivity's prediction by the DTR and RFR may be because the DTR takes all possible consequences into account before the ultimate prediction based on a comprehensive analysis [18]. Those possible consequences can vary negatively with the variation of heat capacity and the thermal resistivity. However, RFR reinforces the diversity of the basic model and improve the prediction by variance reduction [19, 20].

The predictions results obtained from the DTR and RFR demonstrated negative values of R^2 which imply that the model has very high error level (Table 2). Because when the coefficient of determination (R^2) is higher than 0.5 and closer to 1, it indicates that the inputs chosen to train the model have great significance on the output. However, our results displayed negative values of R^2 resulting in the bad fitting obtained from DTR and RFR models showing the inefficiency of the used models in this kind of prediction. These results align with previous work [20], where RF is supposed to reinforce the diversity of basic models and improves prediction accuracy by variance reduction. Meanwhile during the present study, the variance reduction wasn't taken into consideration.

Table 2. Evaluation metrics and efficiency comparison of the DTR and RFR models for the prediction of the thermal conductivity

Errors/Models	R^2	RMSE	MSE	MAE
Decision Tree	−0.26	0.077	0.006	0.05
Random Forest	−17.7	0.197	0.004	0.119

4 Conclusion

Earthen materials constitute the best alternative in terms of eco-friendliness, durability and low-energy consumption during production and life service to conventional construction materials. Depending on the manufacturing process, the earthen material can be used partially or completely alone as a construction material. In this study, the application of ML models for the prediction of thermal conductivity was assessed. In the initial part of the study, the authors created a database from experimental results carried out on manufactured specimens. The specimens were produced from earthen matrix reinforced with natural Borassus as strengthening process and produced through one-part alkali activation. The manufactured specimens were cured for different periods. The primary database created by the authors consisted of thermal conductivity of unreinforced and reinforced composite obtained at different curing days. Generated inputs parameters by the authors varied based on their signification on the output parameter. Two (2) machine learning (ML) models: decision tree regression (DTR) and random forest regressor (RF) were used in this study to predict/evaluate the thermal conductivity of the manufactured composite. A comparison of the two models was carried out based on their efficiency and that was evaluated using four (4) evaluation metrics: coefficient of determination (R^2), mean square error (MSE) root mean square error (RMSE), and mean absolute error (MAE). The results obtained from the various metrics show that the inputs variable chosen to predict the thermal conductivity through those models are not significant or efficient. It shows that the models used are not very efficient in predicting similar property. This may be due to the limited dataset or from the correlation existing between the inputs and output. Thus, it can be concluded that DTR and RFR models are not efficient to predict the thermal conductivity of earthen composite with the inputs variables selected during this investigation. Therefore, for future work the exploration of different inputs can be taken into consideration, or the dataset size should be increased significantly. Also, the various properties of the novel composite can be assessed differently because the composite is made of novel materials however the testing procedure used are the standard used for conventional materials. Henceforth, by altering the testing standard different results can be obtained too.

In sum, this study demonstrates the capability of DTR and RFR models in the prediction of novel composite manufactured from earthen materials, reinforced with vegetal natural fiber, and produced through geo-polymerization technique. The results demonstrated the inefficiency of the models for such prediction. This inefficiency can be attributed to many factors such weak correlation between input-output, limited dataset or inappropriate testing standard used for such novel composite Therefore, for future prediction of similar composite other models can be used for better performance or the input parameters should be changed for a larger dataset. The main limitation of this study is the inexistant literature because significant research has not been carried out in the application of ML models for novel earthen composite reinforced with natural fiber. Also, the geo-polymerization is a technique that is not widely used the sub-Saharan Africa.

Acknowledgments. This investigation is funded by the L'Oréal-UNESCO *for women in science*, sub-Saharan Africa young talents postdoctoral fellowship 2022.

References

1. Mahamat, A., et al.: Alkali activation of compacted termite mound soil for eco-friendly construction materials. Heliyon **7**(3) (2021). https://doi.org/10.1016/j.heliyon.2021.e06597
2. Abdollahnejad, Z., Mastali, M., Luukkonen, T., Kinnunen, P., Illikainen, M.: Fiber-reinforced one-part alkali-activated slag/ceramic binders. Ceram. Int. **44**(8), 8963–8976 (2018). https://doi.org/10.1016/j.ceramint.2018.02.097
3. Mahamat, A., et al.: Assessment of hygrothermal and mechanical performance of alkali activated Borassus fiber reinforced earth-based bio-composite. J. Build. Eng. **62**, 105411 (2022). https://doi.org/10.1016/j.jobe.2022.105411
4. Barone, G., Mazzoleni, P., Finocchiaro, C.: FT-IR study of early stages of alkali activated materials based on pyroclastic deposits (Mt. Etna, Sicily, Italy) using two different alkaline solutions. Constr. Build. Mater. **262** (2020). https://doi.org/10.1016/j.conbuildmat.2020.120095
5. Shoar, S., Chileshe, N., Edwards, J.D.: Machine learning-aided engineering services' cost overruns prediction in high-rise residential building projects: application of random forest regression. J. Build. Eng. **50** (2022). https://doi.org/10.1016/j.jobe.2022.104102
6. Mahamat, A.A., Boukar, M.M.: Machine learning techniques versus classical statistics in strength predictions of eco-friendly masonry units. In: 16th International Conference on Electronics Computer and Computation (ICECCO 2021) (2021)
7. Xie, J., Huang, J., Zeng, C., Huang, S., Burton, G.J.: A generic framework for geotechnical subsurface modeling with machine learning. J. Rock Mech. Geotech. Eng. (2022). https://doi.org/10.1016/j.jrmge.2022.08.001

8. Ben Chaabene, W., Flah, M., Nehdi, M.L.: Machine learning prediction of mechanical properties of concrete: critical review. Constr. Build. Mater. **260**, 119889 (2020). https://doi.org/10.1016/j.conbuildmat.2020.119889
9. Tosee, S.V.R., et al.: Metaheuristic prediction of the compressive strength of environmentally friendly concrete modified with eggshell powder using the hybrid ANN-SFL optimization algorithm. Materials **14**(20) (2021). https://doi.org/10.3390/ma14206172
10. Mahamat, A., et al.: Machine learning approaches for prediction of the compressive strength of alkali activated termite mound soil. Appl. Sci. **11**(11), 4754 (2021). https://doi.org/10.3390/app11114754
11. Khan, K., Ahmad, W., Amin, M.N., Ahmad, A.: A systematic review of the research development on the application of machine learning for concrete. Materials **15**(13) (2022). https://doi.org/10.3390/ma15134512
12. Mangalathu, S., Jang, H., Hwang, S.H., Jeon, J.S.: Data-driven machine-learning-based seismic failure mode identification of reinforced concrete shear walls. Eng. Struct. **208** (2020). https://doi.org/10.1016/j.engstruct.2020.110331
13. Breiman, L.: Classification and Regression Trees, 1st edn. Routledge, New York (1984). https://doi.org/10.1201/9781315139470
14. Shah, H.A., et al.: Application of machine learning techniques for predicting compressive, splitting tensile, and flexural strengths of concrete with metakaolin. Materials **15**(15), 5435 (2022). https://doi.org/10.3390/ma15155435
15. Moussa, H.S., Nshimiyimana, P., Hema, C., Zoungrana, O., Messan, A., Courard, L.: Comparative study of thermal comfort induced from masonry made of stabilized compressed earth block vs conventional cementitious material. J. Miner. Mater. Charact. Eng. **07**(06), 385–403 (2019). https://doi.org/10.4236/jmmce.2019.76026
16. Anysz, H., Brzozowski, Ł., Kretowicz, W., Narloch, P.: Feature importance of stabilised rammed earth components affecting the compressive strength calculated with explainable artificial intelligence tools. Materials **13**(10) (2020). https://doi.org/10.3390/ma13102317
17. Kang, M.C., Yoo, D.Y., Gupta, R.: Machine learning-based prediction for compressive and flexural strengths of steel fiber-reinforced concrete. Constr. Build. Mater. **266** (2021). https://doi.org/10.1016/j.conbuildmat.2020.121117
18. Chen, L., Tran, H., Batra, R., Kim, C., Ramprasad, R.: Machine learning models for the lattice thermal conductivity prediction of inorganic materials. Comput. Mater. Sci. **170** (2019). https://doi.org/10.1016/j.commatsci.2019.109155
19. Bang, H.T., Yoon, S., Jeon, H.: Application of machine learning methods to predict a thermal conductivity model for compacted bentonite. Ann. Nucl. Energy **142** (2020). https://doi.org/10.1016/j.anucene.2020.107395
20. Li, K.Q., Liu, Y., Kang, Q.: Estimating the thermal conductivity of soils using six machine learning algorithms. Int. Commun. Heat Mass Transfer **136** (2022). https://doi.org/10.1016/j.icheatmasstransfer.2022.106139