Predicting Energy Performances of Buildings' Envelope Wall Materials via the Random Forest Algorithm

Aseel Hussien[1*], Wasiq Khan[2], Abir Hussain[1&2], Panos Liatsis[3] Ahmed Al-Shamma'a[4] and Dhiya Al-Jumeily[2]

[1] College of Engineering, University of Sharjah, UAE
[2] Department of Computer Science, Liverpool John Moores University, UK
[3] Electrical Engineering and Computer Science, Khalifa University of Science and Technology, UAE
[4] Chancellor/ University of Khorfakkan UAE

*Correspondence: ahussien@sharjah.ac.ae

Abstract:

Purpose: Numerous simulation software has been used to evaluate energy performance with 12% of the research focusing on long-term energy consumption prediction. This paper aims to utilize machine learning to predict the energy performance of building envelope wall materials over extended periods.

Methodology: In our work, machine learning model learns from a large set of building envelopes simulated using the Integrated Environmental Solutions Virtual Environment as follows:

- The data generation via building performance analysis applications IES-VE software using different wall construction scenarios and stored as a CSV file for the second stage.
- The dataset is partitioned into training (70%) and testing (30%) samples using 10-fold cross-validation.
- Extensive simulations are performed to optimize the model parameters in predicting the energy efficiency of buildings via ML algorithms.

Findings: Machine Learning models can also be used to predict the impact of building design and construction characteristics on energy consumption, showing that factors such as wall thickness, orientation, and thermal mass indicated lower relative standard error (<0.001); however, not all of them were statistically significant ($p>0.05$). While the overall model indicates statistical significance ($p = 2e-16$), the multivariate linear regression model produces R^2 value of 0.42, indicating a weak relationship between predictor variables and target attributes.

Originality: The utlisation of Random forset algorithm for the wall enveolp energy consuption

Research Implecation: different to other technqiues, our proposed approach addressed the issue related to building envelop for new constructions to assist professional from construction industry.

Keywords: Energy; Building Envelop, Indoor Environment Quality, Machine Learning; Multivariate Linear Regression, Random Forest

1. Introduction

Buildings are constructed to improve occupant health and well-being while consuming minimal construction materials and energy usage. Globally, indoor air quality is one of the crucial health matters, as individuals spend around 85–90% of their daily activities in indoor spaces for different purposes (Dong, et al., 2022). In this context, the choice of building materials and design approach affect thermal comfort and the entire indoor environment quality. In the global market, where potential energy may be significantly costly, guidance for the Architecture, Engineering, and Construction industry (AEC) will be crucial for refurbishment and new buildings (Hussien, et al., 2023). Thus, using appropriate construction materials will considerably reduce energy demands, lower CO_2 emissions, and improve thermal comfort in buildings.

According to the geographical climate zone, building materials can be categorized into insulating and preserving (Jannat, et al., 2020; Aditya, et al., 2017). The former prevents outer heat gain (hot climate zones), while the latter targets renal heat loss (cold climate zones). Indeed, the appropriate use of building materials is critical, effectively managing the temperature gradients between indoor and outdoor environments. In the heat insulation climate zone, where high humidity, high temperature, and solar radiation are the main stressors, materials provide shading and heat protection

(Sudhakar, et al., 2019). Several researchers argued the possibility of controlling the indoor thermal comfort environment without using any mechanical system by maintaining the indoor temperature within adaptive comfort levels (De Dear & Brager, 2022). In the latter case, the building's envelope and other variables relating to the wall thickness and climate condition play an essential role in adjusting indoor air temperature (Deshko, et al., 2020). Thus, to achieve average indoor thermal comfort, suitable materials for the building envelope would be a legitimate option (Jannat, et al., 2020; Hussien, et al., 2023). A research study by (Hwang & Chen, 2022), discussed the effect of glazed facades on energy consumption and thermal comfort in office buildings in a hot climate. The results showed that it is essential to link the facades' design with the external environment to ease energy consumption and improve thermal comfort. Different locations offered different thermal comfort when using identical glazed facades.

The study demonstrated an evident lack of connection between the façade design and the building environment which led to increased energy usage and occupant discomfort. The researchers in (Wu, et al., 2020) evaluated the energy consumption between two office rooms with solar radiation and a controlled Heating, Ventilation Airconditioning (HVAC) system via Predicted Mean Vote (PMV) based; the results verified that the PMV-based control provided better thermal comfort in the highly glazed office room, and still shows savings in the energy usage. While a study by (Merabet, et al., 2021) discussed the conflict objectives when trying to reduce energy consumption while maintaining comfortable conditions in buildings, this required intelligent system design. Thus, the study implemented artificial intelligence (AI) techniques to find the AI-based control between energy usage and appropriate indoor comfortable level for buildings' occupants. On the other hand, (Alghamdi, et al., 2022) aimed to determine the most influential architectural design parameters that improve thermal comfort and reduce energy usage. The result showed that the different selections of construction materials assisted in reducing the temperature and improved occupant comfort by 20%, reducing energy consumption by 41%. The study by (Nematchoua, et al., 2014) compared the energy consumption and thermal comfort between traditional and modern buildings. The results demonstrated that traditional buildings were more comfortable with less humidity, while the modern houses required more air ventilation, leading to more energy consumption. This was due to the different construction materials used for both types of houses. The effect of the building envelop on energy usage and thermal comfort was discussed by (Mirrahimi, et al., 2016) in a multi-story building in Malaysia; the results showed that buildings' wall thickness has a crucial impact in determining the energy consumption due to the thermal resistance and the rate of the heat transmission from outdoor to indoor. Despite global efforts to decrease energy usage and reduce CO_2 emissions, Green House Gag (GHG) emissions and energy usage have risen over the years. Among all industries, the construction industry has a crucial role in global energy demand. As such, improving building efficiency has become a vital objective in decreasing gas emissions and consumption of fossil fuels, in addition to lowering 'buildings' CO_2 emissions at an early design stage. Moreover, efficient energy management could enhance the performance of existing stock. For instance, a potential solution would involve detailed energy estimations for ideal decision-making.

Nevertheless, explanatory energy consumption in buildings remains challenging due to the variety of factors that impact consumption. These include physical properties, weather conditions, and the energy usage performance of the building occupants. Furthermore, numerous tools and software have been developed to perform energy simulations in buildings, relying on physical models that require thermodynamic inputs for detailed energy modeling and analysis. For example, EnergyPlus and Design builder have widely used software platforms that calculate energy consumption based on environmental parameters, construction details, HVAC systems, and climate conditions (Jannat, et al., 2020). However, these software have limitations as comprehensive data may not always be available for all building properties and environmental parameters, leading to weaker energy predictions (Mechouet, et al., 2021). In (Al-Homoud, 2001), the author reviewed the energy analysis methods and the possible applications for energy simulation that help decision-making; the results demonstrated several difficulties, like; the non-linear correlation among building elements that affect the selection of the appropriate optimization techniques, in addition to the buildings being composed of several components, and an optimal solution of all components is not feasible. To be achieved, it should be noted that these software need a series of individual optimum decisions resulting in complicated interpretations of the problem and incompetence to reach an actual optimum. As such, these methods calculate energy consumption depending on current environmental parameters. However, some of these data may not be available during the simulation, leading to errors in the predication of energy usage, as data may not be available during the simulation.

On the other hand, the utilization of Machine Learning (ML) to predict energy consumption does not require the utilization of comprehensive and complete information in a similar way to the current available building software; since it learns from available/historical data, and as a result it gained many research interests recently including project management improvement, reduce risk, site safety management, cost estimation, and schedule management (Jannat, et al., 2020). However, our extensive literature review indicated that there is limited literature available on the use of ML in the context of building envelopes and construction materials.

Akinosho et al. (2020) compared two buildings, where data was simulated for their energy usage generated via Energy Plus and Ecotec. In this case, the authors studied high-performance ML models for predicting the cooling and heating loads of buildings, and optimizing their hyperparameters in regard to the specific application. The study implemented grid search coupled with a cross-validation method to evaluate the performance of the combined model parameters. The results revealed the ability to recognize insignificant variables, resulting in efficient model fitting. Another study by (Olu-Ajayi, et al., 2022) presented a comprehensive overview of ML approaches, including artificial neural networks, support vector machines, and Gaussian process regression, which are popular in explanatory prediction and enhancement of energy performance in buildings. The study revealed that determining the best ML model is not straightforward and model performance depends on the sample number and the appropriate tuning of the ML method parameters. Another study (Mousavi, et al., 2022) focused on analyzing the advances in the state-of-the-art machine learning models combined with Building Information models (BIM), computer vision, and their related technologies to enable the digital evolution of tunneling and underground construction. The results revealed a significant gap in the literature related to the use of ML in BIM.

Bhamare et al (2021) employed ML model to predict the thermal performance of building roofs. Five hundred data were generated via numerical simulation using five ML models, including extreme gradient, extra trees regression, random forest regression, gradient boosting, and gate boost. The results showed that gradient boosting was the best-performing model providing training efficiency, good generalization in the testing data, and robustness. In another study by (Zhang, et al., 2022) , ML was used to predict shear walls' strength and deformation capacity. Their simulation results revealed that RF model was the best prediction method for shear wall strength and deformation capacity estimation.

Despite the extensive utilization of ML in predicting energy usage, our literature review indicated that there is limited literature available on the use of ML in the context of building envelopes and construction materials.

This study focuses on the various characteristics of the wall materials used in building envelopes, which affect energy efficiency and indoor' thermal comfort in the UAE's hot climate aiming to utilize ML to predict the energy performance of building envelope wall materials over extended periods.,

As such, the novelty proposed in this paper involves:

- To the best knowledge of the authors, we are the first to implement ML models for building envelope energy prediction.
- Develop parametric simulation analysis via the IES-VE platform, addressing the findings and evaluating the thermal reactions of different construction materials for moderating indoor temperature with the possibility of energy-saving.
- Presenting the interpretable ranking for energy reduction, lower CO_2 emissions, and improved thermal comfort that could be useful for experts in the field.

The remainder of the paper is organized as follows. Section 2 presents the detailed proposed methodology along with the dataset utilized in this study. Section 3 demonstrates the statistical outcomes and performance evaluation of various ML models used to predict energy consumption. Finally, Section 4 concludes the study outcomes and presents future directions.

2. Methods

The proposed approach for the Buildings' energy performances comprises a composite of several data analysis and ML algorithms, including: a) Dataset compilation, consisting of building materials, wall thickness, construction configuration, orientation, and construction shape prepared using IES-VE software b,) pre-processing, to eliminate the noisy samples and to standardize the dataset to the required form c) predictive modeling, to predict the energy consumption in buildings and d) model interpretation particularly, identifying the importance of input factors and performance evaluation of predictive models. Figure 1 shows an overview of the research methodology for this work. Further details about each component are provided in the following sub-sections.

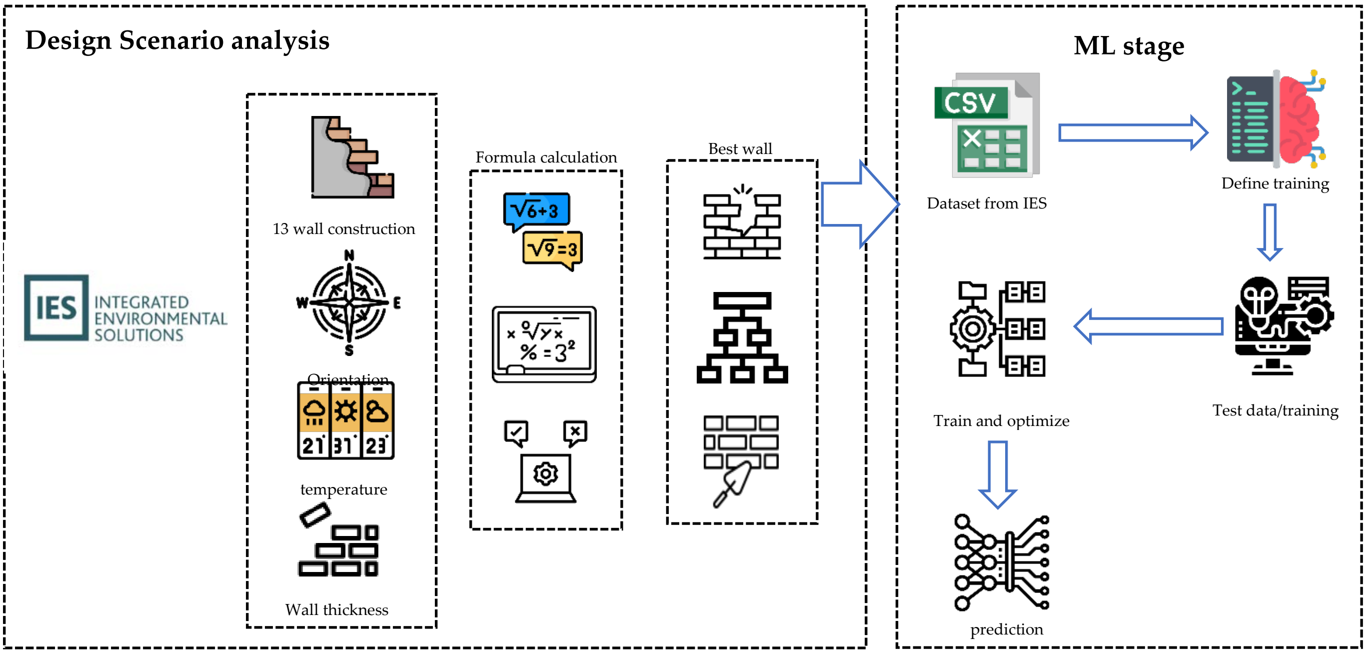


Figure 1 Overview of the research methodology.

As shown in Figure 1, the proposed system consists of 2 stages, first, the associated models are generated, and simulations are performed via IES-VE ApacheSim. The data is stored as a CSV file in the second stage and split into training (70%) and testing (30%) samples using K-fold cross-validation (K=10) (Wong & Yeh, 2019). Extensive simulations are then performed to optimize the model parameters in predicting the energy efficiency of buildings. One hotkey encoding is used to transform the categorical variables of the dataset into numerical form. Once data conversion is completed, data cleansing (e.g., outlier removal, data imputation) and standardization (i.e., normalization) are performed. Outliers are removed using the interquartile range (IQR) (Han, et al., 2021).

2.1. Dataset

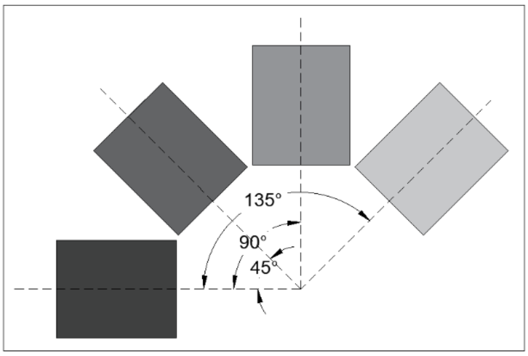
The dataset was simulated using the IES-VE software to measure the thermal performances of the building envelope in the context of energy efficiency improvements. Simulation stages included in Table 1 which shows the different options used in the simulation:
Stage 1: Different Wall Materials
Stage 2: Different Thicknesses of the Wall
Stage 3: Different Orientation of the Space
Stage 4: Different Wall Construction
Stage 5: Different Shape Factors

For example, a wall built with HCB was simulated with 200mm thickness and 0 orientation (North), with wall construction W1 and shape A. then HCB was simulated with 200mm thickness and 0 orientation (North), with wall construction W1 and shape B.

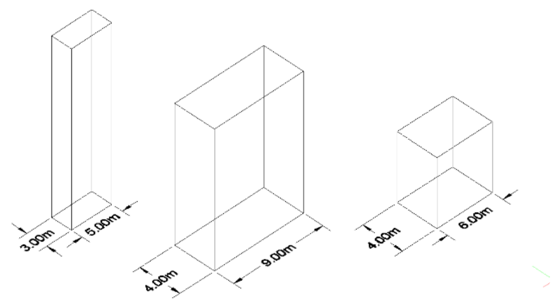
Table 1 The simulation input details used for use IES-VE software

Materials	Thicknesses of the Wall (222-300)	Orientation of the Space (0, 45, 90, 135)				Wall Construction Types (13 types)	Shape Factors (total 3shapes)		
Heavy-weight Concrete Block(HCB)	222	0	45	90	135	1-13 (figure 1)	A	B	C
	300	0	45	90	135	1-13 (figure 1)	A	B	C
Aerated Concrete Block (ACB)	222	0	45	90	135	1-13 (figure 1)	A	B	C
	300	0	45	90	135	1-13 (figure 1)	A	B	C
Common Fried Brick (CFB)	222	0	45	90	135	1-13 (figure 1)	A	B	C
	300	0	45	90	135	1-13 (figure 1)	A	B	C
Unfired Brick (UFB	222	0	45	90	135	1-13 (figure 1)	A	B	C

	300	0	45	90	135	1-13 (figure 1)	A	B	C



Orientation diagram



Building shapes A, B, and C

A total of 1119 data were stimulated according to the different options for each wall. The outcomes related to the samples of the data used are presented in Table 2.

Table 2: a sample of walls details used in the simulation with different Thermal Performance and energy consumption.

Wall type	Wall thickness (mm)	Orientation	Material	Shape	Thermal transmittance (U-Value) m_2K/W	Thermal resistance (R-Value)W/m^2K	Decrement factor/Summer Day	Thermal Mass kj/m2k	Yearly Energy Consumption MWh
W-1	222	N	ACB	A	1.64	0.43	0.33	136.89	11.74
W-2	222	N	ACB	A	0.91	0.53	0.13	8.53	10.72
W-3	222	N	ACB	A	0.96	0.72	0.17	156.80	9.72
W-4	222	N	ACB	A	1.83	0.96	0.52	154.74	12.25
W-5	222	N	ACB	A	1.22	0.83	0.63	202.10	9.52
W-6	222	N	ACB	A	0.93	0.98	0.42	174.92	8.34
W-7	222	N	ACB	A	0.63	1.26	0.53	90.63	8.62
W-8	222	N	ACB	A	0.74	1.96	0.64	264.95	11.71
W-9	222	N	ACB	A	0.61	1.48	0.52	207.92	10.00
W-10	222	N	ACB	A	0.64	1.38	0.17	73.05	10.31
W-11	222	N	ACB	A	0.84	1.31	0.09	72.77	9.26
W-12	222	N	ACB	A	0.71	1.23	0.21	73.05	9.62
W-13	222	N	ACB	A	0.63	0.88	0.12	78.23	9.81

The building envelope walls and energy-use case data in this research study were simulated using the IES-VE software via a range of applications, including ApacheSim, to evaluate heat deviation, cooling and heating, heat transfer, and energy demand. In addition, a variety of brick materials were used to model wall spaces, including Heavyweight Concrete Block (HCB), Aerated Concrete Block (ACB), Common Fried Brick (CFB), and Unfired Brick (UFB). At the same time, wall thickness varies from 222 to 300 mm depending on the materials used and the construction configurations. Furthermore, the spaces were simulated with various construction methods and orientations, including 0 (representing North), 45, 90, and 135 degrees, to investigate the impact of the orientation of the building on its energy demand. Figure 2 shows the details of the 13 walls used in the simulations to determine the thermal efficiency of materials via various wall constructions. The objective of the simulations is to assess the impact of the construction method and wall layers on the thermal performance of buildings. Table 3 shows examples of the different ApacheSim simulations with the associated equations used in this research study.

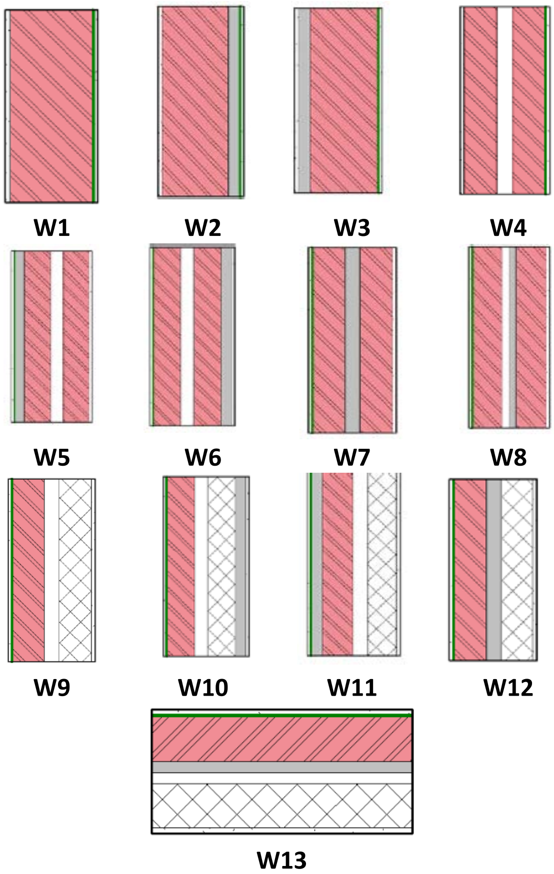


Figure 2. The details of the 13 walls used in the IES-VE A ApacheSim simulations.

Table 3. Examples of the different ApacheSim simulations with their equations

Calculation	Equations	Units	Reasons
thermal diffusivity (m²/s)	$\alpha = \lambda/\rho C\rho$ (Eq. 1)	α is the thermal conductivity (m²/s) λ is thermal conductivity (W/mK) ρ is the density (kg/m³) C_ρ is the specific heat capacity J/(kgK).	To find how quickly the temperature of the material reaches thermal equilibrium.
thermal effusively Ws$^{1/2}$/(m²K)	$\tau = \sqrt{\lambda\rho C_\rho}$ (Eq. 2)		To find the speed at which the surface of the materials gets warm.

Heat differentiation	$\frac{T_{n-1}-2T_n+T_{n+1}}{\delta_n^2} = (\rho C_\rho/\lambda)\partial T/\partial t$ (Eq. 3)	(T_n) is the temperature (°C), at a node n and δ_n is the local node spacing (m).	To find the heat differentiation and to ensure accuracy. Several nodes distributed within the layers ensure accurately modeling the transfer of heat and storage characteristics for the chosen time-step.
temperature time derivative (°C/s)	$T_n^j = (T_n^{j+1} - T_n^j)/\Delta$ (Eq. 4)	(T_n^j) is the temperature (°C) (T_n^j) is the temperature time derivative (K/s) at node n and time-step (j).	To express the temperature time derivative $\partial T / \partial t$ at present time: Explicit methods use a forward-difference scheme, which uses present and future values of the nodal temperature
Energy demand for space (MJ)	$Q_{HN} = Q_{L,H} - (\eta_{G,H} \times Q_{G,H})$ (Eq. 5)	Q_{HN} and Q_{NC} Are the energy demand for heating and cooling of the building zone (MJ), respectively. $Q_{L,H}$ and Q_{LC} are the total heat losses.	To calculate the energy demand for cooling and heating of the buildings
Total heat transfer (MJ)	$Q_L = Q_T + Q_V$ (Eq. 6)	QL is the total heat transfer (MJ), Q_T is the total heat transfer by transmission (MJ), and Q_V is the total heat transfer by ventilation (MJ).	To calculate the total heat transfer

The heat transfer ratio via building structure, whether composite or single, is known as the thermal transmission or U-value, which is normalized by the variations in the temperature across the building structure, depending on the thermal resistance (R-value). The latter depends on the thickness of the materials and their thermal conductivity. The lower the heat loss is, the higher the resistance (Jannat, et al., 2020). Thermal diffusivity relates to the status of heat transfer, indicating how fast the temperature of a material achieves a thermal balance with the adjacent temperature. A higher thermal diffusivity value reveals faster heat transmission through the material (Zhang, et al., 2019). Figure 3 shows the monthly and annual minimum, maximum, and mean air temperatures in the UAE, where the data for this study is set..

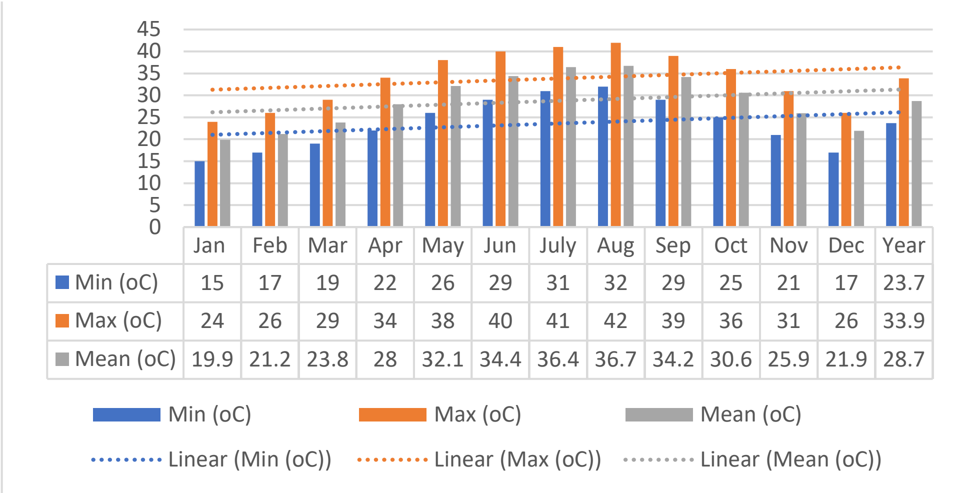


Figure 3. A depiction of the UAE minimum, maximum, and mean air temperatures per month and year (Alqasemi, 2022).

2. 2. Machine leaning methods for the energy consumption prediction

In this section two machine learning models are utilized for the prediction of energy consumption of building envelops using MLR and RF since they represent the state of the art models that showed good performance in various applications.

2.2.1. Multivariate Linear Regression (MLR)

Multivariate Linear Regression (MLR) is a popular regression technique where multiple explanatory variables contribute to the dependent variables. The research study used MLR for the multi-attribute analysis to investigate the inter-relationships of multiple factors, shown in Table 4, and the target variable (i.e., energy consumption). The algorithm calculates a set of coefficients for the independent explanatory variables, which can be written as:

$$Y_i = \alpha + \beta_1 x_i^{(1)} + \beta_1 x_i^{(2)} + \ldots.. + \beta_1 x_i^{(n)} \qquad \textit{(Eq. 7)}$$

where Y_i is the estimated value of ith component of the dependent variable Y (Energy consumption), n is the number of independent variables (n=26, in this study), x_i^j is denoted by the i^{th} component of j^{th} explanatory variable.

The parameter α in Eq. 7 represents the constant term, while the parameter β_i represent the corresponding coefficients for each explanatory variable, which are given in Table 4.

The associated cost function (i.e., model error, also known as residuals) is given by:

$$E(\alpha, \beta_1, \beta_2 \ldots.., \beta_n) = \frac{1}{2m} \sum_{i=1}^{m} (y_i - Y_i) \qquad \text{(Eq. 8)}$$

Where m is the number of training data patterns (m=936 in this study) and y_i is the dependent variable's observed data (i.e., ground truth). Table 4 summarizes the parameters defined in Eq.8 in the context of the proposed model. It can be noticed that β (i.e., slope coefficient for each explanatory variable) varies for different explanatory variables. A higher positive and negative β indicates strong positive and negative relationship with the target variable (Y), respectively. For instance, $\beta = 0.4$ (for thermal mass) in Table 4 shows that the target variable (Y) will increase by 0.4 units if the thermal mass increases by 1 unit, indicating a positive relationship between these variables. Likewise, the energy consumption will decrease by 0.15 units with 1 unit rise in Construction type-WL (indicating a negative relationship with target variable).

Table 4. The intercept and coefficient values for the proposed MLR model for the prediction of yearly energy consumption

Explanatory Variable (x)	β	Standard Error
Construction type_WA	-0.04	0.033
Construction type_WB	-0.06	0.034
Construction type_WC	-0.08	0.033
Construction type_WD	-0.14	0.033
Construction type_WE	-0.08	0.034
Construction type_WF	-0.04	0.033
Construction type_WG	-0.10	0.033
Construction type_WH	-0.08	0.034
Construction type_WI	-0.11	0.033
Construction type_WJ	-0.12	0.034
Construction type_WK	-0.12	0.033
Construction type_WL	-0.15	0.034
Wall thickness	-0.04	0.025
Orientation	0.03	0.025
Material_ACB	-0.003	0.031
Material_CFB	0.022	0.031
Material_HCB	-0.02	0.031
Shape_A	-0.0001	0.028
Shape_B	0.01	0.029
Shape_C	0	0
Thermal transmittance	0.17	0.028
Thermal resistance	0.02	0.027
Decrement factor/ summer day	0.31	0.028
Thermal Mass	0.40	0.027
Constant term (α)	0.008	0.025

Further details and mathematical formulation of MLR can be found in (Maulud & Abdulazeez, 2022). The magnitude of the altered R^2, the relative standard error (RSE) for the regression, and the significance value results of the t-test for the individual explanatory variables can be used as the performance measure of the MLR. Furthermore, visualizations of the diagnostic plots for the residuals versus the fitted values and R^2 are useful to visualize the performance of the regression model.

2.2.2 Random Forest (RF)

Random Forest (RF) can be thought of as a potentially non-linear alternative to MLR. RF is a very popular method used in various classification and regression tasks. The RF algorithm can effectively and efficiently produce high-dimensional feature partitions based on the strategy of divide-and-conquer, where the distribution probability is situated. Additionally, it enables density estimation for random functions, which can be used in classification, clustering, and regression tasks. The main concept is to use multiple weak models (in the form of decision trees), where each model is trained over bootstrapped samples from the dataset. In Bootstrap process, random samples of subsets are generated from a given dataset (for specific number of iterations) for given number of variables. Bootstrap RF aggregation in this context, combines ensemble learning methods leading to reduced variance of the learning model, which is a major challenge in regression tasks. In the aggregation step, the average outcome of each observation across all models is calculated as the final prediction for the corresponding observation (i.e., test data sample). This study uses the Random Subspaces method and bagging to achieve robust regression results (i.e., utilizing the concept of bagging predictors).

It is important to note that identifying the best trade-off between bias and variance (i.e., low bias-low variance) is one of the common challenges with ML models, particularly decision tree. A single model (e.g., a single decision tree) can possibly have either low-bias high-variance, high-bias low-variance, or high-bias high-variance, which is controlled in the present study by utilizing an ensemble modelling approach with bagging. Using the bagging technique, multiple regressors with varying structures (i.e., with different depths, number of trees, different subsets of input attributes and different samples in each bag for each regressor) are created to identify the optimal regression points. Based on the results from multiple regressors, prediction error from each tree is aggregated to reduce the overall prediction error, thus simultaneously reducing the bias and variance in model predictions. In addition, bootstrapping (sampling with replacement) helps perform the splits among different input attributes at varying levels with lowest Entropy/highest information gain which makes the decision trees highly uncorrelated and strong regressors as compared to non-ensemble learning. We further perform model tuning (using randomized search and cross-validation) with varying parameters (e.g., number of trees, no. of variables in a split, tree depth etc.,), to achieve optimal outcomes. Mathematically, the mean squared error (MSE) from a RF regression model can be represented as:

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(f_i - y_i)^2 \qquad (Eq.\ 9)$$

Where N is number of data points, f_i is the model output, and f_y $is\ the\ actual\ value\ of\ data\ point\ \ i.$ Equation 9 calculates the distance of each node from the model predicted actual value to identify which tree provides better decision in the forest. Further technical details, including explanations of the decision tree structures and the feature bagging approach can be found in related works (Khan, et al., 2021; Breiman, 2001).

3. Results and Discussions

Experiments were conducted using the dataset described in Section 2.1, and multiple predictive ML models to predict the energy consumption as a target variable. For evaluation and comparison purposes, we used R^2 and the relative standard error (RSE) criteria for MLR, while the variance in the proportion and RSE criteria were chosen for RF, as suggested in similar works (Oukawa, et al., 2022). Furthermore, we used 10-fold cross-validation method to evaluate the model generalization and reliability on the unseen samples. The first set of experiments utilized MLR to model the energy consumption, as shown in Table 1. The RSE values ranged from 0001 to 0.4. It can be noticed that wall thickness, orientation, and thermal mass indicated lower RSE (<0.001); however, not all of them were statistically significant (p>0.05). For instance, the RSE for the orientation is 0.001, whereas the statistical significance test did not indicate a strong relationship with the target attribute (p=0.23). In contrast, the feature of DF summer day showed a relatively higher RSE (0.23). However, it indicated statistical significance for the relationship with the output (p = 2e-16). While the overall model indicated statistical significance (p = 2e-16), the overall MLR model produced an R^2 value of 0.42, indicating a weak relationship between explanatory variables and target attributes. We used 25% of the data to predict using the MLR model, which resulted to a mean squared error (MSE) of 0.6. The statistical outcomes of the MLR model in predicting energy consumption are presented in Table 5.

Table 5. Overview of the MLR model's statistical outcomes in predicting energy consumption.

Variable	RSE	Significance (p-value)
Overall	0.70	0.73
construction.type_WA	0.41	0.52
construction.type_WB	0.40	0.26
construction.type_WC	0.40	0.0.05
construction.type_WD	0.41	0.0003
construction.type_WE	0.40	0.23
construction.type_WF	0.41	0.45
construction.type_WG	0.41	0.004
construction.type_WH	0.41	0.001
construction.type_WI	0.41	0.002
construction.type_WJ	0.40	0.0001
construction.type_WK	0.41	0.001
construction.type_WL	0.40	0.0001
Wall thickness	0.002	0.24
Orientation	0.001	0.23
Material_ACB	0.22	0.88
Material_CFB	0.22	0.53
Material_HCB	0.22	0.32

Shape_A	0.19	0.75
Shape_B	0.19	0.98
Thermal transmittance	0.13	8.27e-08
Thermal resistance	0.16	0.28
Decrement factor/summer day	0.23	2e-16
Thermal Mass	0.001	2e-16

As we used one-hot encoding for several explanatory variables (e.g., construction type), we further performed a correlation analysis to visualize the inter-relationship between the encoded variables. Figure 4 shows the correlation plot for the entire dataset. It can be noticed that there is no inter-relationship between the sub-types of construction wall type, while there was a weak negative relationship between material sub-types (i.e., the encoded variables). On the other hand, the thermal mass, U-value, and DF variables indicated a weak positive inter-relationship with the target variable. This also aligned with the outcomes from the MLR model (in Table 5), which indicated the statistical significance of these variables.

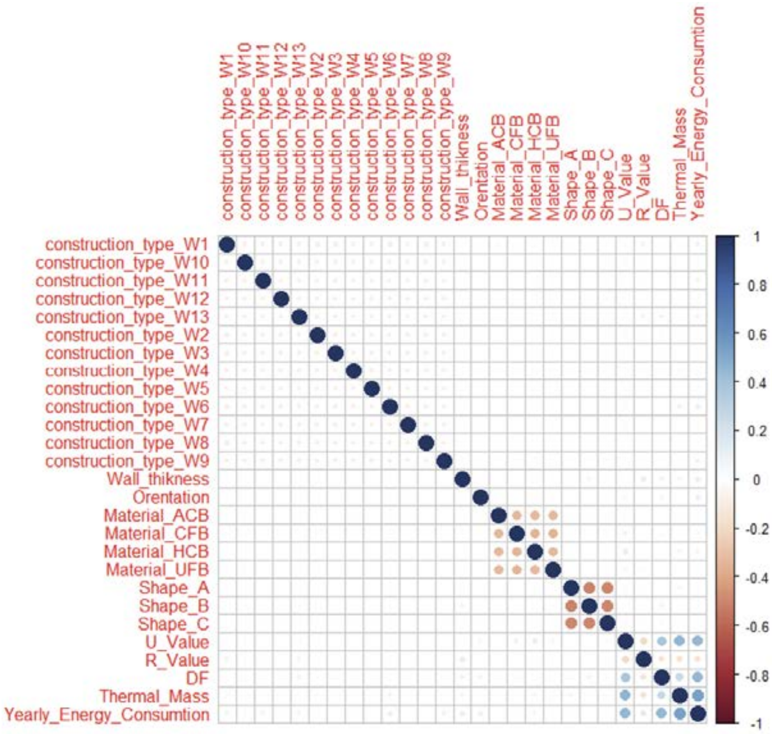
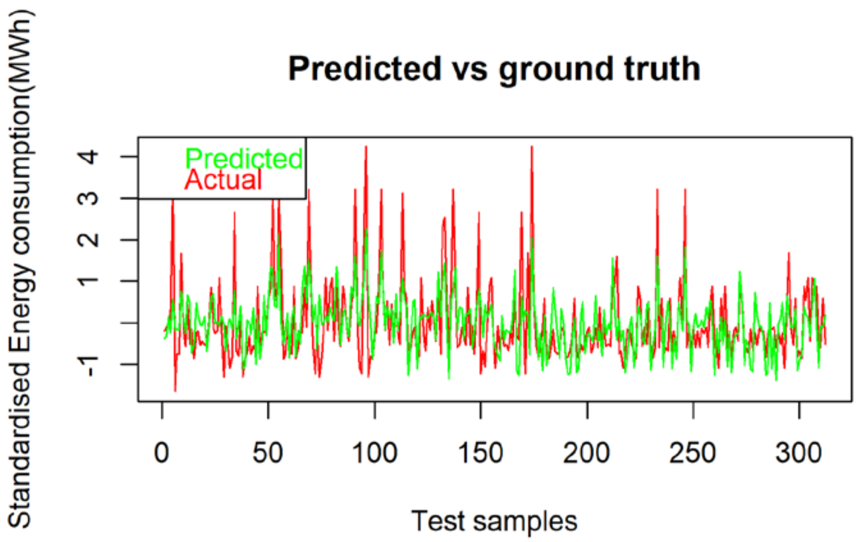
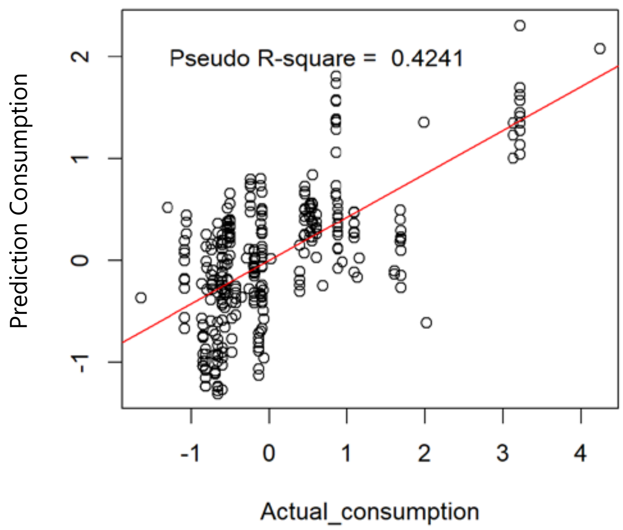


Figure 4. Correlation analysis between explanatory and target variables.

Figure 5(a) shows the prediction performance of the MLR model for the unseen proportion of the dataset (30% of the dataset). It can be noticed that the overlapping between the ground truth (red line) and the predicted values (green line) has substantial deviations, which shows the poor prediction of the MLR on the unseen data. By further observing the performance of the MLR model, it can be concluded that it followed the trend of the data. However, it cannot precisely track the extreme energy consumption values, possibly due to its linear nature. Moreover, we employed the pseudo-r-square tool (Di Franco & Santurro, 2021), a well-known statistical measure for the goodness of model fit. Figure 5(b) shows that the predicted values are dispersed from the actual values (ground truth), producing an R^2 value of 0.42, which indicates poor MLR model prediction on the testing set.



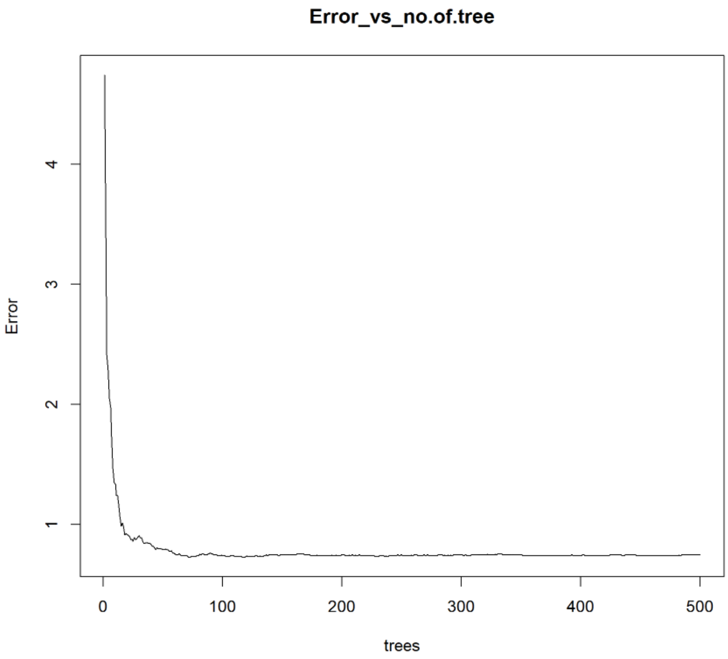
(a)



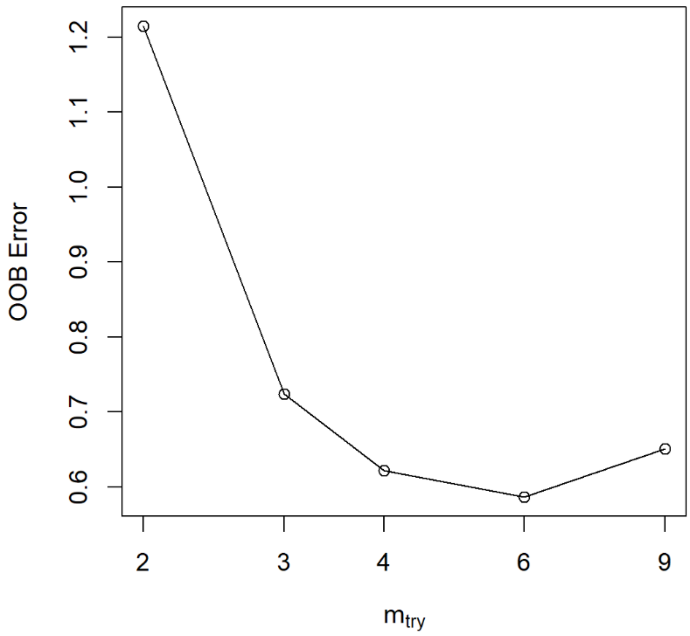
(b)

Figure 5. MLR performance on the testing set. (a) Plot of predicted vs. actual energy consumption. (b) R-square fitness test.

Alternatively, we evaluated the predictive performance of Random Forest approach (as described in Section 2.3). Figure 6(a) shows the configuration of the number of trees used in RF, indicating a stable error value for 100 trees (approximately). Likewise, Figure 6(b) shows the optimal selection of several variables randomly sampled as candidates at each split (mTry = 6), producing a minimum out-of-bag error (OOB error).



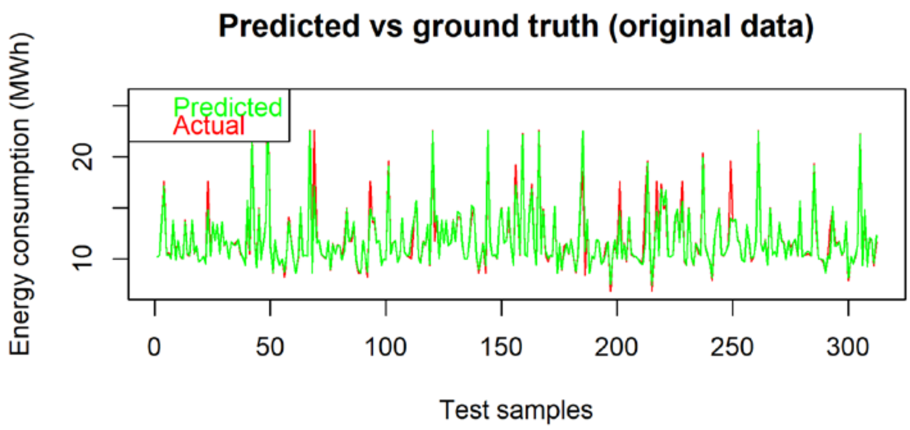
(a)



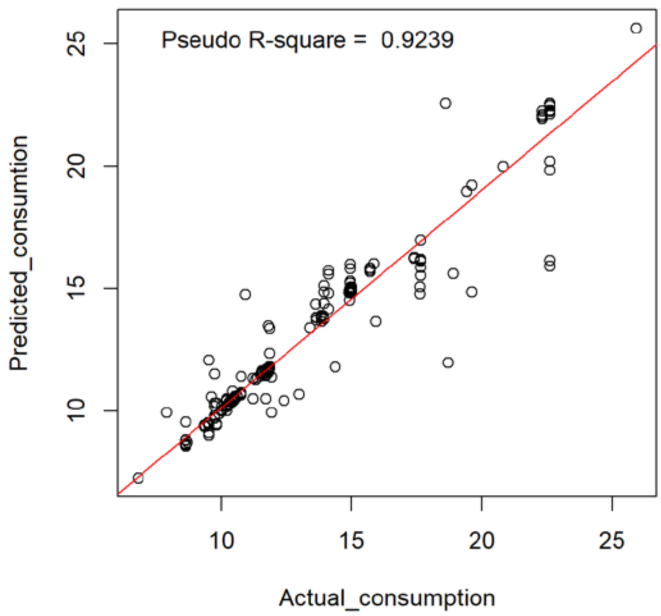
(b)

Figure 6. Random forest performance. (a) Plot of the mean squared versus the number of trees used in RF, (b) Plot of the Out-of-Bag error vs candidates at each split.

One of the main benefits of tree-based ML models is their ability to deal with various data types, including categorical data. Thus, the RF model was trained over the original dataset, i.e., without using any data conversion (one hotkey encoding). Similar to MLR model, the RF model was trained on 70% of the data and evaluated on the remaining 30% of unseen instances. Figure 7(a) shows the RF model's prediction performance for the dataset's unseen proportion. It can be noticed that there is a high degree of overlap between the ground truth (red line) and the predicted values (green line), which demonstrated the superior prediction performance of the RF model on the unseen data. Statistical analysis of outcomes indicated a 92.5% variance, explained by the model, resulting in an MSE of 0.07 in the training set. Figure 7(b) shows the statistical measure of goodness of fit, indicating the variation of energy consumption, explained by the explanatory variables in the RF regression model. It can be noticed that for the testing set, the model produced a pseudo $R^2 = 0.92$, which indicated the robustness of the RF as compared to MLR (with $R^2 = 0.42$).



(a)



(b)

Figure 7. RF performance on the testing set using the original data representation. (a) The plot of predicted vs. actual energy consumption. (b) R-square fitness test.

Figure 8 shows the importance of variables in the prediction task, which is one of the powerful features of the RF model, i.e., explainability. It should be noted that the Y-axis represents the variable's name, while the x-axis represents the increase of %IncMSE in prediction.

The outcomes showed that the parameters of DF summer day, R-value, thermal mass, and U-value are significant contributors to the regression model. Interestingly, this aligned with the MLR outcomes of statistical significance (with p-values<0.05) for most cases (e.g., thermal mass, DF, thermal transmission) except R-value (i.e., thermal resistance), which was identified as significant in the RF model but not in the MLR model. The R-Value is related to the wall thickness and measures heat flow resistance through different construction materials. Therefore, the greater the R-value, the better the material's thermal resistance. Thus, this is considered a significant contributor to the building wall envelopes.

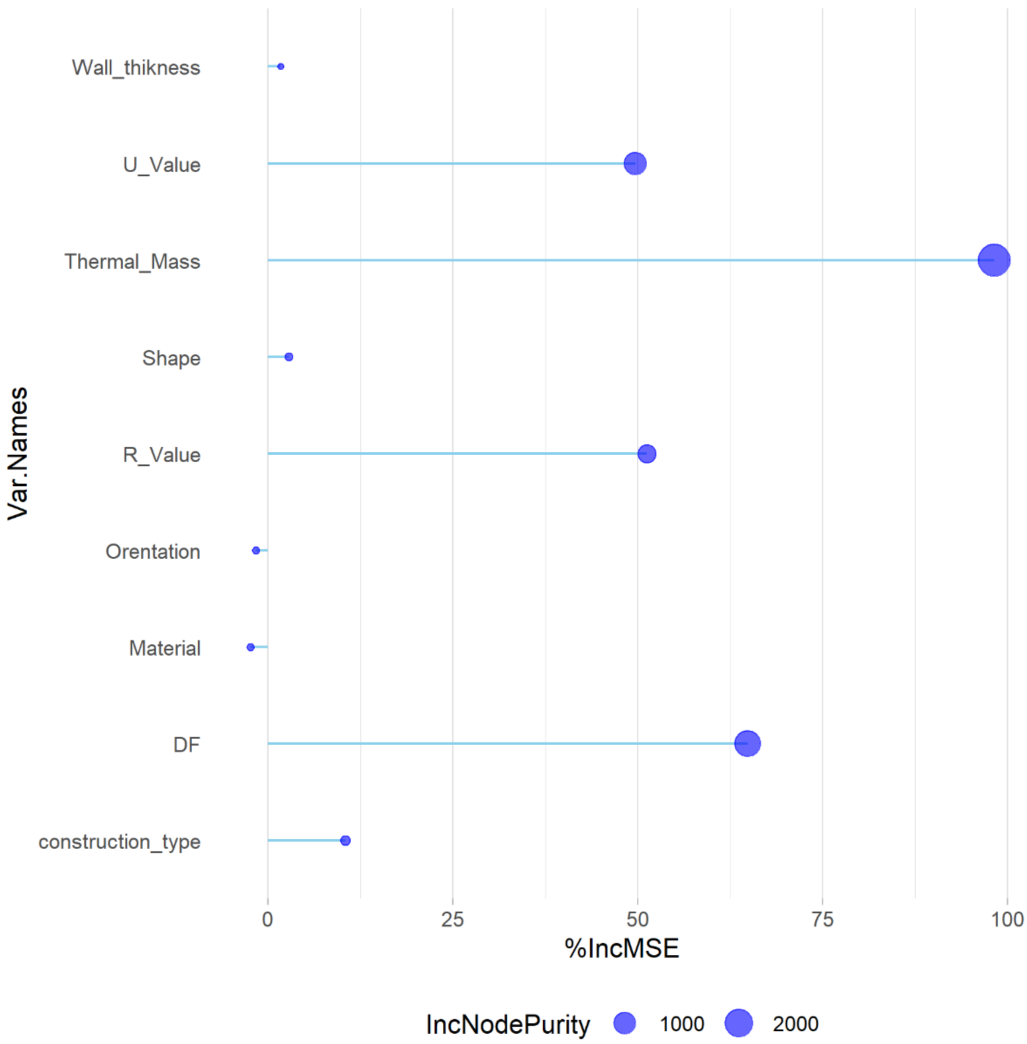
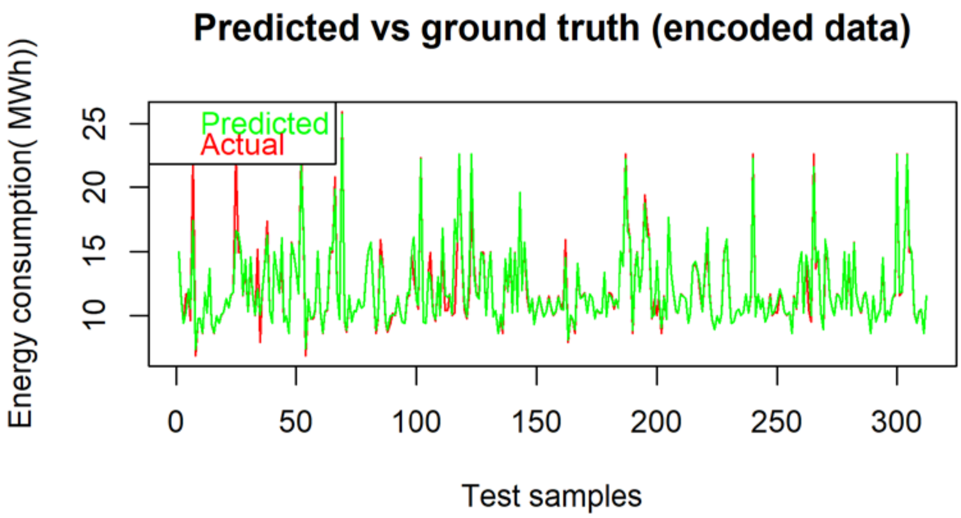
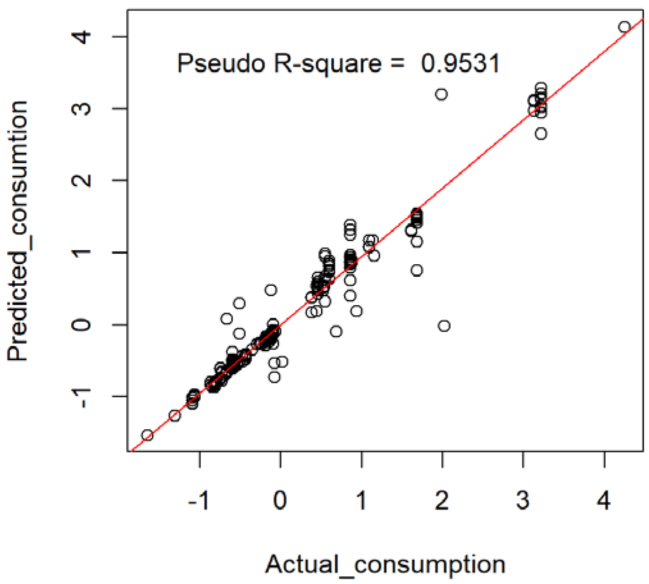


Figure 8. Attribute importance to the original data representation (without encoding) using the RF model.

While the RF model produced promising regression outcomes, it is essential to note that the original data representation was used (i.e., without the use of one-hot key encoding). For fairness of comparison with the MLR model, further experiments were conducted using the transformed dataset, i.e., the data configuration used in the MLR model. Figure 8 shows the prediction performance of the RF model in the case of unseen samples, indicating a perfect overlap between ground truth and predictions. It can also be noticed that the R^2 score has also significantly improved to 0.95, indicating the optimal fitness of the regression model.



(a)



(b)

Figure 9. RF performance on the testing set using the transformed data representation. (a) Plot of predicted vs actual energy consumption. (b) R-square fitness test.

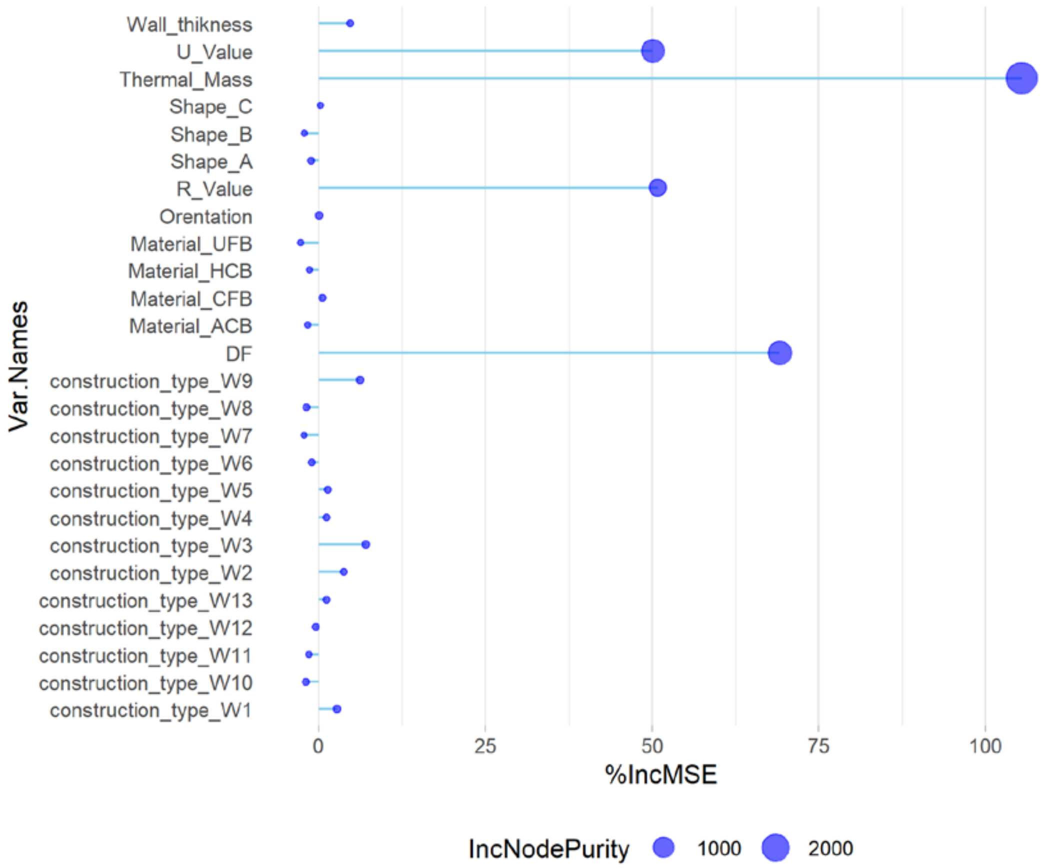


Figure 10. Using the RF model, attribute importance for the transformed data representation (with one-hot key encoding).

Figure 10 shows the variable importance of the encoded data retrieved from the RF model. Similarly to Figure 9 (a, and b), which presented the variable importance identified by the RF model on the original data representation, the variables of DF summer day, R2 value, thermal mass, and U-value were significant contributors to the regression model. In addition, we can easily identify the attribute significance for the discrete level factors (e.g., construction types W1-W13). For instance, construction type W1 contributes more towards the model prediction than W2, W6, and W11.

Table 6. Performance comparison and findings between the use of the RF model on the original vs. the one-hot key encoded data.

Parameter	Original data	Encoded data
Type of RF	Regression	Regression
mTry	6	19
No. of trees	100	100
Var explained	92.5	93.15
MSE	0.073	0.06
Attribute importance	Df summer day, R-value, thermal mass, and U-value	Df summer day, R-value, thermal mass, and U-value
Thermal resistance	0.68	0.95
R^2	0.92	0.95

Finally, Table 6 compares the statistical outcomes from the RF model for both scenarios (original data vs. encoded data). It can be noticed that the RF performed better for the encoded data, indicating a slight reduction in MSE (from 0.073 to

0.06) and an increase in R^2 (0.92 to 0.95). Similar outcomes are shown from the test of significance, where the R-value increased from 0.68 (for the original data) to 0.95 (for the encoded data), indicating higher similarity between ground truth samples and model prediction. The attribute importance remained unchanged in both scenarios; however, encoded data was useful in identifying attribute importance at a discrete level (i.e., for sub-types of attributes, such as construction wall types).

The research study by (Bhamare, et al., 2021), implemented five different ML-based models (XGBR, RFR, GBR, ETR, and CatBoost Regression) to predict the PCM integrated roof envelopes. The data set for the study involved melting enthalpy, density, specific heat, thermal conductivity, and melting temperature for the direct building thermal performance calculations via 500 data generated via the MATLAB algorithm. The results showed that the GBR model performed best among the other 4 ML models. It showed the root mean squared error values and a least mean absolute error of 0.3741 and 0.23, respectively, compared with the other ML models. They demonstrated an R^2 value of 97.92%. CatBoost Regression and RFR have indicated decent accuracy prediction with R-Square values of 97.54% and 97.70%, respectively. The ETR model showed a root mean absolute error and root mean squared error values of 0.34 and 0.36, respectively, and a minimum R^2 value of 94.56%.

Table 2. Benchmark of various ML learning algorithms for energy consumption

Reference	ML	Utilization	Results
(Bhamare, et al., 2021)	XGBR, RFR, GBR, ETR, and CatBoost Regression	PCM integrated roof envelopes	least mean absolute 0.3741
(Quevedo, et al., 2023)	SVM, ANN, MLR	Energy benchmark for a university building in Brazil	The SVM method had the lowest mean absolute error, root mean absolute error, and the highest R2 value
(Ngo, et al., 2022)	SVR, GWO, RF, M5P, REPTree	Energy use forecast for a commercial building in Vietnam	The average root-mean-square error (RMSE) of the WIO-SVR was 2.02 kWh which was more accurate than those of the SVR model with 10.95 kWh, the RF model with 16.27 kWh, the M5P model with 17.73 kWh, and the REPTree model with 26.44 kWh.
(Mohammadiziazi & Bilec, 2022)	RF, MLR, XGBoost, single regression tree	climate change analysis in USA	RF model provided better performance and reduced the mean absolute error by 4%, 11%, and 12% compared to XGBoost, single regression tree, and MLR, respectively.
(Dong, et al., 2021)	ANN, SVR, MLR	Hourly energy consumption prediction of an office building	*The Method* not only improved the prediction accuracy of SVR, ANN and stacking model but also maintaining the stability of SVR algorithm. The limitation of this strategy is that energy consumption patterns are based on the results of an analysis of particular building usage over some time. When the research subject changes or a new running pattern appears, it cannot be updated in time
Our proposed technique	RF, and MLR	Buildings' Envelope Wall Materials	RF performed better for the encoded data, indicating a slight reduction in MSE (from 0.073 to

			0.06) and an increase in R2 (0.92 to 0.95).

As it can be shown in Table 7, various ML models have been utilized in the context of building energy consumption prediction in which the researcher works focused on integrated roof envelopes, energy forecast for commercial buildings, climate change analysis, hourly energy consumption prediction, which were specific and narrowed down experiments that did not take extensive considerations to the surrounding environments, various construction materials, wall thickness, orientation, wall configuration, and construction shapes at the same time. This is different from our proposed approach, which utilized RF and looked extensively at all these parameters that affect energy consumption in buildings. In addition, all ML techniques provided in Table 7 looked at existing building conditions to benchmark the energy usage, forecast energy consumption and climate change analysis while our proposed approach addressed the issue related to building envelop for new constructions to assist professional from construction industry to select the optimum wall thickness, orientation, construction methods and shape in advance and before the construction of the building. This is significant since it provides an improved carbon footprint for new buildings.

4. Conclusion

This research aimed to develop a thermal performance prediction model for wall envelopes in buildings in the UAE via ML methods. RF was utilized in this work to predict energy consumption, in which our extensive simulation results indicate that RF outperformed conventional approaches, including MLR. The proposed models were trained and validated using a dataset simulated in the IES-VE to assess their thermal performances as potential energy efficiency improvements. The RF model was evaluated using one-hot key encoding, in which the model demonstrated superior performance. In addition, the proposed RF-based energy consumption prediction is useful for interpreting and visualizing the attributes' ranking as a feature extraction model. This might be useful for building life cycle assessments as they can lead to more efficient and effective utilization of energy resources. The findings can help to reduce energy consumption, resulting in cost savings and lower carbon emissions.

The findings of our work contribute and provide valuable inputs as follows.

- A ML framework is proposed as a powerful prediction methodology to improve energy performance in the construction industry and enhance the energy efficiency of buildings.
- RF can be suggested as an efficient method for explanatory energy consumption in the building sector.
- Prediction results were evaluated with statistical tests. These were identified as significant in the case of the RF model but not for the MLR model. The prediction performance of the RF model over unseen samples demonstrates a perfect overlap between ground truth and predictions. It can be noticed that RF generates R^2 value equal to 0.95, indicating the optimal fitness of the regression model. In the case of the MLR, the results showed an R^2 value of 0.42, which indicates poor prediction performance on the testing set.

Although ML models have been implemented widely within the construction sector, various gaps and limitations were discovered. For example, some studies are yet in the research and development stage, while the sector has approved very few for use in real-world settings. This could be due to deficiencies in the training data. In addition, if implemented, the cost benefits of ML cannot be accurately estimated because a potential model working well for number of buildings may not be appropriate for others.

Future works will involve using larger and more diverse data sets to train various ML models, including but not limited to recurrent neural networks, long short-term memory (LSTM), and adaptive models.

Author Contributions: "Conceptualization, Aseel Hussien and Abir Hussain Methodology, Wasiq Khan; software, Wasiq Khan, Aseel Hussien validation, Wasiq Khan, Panos Liatsis and Dhiya Al-Jumeily.; formal analysis, Ahmed Al-Shamma'a.; investigation, Aseel Hussien resources, Aseel Hussien data curation, Aseel Hussien .; writing—original draft preparation, Aseel Hussien, Panos Liatsis, Wasiq Khan writing—review and editing, Abir Hussain, Panos Liatsis, Dhiya Al-Jumeily visualization, Ahmed Al-Shamma'a supervision, Abir Hussain, Aseel Hussien .; project administration, Dhiya Al-Jumeily funding acquisition, Aseel Hussien.

Conflicts of Interest: "The authors declare no conflict of interest."

Abbreviations

A list of the abbreviations used in this paper is given as follows.

IES-VE	Integrated Environmental Solutions- Virtual environment
CSV	Excel Sheets
ML	Machine Learning

R^2	Regression correlation Coefficient
AEC	Architecture Engineering and construction
CO_2	Carbon dioxide
PMV	Predicted Mean Vote
AI	artificial intelligence
GHG	Green House Gag
HVAC	Heating, ventilation and airconditioning
BIM	Building Information models
RF	Random Forest
MLR	Multivariate Linear Regression
HCB)	Heavy-weight Concrete Block(
ACB	Aerated Concrete Block
CFB	Common Fried Brick
UFB	Unfired Brick
U-value	the rate of transfer of heat through a structure
R-value	the ability of insulation material to resist heat flow
R-Square	the coefficient of determination
RSE	the relative standard error
MSE	mean squared error
OOB error	out-of-bag error
XGBR	Extreme Gradient Boosting
RFR	Recursive Feature Elimination
GBR	Gradient boosting
SVM	support vector machine
ANN	Artificial neural network
GWO	grey wolf optimization

References

Aditya, L., Mahlia, T.M.I., Rismanchi, B., Ng, H.M., Hasan, M.H., Metselaar, H.S.C., Muraza, O. and Aditiya, H.B., 2017. A review on insulation materials for energy conservation in buildings. *Renewable and sustainable energy reviews*, *73*, pp.1352-1365.

Akinosho, T.D., Oyedele, L.O., Bilal, M., Ajayi, A.O., Delgado, M.D., Akinade, O.O. and Ahmed, A.A., 2020. Deep learning in the construction industry: A review of present status and future innovations. *Journal of Building Engineering*, *32*, p.101827.

Alghamdi, S., Tang, W., Kanjanabootra, S. and Alterman, D., 2022. Effect of architectural building design parameters on thermal comfort and energy consumption in higher education buildings. *Buildings*, *12*(3), p.329.

Al-Homoud, M., 2001. Computer-aided building energy analysis techniques. Building and Environment, 36(4), pp.421-433.

Alqasemi, A. 2022. Retrieval of monthly maximum and minimum air temperature using MODIS aqua land surface temperature data over the United Arab Emirates.. Geocarto International, 37(10), pp.2996-3013.

Bhamare, D.K., Saikia, P., Rathod, M.K., Rakshit, D. and Banerjee, J., 2021. A machine learning and deep learning based approach to predict the thermal performance of phase change material integrated building envelope. *Building and Environment*, *199*, p.107927.

Bhamare, D.K., Saikia, P., Rathod, M.K., Rakshit, D. and Banerjee, J., 2021. A machine learning and deep learning based approach to predict the thermal performance of phase change material integrated building envelope. *Building and Environment*, *199*, p.107927.

Breiman, L., 2001. " Random Forests Machine Learning" vol. 45 pp. 5-32.

De Dear, R.J. and Brager, G.S., 2002. Thermal comfort in naturally ventilated buildings: revisions to ASHRAE Standard 55. *Energy and buildings*, *34*(6), pp.549-561.

Deshko, V., Buyak, N., Bilous, I. and Voloshchuk, V., 2020. Reference state and exergy based dynamics analysis of energy performance of the “heat source-human-building envelope” system. *Energy*, *200*, p.117534.

Di Franco, G. and Santurro, M., 2021. Machine learning, artificial neural networks and social research. *Quality & quantity*, *55*(3), pp.1007-1025.

Dong, Z., Liu, J., Liu, B., Li, K. and Li, X., 2021. Hourly energy consumption prediction of an office building based on ensemble learning and energy consumption pattern classification. *Energy and Buildings*, *241*, p.110929.

Dong, Z., Yun, Y., Sun, Y. and Wang, F., 2022. Building use-wellbeing associations: A spatial perspective. *Cities*, *121*, p.103493.

Gangu, S.C., Bandi, M.N., Viswanadham, S., Sivaji, C.C. and Kiran, T.S., 2022. Edibility detection of mushroom using Logistic Regression and PCA. *International Journal of Advanced Research in Computer Science*, *13*(3).

Guo, J., Jiang, Y., Wang, Y. and Zou, B., 2020. Thermal storage and thermal management properties of a novel ventilated mortar block integrated with phase change material for floor heating: an experimental study. *Energy conversion and management*, *205*, p.112288.

Guo, J., Jiang, Y., Wang, Y. and Zou, B., 2020. Thermal storage and thermal management properties of a novel ventilated mortar block integrated with phase change material for floor heating: an experimental study. *Energy conversion and management*, *205*, p.112288.

Han, C., Zhang, W. and Ma, T., 2022. Data cleaning framework for highway asphalt pavement inspection data based on artificial neural networks. *International Journal of Pavement Engineering*, *23*(14), pp.5198-5210.

Hussien, A., Saleem, A.A., Mushtaha, E., Jannat, N., Al-Shammaa, A., Ali, S.B., Assi, S. and Al-Jumeily, D., 2023. A statistical analysis of life cycle assessment for buildings and buildings’ refurbishment research. *Ain Shams Engineering Journal*, p.102143.

Hwang, R.L. and Chen, W.A., 2022. Creating glazed facades performance map based on energy and thermal comfort perspective for office building design strategies in Asian hot-humid climate zone. *Applied Energy*, *311*, p.118689.

Jannat, N., Hussien, A., Abdullah, B. and Cotgrave, A., 2020. A comparative simulation study of the thermal performances of the building envelope wall materials in the tropics. *Sustainability*, *12*(12), p.4892.

Jannat, N., Hussien, A., Abdullah, B. and Cotgrave, A., 2020. Application of agro and non-agro waste materials for unfired earth blocks construction: A review. *Construction and Building Materials*, *254*, p.119346.

Johra, H., Heiselberg, P. and Le Dréau, J., 2019. Influence of envelope, structural thermal mass and indoor content on the building heating energy flexibility. *Energy and Buildings*, *183*, pp.325-339.

Khan, W., Crockett, K., O'Shea, J., Hussain, A. and Khan, B.M., 2021. Deception in the eyes of deceiver: A computer vision and machine learning based automated deception detection. *Expert Systems with Applications*, *169*, p.114341.

Maulud, D. and Abdulazeez, A.M., 2020. A review on linear regression comprehensive in machine learning. *Journal of Applied Science and Technology Trends*, *1*(4), pp.140-147.

Mechouet, A., Oualim, E.M. and Mouhib, T., 2021. Effect of mechanical ventilation on the improvement of the thermal performance of PCM-incorporated double external walls: A numerical investigation under different climatic conditions in Morocco. *Journal of Energy Storage*, *38*, p.102495.

Merabet, G.H., Essaaidi, M., Haddou, M.B., Qolomany, B., Qadir, J., Anan, M., Al-Fuqaha, A., Abid, M.R. and Benhaddou, D., 2021. Intelligent building control systems for thermal comfort and energy-efficiency: A systematic review of artificial intelligence-assisted techniques. *Renewable and Sustainable Energy Reviews*, *144*, p.110969.

Mirrahimi, S., Mohamed, M.F., Haw, L.C., Ibrahim, N.L.N., Yusoff, W.F.M. and Aflaki, A., 2016. The effect of building envelope on the thermal comfort and energy saving for high-rise buildings in hot–humid climate. *Renewable and Sustainable Energy Reviews*, *53*, pp.1508-1519.

Xu, L., Tong, S., He, W., Zhu, W., Mei, S., Cao, K. and Yuan, C., 2022. Better understanding on impact of microclimate information on building energy modelling performance for urban resilience. *Sustainable Cities and Society*, *80*, p.103775.

Mousavi, M., TohidiFar, A. and Alvanchi, A., 2022. BIM and machine learning in seismic damage prediction for non-structural exterior infill walls. *Automation in Construction*, *139*, p.104288.

Nematchoua, M.K., Tchinda, R. and Orosa, J.A., 2014. Thermal comfort and energy consumption in modern versus traditional buildings in Cameroon: A questionnaire-based statistical study. *Applied Energy*, *114*, pp.687-699.

Ngo, N.T., Truong, T.T.H., Truong, N.S., Pham, A.D., Huynh, N.T., Pham, T.M. and Pham, V.H.S., 2022. Proposing a hybrid metaheuristic optimization algorithm and machine learning model for energy use forecast in non-residential buildings. *Scientific Reports*, *12*(1), p.1065.

Olu-Ajayi, R., Alaka, H., Sulaimon, I., Sunmola, F. and Ajayi, S., 2022. Building energy consumption prediction for residential buildings using deep learning and other machine learning techniques. *Journal of Building Engineering*, *45*, p.103406.

Oukawa, G.Y., Krecl, P. and Targino, A.C., 2022. Fine-scale modeling of the urban heat island: A comparison of multiple linear regression and random forest approaches. *Science of the total environment*, *815*, p.152836.

Quevedo, T.C., Geraldi, M.S. and Melo, A.P., 2023. Applying machine learning to develop energy benchmarking for university buildings in Brazil. *Journal of Building Engineering*, *63*, p.105468.

Seyedzadeh, S., Rahimian, F.P., Rastogi, P. and Glesk, I., 2019. Tuning machine learning models for prediction of building energy loads. *Sustainable Cities and Society*, *47*, p.101484.

Shoar, S., Chileshe, N. and Edwards, J.D., 2022. Machine learning-aided engineering services' cost overruns prediction in high-rise residential building projects: Application of random forest regression. *Journal of Building Engineering*, *50*, p.104102.

Sudhakar, K., Winderl, M. and Priya, S.S., 2019. Net-zero building designs in hot and humid climates: A state-of-art. *Case Studies in Thermal Engineering*, *13*, p.100400.

Van, T.N. and Quoc, T.N., 2021. Research trends on machine learning in construction management: a scientometric analysis. *Journal of Applied Science and Technology Trends*, *2*(03), pp.96-104.

Veiga, R.K., Veloso, A.C., Melo, A.P. and Lamberts, R., 2021. Application of machine learning to estimate building energy use intensities. *Energy and Buildings*, *249*, p.111219.

Wong, T.T. and Yeh, P.Y., 2019. Reliable accuracy estimates from k-fold cross validation. *IEEE Transactions on Knowledge and Data Engineering*, *32*(8), pp.1586-1594.

Wu, J., Li, X., Lin, Y., Yan, Y. and Tu, J., 2020. A PMV-based HVAC control strategy for office rooms subjected to solar radiation. *Building and Environment*, *177*, p.106863.

Zhang, H., Cheng, X., Li, Y. and Du, X., 2022. Prediction of failure modes, strength, and deformation capacity of RC shear walls through machine learning. *Journal of Building Engineering*, *50*, p.104145.