Applied Chemical Engineering (2025) Volume 8 Issue 1 ISSN: 2578-2010 (O)
doi: 10.59429/ace.v8i1.5593

ORIGINAL RESEARCH ARTICLE

Comparative study of hybrid machine learning models to predict the energy consumption of buildings enhanced with PCM: A case study in Morocco

Kaoutar JRAIDA[1*], Amina MOURID[2,3], Youness EL MGHOUCHI[1], Chadia HAIDAR[1], Mustapha EL ALAMI[3]

[1] *Department of Energetics, ENSAM, Moulay Ismail University, Meknes, 15290, Morocco*

[2] *Laboratory of Advanced Materials Studies and Applications, FS-EST, Moulay Ismail University, Meknes, 3103, Morocco*

[3] *LPMMAT, Physics Department, Faculty of Sciences Ain Chock, Hassan II University, Casablanca, 5366, Morocco*

***Corresponding author:** Kaoutar JRAIDA, kaoutar.jraida@gmail.com

ABSTRACT

Predicting the thermal performance of buildings is a key research area in the context of improving energy efficiency and reducing environmental impacts. Several approaches have been developed to model and predict thermal performance. Among these approaches, machine learning techniques are distinguished by their ability to exploit large amounts of data and model complex systems, but their effectiveness remains to be demonstrated in different contexts. This work therefore explores the application of hybrid machine learning models. Six different models, including ANN-LR, ANN-RR, ANN-RF, ANN-GB, ANN-DT, and ANN-ELM were evaluated and compared to the standalone model (ANN) based on the statistical metrics. Using the TRNSYS tool, the dynamic simulation of a building enhanced with PCM into the roof was performed to generate the data. The findings proved the effectiveness of the hybrid machine learning techniques, with ANN-LR and ANN-GB emerging as the most reliable hybrid approaches for accurate prediction, showcasing their robustness and suitability for complex prediction tasks, while ANN-RR model proved to be the least effective. Furthermore, the performance of the models varied considerably depending on the target, with total energy consumption appearing more complex and challenging for prediction.

Keywords: buildings; hybrid machine learning; energy performance; phase change material; energy simulation

ARTICLE INFO

Received: 7 January 2025
Accepted: 18 February 2025
Available online: 21 February 2025

1. Introduction

Energy management in buildings represents a major challenge in the current energy transition context[1]. Buildings consume a significant share of the world's energy[2], and their consumption is due to a complex interaction between several architectural, technological, and behavioral factors[3–5]. According to the Global Status Report for Buildings and Construction, the global buildings sector represented about 30% of the final energy demand in 2022[6]. Specifically, operational energy demand, including space heating, cooling, water heating, lighting, and other uses, was around 132 EJ. Additionally, an increase in the segment of electricity energy use of 5% is observed between 2010 and 2022[6]. Within the last few decades, numerous strategies, primarily centered on enhancing energy efficiency and integrating innovative technologies, have been developed in order to mitigate the energy consumption of buildings[7–11]. Phase change materials (PCMs) are one of the thermal management technologies used today in building envelopes[12]. These materials are able to change their physical status as a result of the changes in the external temperature, making it possible to reduce gaps between peak and off-peak thermal loads, and optimize interior thermal comfort in buildings[13].

Due to several interconnected factors, predicting the energy consumption of buildings enhanced with PCMs remains a complex challenge[14]. The use of machine learning techniques can offer an effective solution to model and predict these complex dynamics[15]. There is a large number of publications on machine learning. However, there are only a few studies in literature dedicated to the prediction of thermal performance of PCM-Enhanced buildings. The bibliometric analysis illustrated in **Figure 1** highlights the main concepts related to the prediction of building energy consumption, including machine learning, phase change materials, and energy efficiency. Although PCMs and machine learning techniques are emerging and well-explored topics individually, the connections between these concepts and key notions such as building envelopes and energy consumption remain limited. This observation reveals a notable gap in the literature regarding the application of machine learning models to predict the energy consumption of buildings incorporating PCMs.

This work therefore aims to address this research gap by proposing an approach that exploits the capabilities of machine learning algorithms to efficiently model and predict the energy consumption of buildings enhanced with PCM. Using the TRNSYS tool, the dynamic simulation of a building enhanced with PCM was performed to generate the data. Then, the prediction performance of six different models, including ANN-LR, ANN-RR, ANN-RF, ANN-GB, ANN-DT, and ANN-ELM were evaluated and compared to the standalone model.

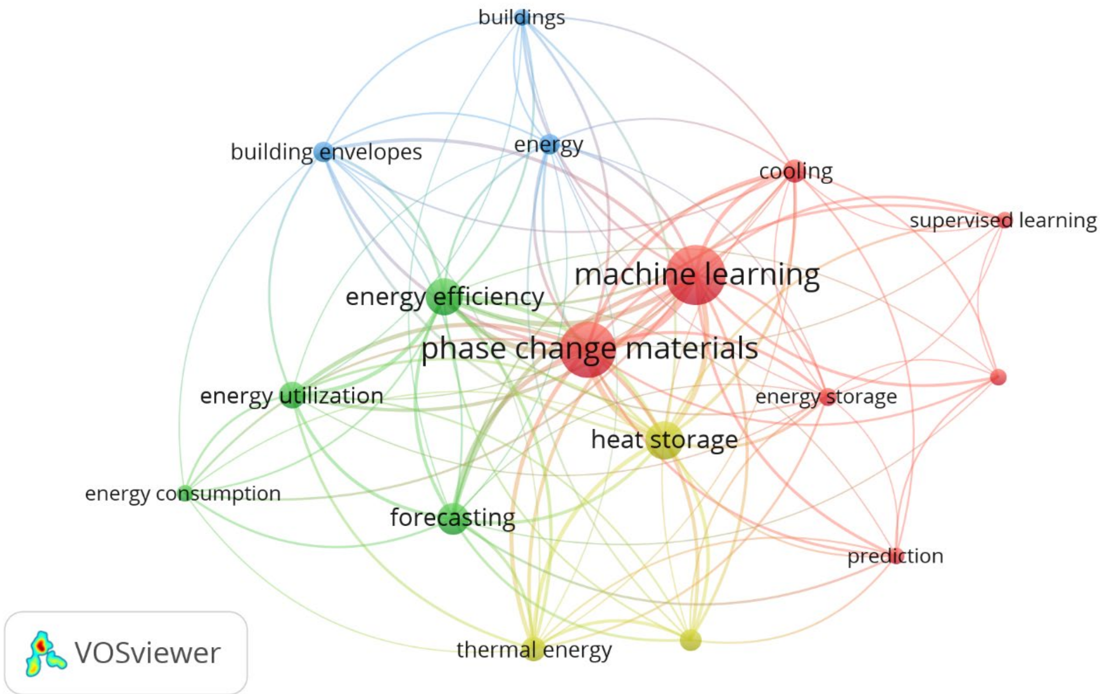

Figure 1. Keywords co-occurrence network.

2. Literature review

As an application of AI, machine learning models establish their relevance in the field of building engineering[16]. Numerous works related to the prediction of a building's energy consumption using machine learning have been thoroughly explored in the literature. For instance, Yan et al.[17] Introduces an innovative prediction framework that employs the stacking ensemble learning algorithm, combining several deep learning models to predict aspects of residential building performance. The algorithm utilizes a set of three base models and integrates multimodal inputs. The results demonstrate that the model exhibited markedly enhanced performance. Shrestha and Shimizu[18] evaluated the thermal performance of traditional Japanese wooden houses. They employed support vector regression and random forest regression models for predicting indoor temperatures based on external factors and adjacent room temperatures. The obtained results indicated that the temperature prediction model for air-conditioned rooms was influenced by the thermal conditions of adjacent non-air-conditioned spaces. In another study[19], the authors developed a hybrid method integrating BIM-DesignBuilder, Grey Wolf Optimization, Random Forest and Non-dominated Sorting Genetic Algorithm II, to optimize building design parameters. The obtained results for a building case show that Random Forest demonstrates strong predictive accuracy for life cycle carbon emissions, economic cost, and predicted mean vote. Furthermore, when applied with well-defined objectives, the multi-objective optimization framework of the RF-NSGA-II model serves as a tool to identify optimal solutions that balance enhanced internal comfort, reduced economic costs, and minimized carbon emission throughout a building's lifespan. Mohebbi et al.[20] employed a novel metaheuristic optimization algorithm, the Tyrannosaurus Rex Optimization Algorithm to enhance the prediction accuracy of three regression techniques. The depicted results proved that the developed models exhibit better performance in predicting residential heating load than traditional methods. Hussein et al.[21], explored the use of Random Forest to predict the energy performance of buildings. The findings revealed that although the machine learning model exhibited considerable promise in predicting energy consumption, based on factors such as wall thickness, orientation, and thermal mass, not all variables attained statistical significance. Roodkoly et al.[22] created a machine learning-based model to predict building energy performance metrics for high-performance building design (HPBD). Four different models, including Artificial Neural Network (ANN), Support Vector Machine (SVM), Random Forest (RF), and K-Nearest Neighbors (KNN), were evaluated. The results show that ANN model outperformed the other algorithms in predicting annual energy consumption, CO2 emissions, and percentage of comfort hours during the design phase. Qin et al.[23] focused on developing accurate prediction models for heating and cooling loads in nearly zero-energy buildings. Machine learning models, including Multivariate Polynomial Regression, Support Vector Regression, Multilayer Perceptron and Extreme Gradient Boosting are implemented and compared. The obtained results show that feature selection significantly improves prediction accuracy while reducing model complexity. In another study, Shen[24] compared various machine learning against measured heating demand. The GPAO model, combining GPR with the Arithmetic Optimization Algorithm, achieves a maximum coefficient of determination of 0.991 and minimum performance errors of 0.955, demonstrating its superiority over other machine learning models.

Some research works investigated the use of machine learning for predicting the thermal performance of PCM-integrated buildings. For example, Abbasian-Naghneh et al.[25] predicted the heating and cooling energy consumption using a neural network algorithm method. The model's results were then introduced into the genetic algorithm to determine the optimal annual energy consumption. Urresti et al.[26,27] evaluated the performance of ANN for the thermal analysis of buildings with PCM integration. They studied the generalization of the ANN and stated that the obtained ANNs didn't show good generalization even when using Bayesian Regularization. Yang et al.[28] predicted the PCM-enhanced building operational energy consumption by developing a stacking model combining eight typical ML models. The Non-Dominated Sorting Genetic Algorithm III (NSGA-III) was then combined with the stacking model. The findings proved that the proposed stacking model gives good results and demonstrated the effectiveness of NSGA-III in achieving high-dimensional multi-objective optimization. In another study[29], the authors compared three different machine learning models Support Vector Machines, Multiple regression, and Artificial Neural Networks to predict the energy demand of PCM-integrated housing. The findings indicate that ANN outperforms other models. Jraida et al.[15] evaluated the performance of thirteen machine learning models for predicting hourly energy consumption. The results demonstrated the potential of PCMs for energy savings in buildings and showed the effectiveness of machine learning, particularly SVM, in predicting building thermal performance across different climatic conditions in Morocco.

3. Methodology

The approach employed in this paper is based on two main steps: the generation of energy data through dynamic simulations and the application of hybrid machine learning models for predicting energy consumption. First, Using the TRNSYS tool, a residential building with PCM integrated into the roof was modeled. The simulations generated hourly data on energy consumption for heating, cooling, and total consumption of the building. In the second step, The data from the TRNSYS simulations were carefully prepared for the machine learning models by being divided into two sets: a training set (70% of the data) to fit the models and a test set (30% of the data) to assess their generalization capacity. The selected input variables include key climate parameters such as dry outdoor temperature, relative humidity, wind speed and direction, and total solar radiation, allowing to capture the climate variations specific to the 24 Moroccan cities studied. Six hybrid machine learning models, including ANN-LR, ANN-RR, ANN-RF, ANN-GB, ANN-DT, and ANN-ELM were evaluated and compared to the standalone model (ANN) based on the statistical metrics.

3.1. Data

In this study, a typical cubic residential building (3m × 3m × 3m) integrating phase change materials (PCMs) into the roof, was simulated under different climatic conditions covering 24 Moroccan cities. The building materials' composition, layer thickness, and thermal properties are presented in **Table 1**. The openings include a single-glazed window with a thermal transmittance of 5.74 W/m² K and a wooden door, all exposed to outdoor conditions. The PCM used, a mixture of paraffin (60%) and ethylene polymer (40%), exhibits a melting temperature of 21.7 to 31 °C, a latent enthalpy of 70 kJ/kg, and a thermal conductivity varying between 0.14 and 0.18 W/m K depending on the condition. The necessary climate data were collected from EPW (EnergyPlus Weather) files to capture the diversity of Moroccan weather conditions, ranging from sub-humid and semi-arid climates in the north to desert climates in the south[30]. The simulations were performed using the TRNSYS tool, coupling type 399, based on the enthalpy method to model heat transfers in PCM layers according to a one-dimensional Cranck-Nicolson numerical scheme, to type 56, used to simulate heat exchanges in multi-volume areas of the building. The simulations performed detailed hourly data on the energy needs for heating, cooling, and total energy consumption of the building.

Table 1. Characteristics of building components[15].

	Layer No.	Description	Thickness (cm)	Conductivity (W/m.K)	Density (kg/m3)	Capacity (kJ/kg.K)
External wall	1(inside)	Mortar	1	1.15	1700	1
	2	Redbrick	7.2	1.15	1700	0.794
	3	Air gap	12.6	0.08	1	1.227
	4	Redbrick	7.2	1.15	1700	0.794
	5(outside)	Mortar	1	1.15	1700	1
External roof	1(inside)	Mortar	2	1.15	1700	1
	2	Heavy concrete	15	1.75	2300	0.92
	3	Mortar	2	1.15	1700	1
	4	Bitumen sheet	3	0.5	1700	1
	5	Mortar	1	1.15	1700	1
	6(outside)	PCM	0.526	0.14-0.18	853	Cpcm(T)
Floor	1(inside)	Mortar	5	1.15	1700	1
	2	Clay	3	0.6	2000	1.5
	3	Polystyrene	2	0.039	25	1.38
	4	Heavy concrete	4	1.75	2300	0.92
	5(outside)	Limestone	10	2.25	2400	0.8

3.2. Machine learning models

To enhance the robustness of the model in predicting energy consumption, six different hybrid machine learning models were developed by combining the artificial neural network (ANN) with different machine learning techniques. Each learning algorithm follows a two-step approach: first, an ANN is used to generate initial predictions, then, these predictions are refined by another model, such as Random Forest (RF), Decision

Tree (DT), Gradient Boosting (GB), Extreme Learning Machine (ELM), Linear Regression (LR) and Linear Ridge Regression (RR), each contributing to correct or improve the errors of the initial ANN predictions. In this section, the different machine learning methods used to predict hourly energy consumption are described.

3.2.1. Artificial neural network (ANN*)*

Artificial Neural Network is based on the principles of human nervous system. Human nervous system includes components like neurons, synapse, and dendrites to process and transfer information between neurons[31]. Analogous to it, ANN contains components such as artificial neurons and weights and transfer functions to process the information. Different network models exist. The simplest one is the feed-forward model[32]. It consists of three layers, the first one is the input layer that receive data, the last layer, called output send the evaluated results and between these two layers, a hidden layer that is used to communicate with other neurons and determine the solution of the problem[33,34].

3.2.2. Decision tree (DT)

Decision tree is a supervised learning method applied to both classification and regression problems. It involves building a model in the shape of a tree structure, with internal nodes representing decisions based on specific variable values. The branches illustrate the possible outcomes of the decision, while the leaf node represents a final classification or prediction. The decision tree is built by starting with the root node and recursively dividing the data into smaller subsets based on the values of the variables. The process continues until the subsets are pure, meaning that they contain only one class or one output value. The final tree can then be used to make predictions on new data by following the decisions and corresponding branches until a leaf node is reached[35].

3.2.3. Extreme learning machine (ELM)

The extreme learning machine (ELM) is a single-hidden layer feedforward neural networks, originally proposed by Reference [36]. The model consists of three layers[37,38]:

The input layer, which imports the sample dataset (x_j, t_j) of input data ($a_j = [a_{j1}, a_{j2}, \dots, a_{jn}]^T \in R^n$) and target data ($t_j = [t_{j1}, t_{j2}, \dots, t_{jm}]^T \in R^m$).

The second layer is the hidden layer and is calculated using the equation below:

$$\sum_{i=1}^{L} \beta_i g(w_i x_j + b_i) = o_j \tag{1}$$

Where w_i is the initial weight randomization, β_i is the output weight relating the hidden nodes and the output nodes, L is the random number of hidden nodes, g(x) is the activation function and o_j is the output value.

The aim of the third output layer step is to find the matrix; via the following equation:

$$H\beta = T \tag{2}$$

Where H is the hidden layer output matrix, β is the output weight, and T is the desired output.

3.2.4. Gradient boosting (GB)

Gradient Boosting is an automatic learning method used in regression and classification[39]. The idea behind boosting is to create a sequence of simple models to address the errors from the preceding models [40]. This learning technique is based on a gradient descent optimization of a function, $F^*(x)$. The objective is to find an approximation, $\hat{F}^*(x)$, by minimizing the value of a differentiable loss function. The prediction of $F^*(x)$ is built as [41]:

$$F_m^*(x) = F_{m-1}^*(x) + p_m h_m(x) \tag{3}$$

Where p_m is the weight of the m^{th} function and h_m is a function to minimize the loss function. Each $h_m(x)$ is trained on the dataset and pseudo-residuals are calculated.

3.2.5. Randon forest (RF)

Random forests are a type of ensemble learning method that combines the predictions of multiple models to improve their overall performance. In this case, the individual models are decision trees, and the ensemble is formed by training many decision trees on different subsets of observations and variables. The idea behind random forests is that by training many decision trees on different subsets of the data, the resulting ensemble will be more accurate and less likely to overfit than any individual decision tree. This is because the errors made by the individual trees will tend to cancel each other out and because the diversity of the trees will capture a wider range of relationships in the data[42].

3.2.6. Linear regression (LR)

The linear regression model is considered one of the simplest machine learning algorithms for beginners in data mining, as it does not require tuning of parameters[42]. Additionally, it needs fewer computing resources and consequently has a faster prediction speed. Due to its simplicity and good prediction performance, this model has been widely used in many fields. The approach of regression involves modeling the linear relationship between the independent variable x and the dependent variable y that is under analysis. One of the most common regression models of linear regression is expressed as follows:

$$y = a_1x_1 + a_2x_2 + \cdots + a_ix_i + \cdots + a_nx_n + \varepsilon \tag{4}$$

Where a is the regression coefficient of the explanatory variables, ε is a random deviation or error term, and n is the dimension of the explanatory variables.

3.2.7. Linear ridge regression (LRR)

The ridge regression was introduced by Hoerl and Kennard (1970) as a technique to address highly correlated regressors and stabilize the solution of the linear regression problem[43]. It is fundamentally a modified least-squared estimation method for the dataset exhibiting multicollinearity[44]. The Key benefit of ridge regression is that it reduces the complexity of the model by penalizing large coefficients, which improves its generalization ability and avoids overfitting.

3.3. Performance measure

In order to assess the performance of ML models, widely used evaluation metrics are calculated. The mathematical equations of these statistical indicators are listed below:

i. Root Mean Square Error (RMSE):

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - y_i'\right)^2} \tag{5}$$

ii. Coefficient of determination (R^2):

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(y_i - y_i'\right)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2} \tag{6}$$

iii. Standard deviation (σ):

$$\sigma = \sqrt{\frac{n}{n-1}\left(RMSE^2 - MBE^2\right)} \tag{7}$$

iv. Mean Absolute Error (MAE):

$$MAE = \frac{1}{n}\sum_{i}^{n}\left|y_i - y_i'\right| \tag{8}$$

Where n is the data set size, y_i is the actual value, y'_i is the predicted hourly energy consumption data and $\overline{y_i}$ is the mean of y_i.

In general, lower values of the RMSE, MAE and σ values, yield superior results. The result of R^2 value around 1 suggests the model generates predictions with no error. The higher the R^2 value, the more favorable the results.

3. Results

The accuracy of the hybrid machine learning models in predicting energy consumption is evaluated using four statistical performance measures; including, mean absolute error (MAE), root mean square error (RMSE), coefficient of determination (R^2), and standard deviation (σ). The results of the hybrid machine learning model evaluations are presented in **Figures 2-8** and listed in **Tables 2-4** for cooling, heating, and total energy consumption.

Figures 2-4 compare the R^2 metric for the standalone model (ANN) with the studied hybrid machine learning models across all targets during the training and testing phases. We can note that the performance of the ANN model for all datasets is always greatly lower than for other hybrid machine learning. When predicting cooling, heating, and total energy consumption the individual model yielded an R^2 of 0.6303, 0.7459, and 0.00 for the training datasets, along with an R^2 of 0.6278, 0.7671 and 0.00 for the testing datasets. For cooling energy consumption, the R^2 for all the hybrid models falls within the range of 0.8089 (for the ANN-RR in the phase of training) to 0.9185 (for the testing phase of the ANN-LR hybrid model). When predicting heating energy consumption, the R^2 for all the hybrid models falls within the range of 0.7567 (for the ANN-RR in the testing phase) to 0.8142 (for the training phase of the ANN-LR hybrid model). For total energy consumption, the R^2 for all the hybrid models falls within the range of 0.7936 (for the ANN-RR in the testing phase) to 0.8586 (for the testing phase of the ANN-LR hybrid model). According to the plots, the ANN-LR model achieves better results as shown by the higher R^2 value for all datasets. The R^2 result is 0.9185, suggesting that the model elucidates about 91.85% of the variability in cooling energy consumption. Subsequently, the ANN-RR model yields lower results than the other hybrid machine learning. The R^2 result is 0.7936 implying that the model elucidates 79.36% of the variation in the total energy consumption, with a relatively small proportion of the variation left unexplained. It is worth noting that despite the ANN-RR model's R^2, it continues outperforming the standalone model (ANN) demonstrating superior predictive performance even with a portion of the variation left unexplained.

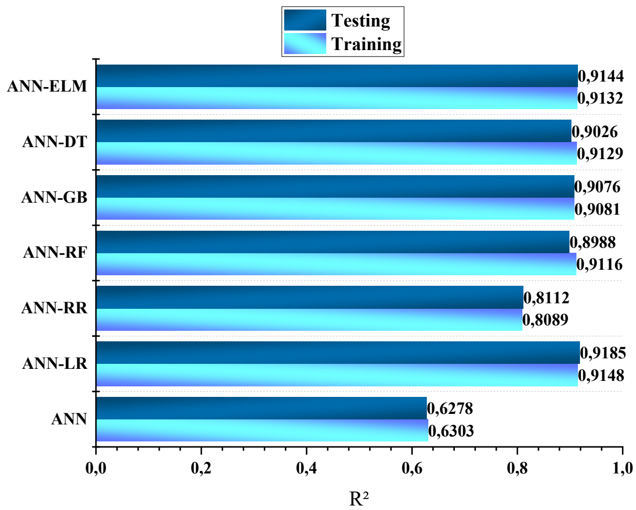

Figure 2. R^2 calculation for the cooling energy consumption prediction.

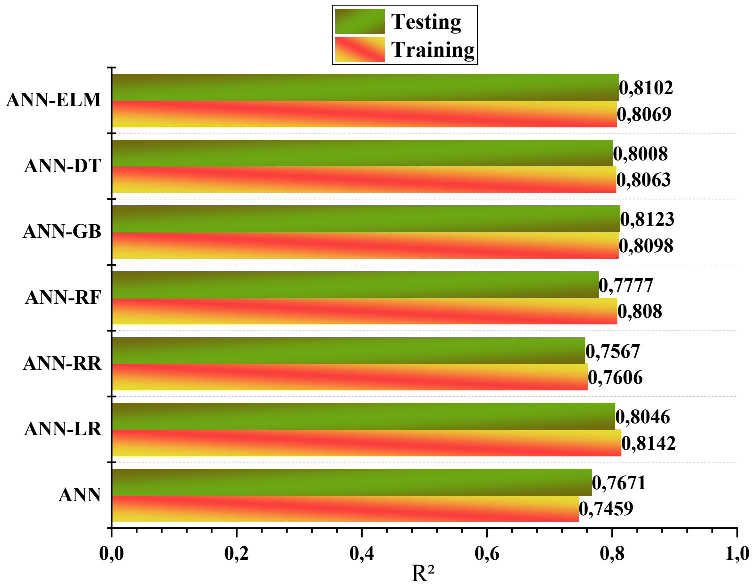

Figure 3. R^2 calculation for the heating energy consumption prediction.

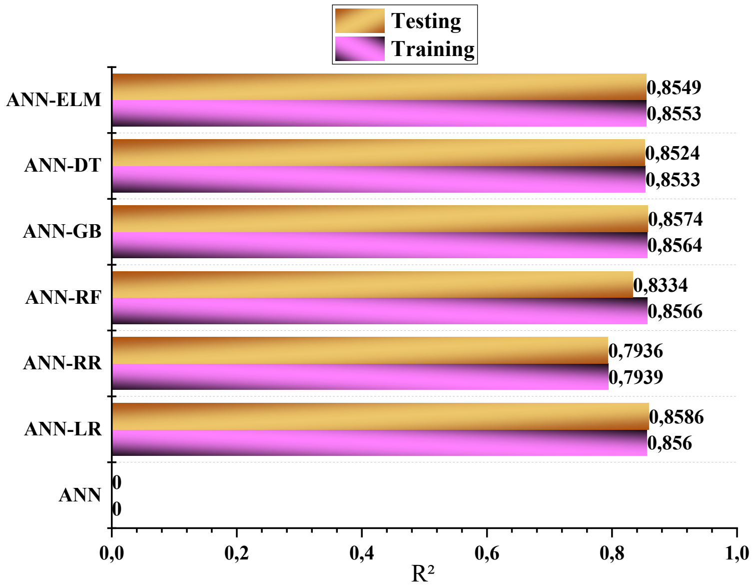

Figure 4. R^2 calculation for the total energy consumption prediction.

Figure 5 presents the results of RMSE achieved by the standalone and hybrid machine learning models for predicting the three outputs during the testing phase. The findings indicate that the hybrid methods outperform the individual method in terms of performance. For cooling energy consumption, the standalone model shows a high RMSE of 237.15, while the hybrid models, especially the ANN-LR (RMSE=94.66), show significantly better results, with values ranging from 94.66 to 143.04. For heating energy consumption, the ANN model shows an RMSE of 106.58, while the hybrid methods show close performances, with an RMSE ranging between 92.39 and 106.14. In particular, the ANN-GB (RMSE = 92.39) shows a slight improvement over the individual method. For total energy consumption, the ANN model shows a high RMSE of 324.81, while the hybrid methods provide much more accurate results, with an RMSE ranging from 136.66 to 163.97.

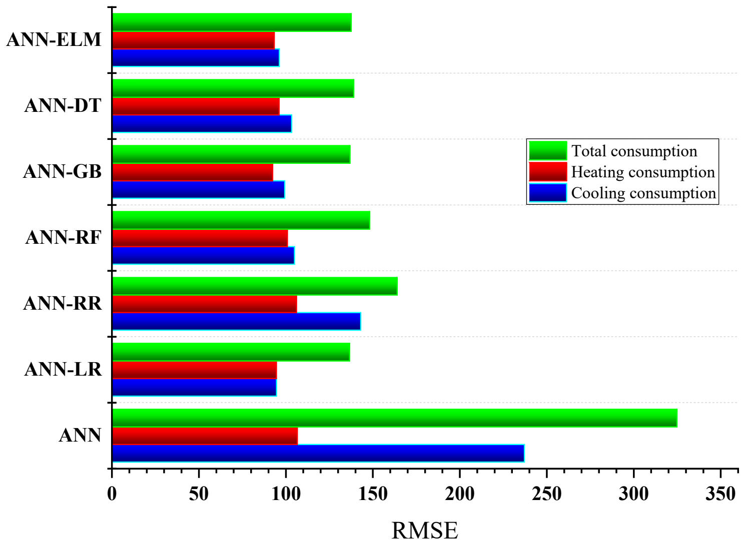

Figure 5. RMSE calculation for the prediction datasets.

Figure 6 presents the scatter plots of the actual cooling energy consumption versus predicted values, while **Figures 7** and **8** present the actual heating energy consumption versus predicted, and the actual total energy consumption versus predicted, respectively. A 45° straight line (y=x) in these graphs represents the perfect fit between the predicted and actual values. As shown in **Figures 6-8** the results indicate a notable difference in the performance of models to predict cooling, heating, and total energy consumption. For cooling energy consumption, the ANN-RF and ANN-GB models exhibit the closest alignment to the ideal line, with the training and testing results showing a relatively tight distribution, which illustrates a strong correlation between predicted and actual values. However, the ANN-DT and ANN-RR models show more scattered predictions, especially for higher values. Which suggests reduced accuracy. For heating energy consumption, the ANN-LR and ANN-ELM models show better alignment with the ideal line, especially for the testing data, which demonstrates their capability to capture heating patterns effectively. The dispersion of points is in general lower for heating than for cooling, which demonstrates that heating predictions are less complex for the models to handle. However, the ANN-RF and ANN-GB models indicate less reliable predictions for extreme cases. They display slightly more spread at higher values. For total energy consumption, the results depicted in **Figure 8** reveal an increased dispersion for all models, especially at higher values, which reflects the increased complexity of predicting total energy consumption. The ANN-LR and ANN-ELM models maintained relatively good alignment, but the scatter was more pronounced compared to heating. However, the ANN-RF and ANN-GB models, which performed well for cooling and heating, struggled to maintain the same level of correlation for total energy consumption, with points deviating more from the ideal line. Overall, the figures suggest that model performance varies significantly depending on the target energy type, with cooling and heating showing good alignment, while total energy predictions appear more challenging, resulting in greater scatter and reduced alignment across all models.

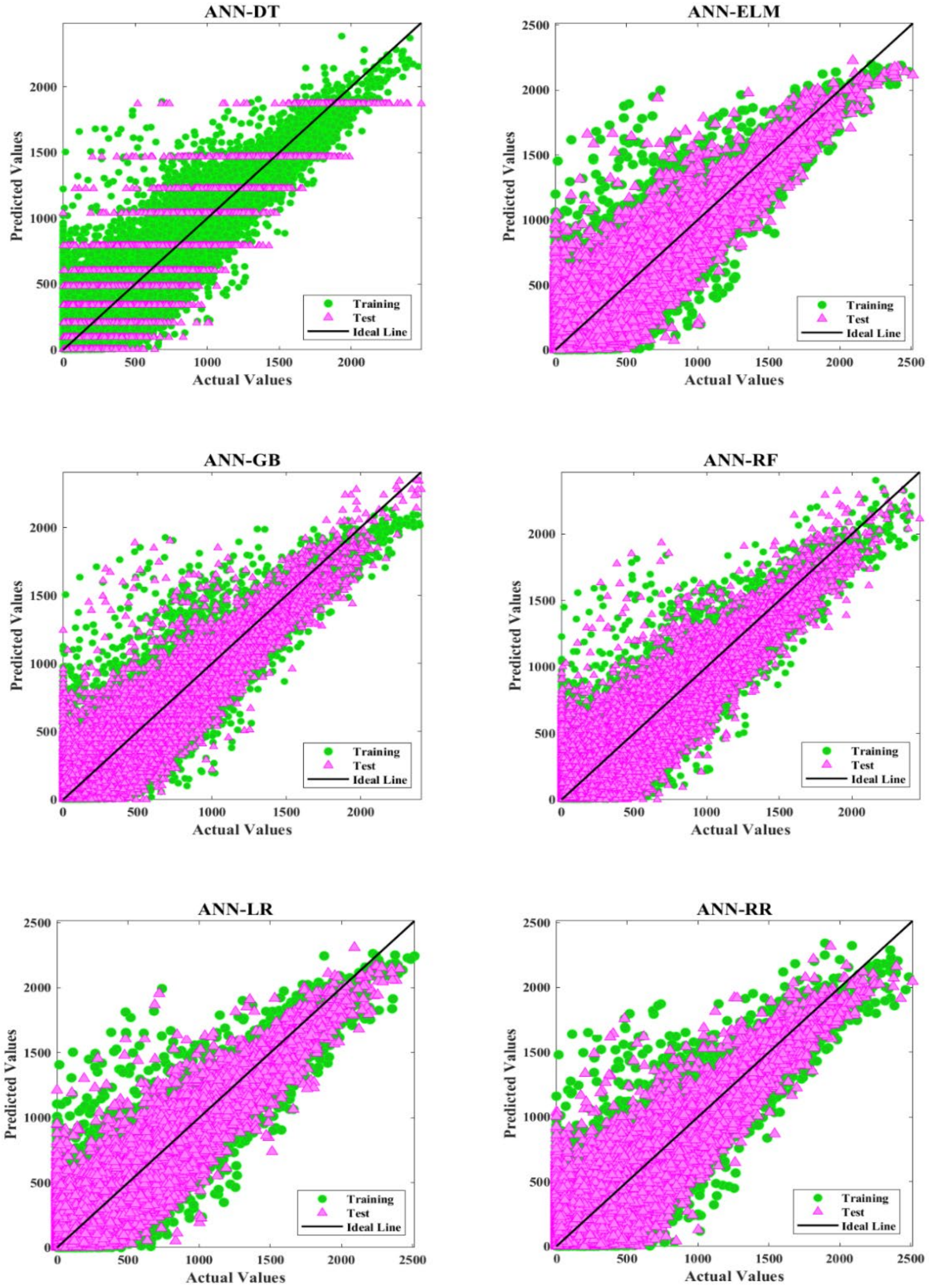

Figure 6. Scatter plot of actual vs predicted cooling energy consumption.

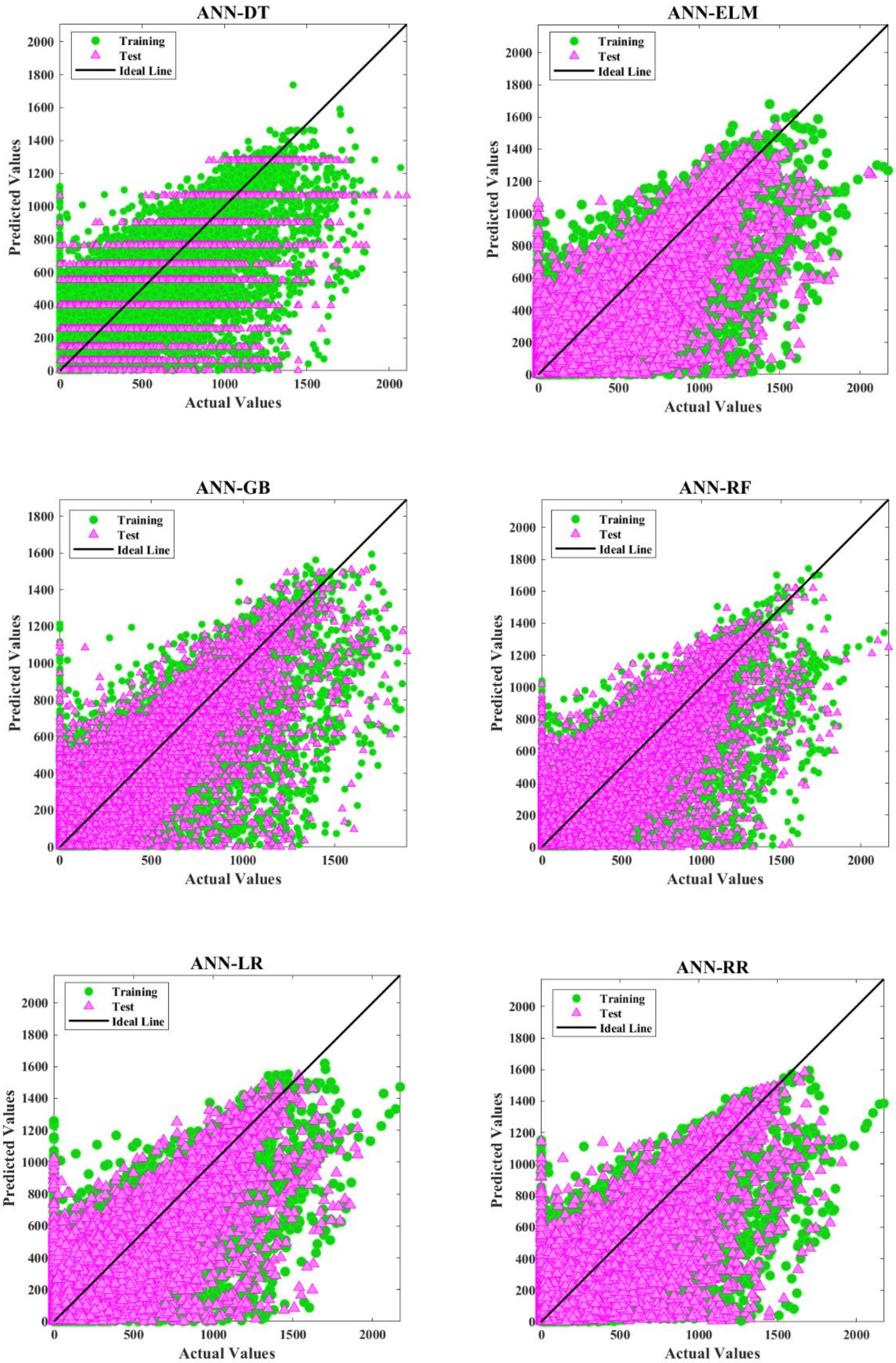

Figure 7. Scatter plot of actual vs predicted heating energy consumption.

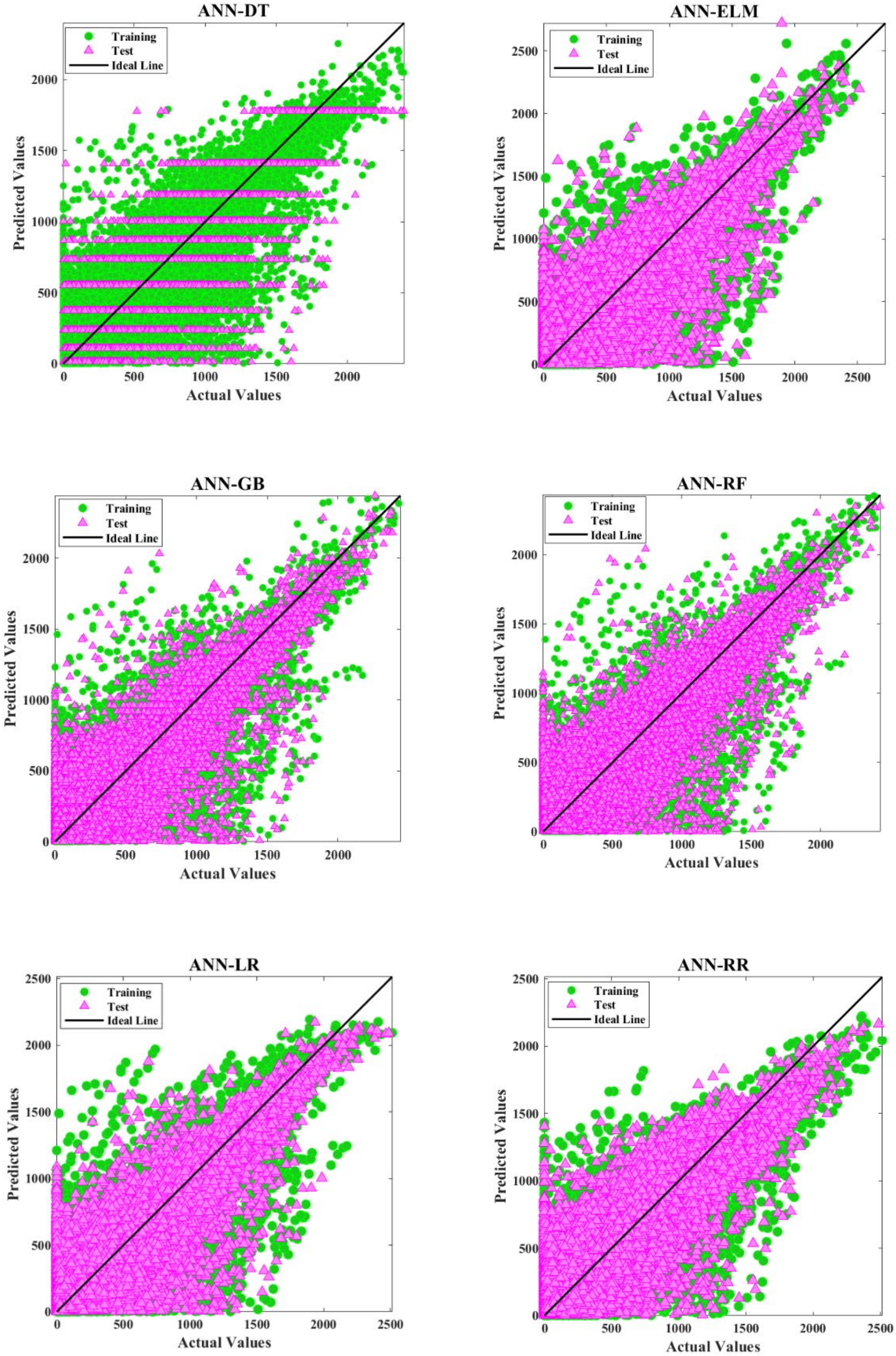

Figure 8. Scatter plot of actual vs predicted total energy consumption.

The outcome of statistical evaluators of all the targets in the training and testing phases, as delineated in **Tables 2-4**, reveals that the hybrid machine learning models show good stability between the training and testing phases, reflecting their ability to generalize. For the cooling energy consumption, the ANN-LR stands out for a relatively low MAE (47.19 in training and 46.92 in testing) and relatively stable standard deviation, with values close to 315 in both phases. It also obtains a Rank 1, both in training and testing, indicating a good performance in terms of precision and generalization capacity. In contrast, the ANN-RR has a relatively higher MAE (118.12 in training and 117.08 in testing), reflecting greater error variability and lower stability, placing it in rank 6 in both phases. Regarding heating energy consumption and total energy consumption, although the ANN-LR maintains a good position with a rank 1 in training, it is the ANN-GB that obtains the best performance in the testing phase with a rank 1, which indicates that it generalizes better than the other methods on the test data, despite standard deviations relatively close to those of the other methods. On the other hand, the ANN-RR, remains the least efficient in rank 6, making it the worst-performing method in all outputs.

Table 2. The statistical metrics for all hybrid ML models for cooling energy consumption.

Method	MAE	RMSE	R^2	σ	RANK	MAE	RMSE	R^2	σ	RANK
ANN-LR	47.192	95.79	0.91489	314.07	1	46.923	94.667	0.91858	315.36	1
ANN-RR	118.12	144.01	0.8089	288.02	6	117.08	143.04	0.81125	290.03	6
ANN-RF	51.293	97.792	0.91169	313.85	4	49.284	105	0.8988	318.48	5
ANN-GB	52.961	100.15	0.90814	314.83	5	46.987	99.344	0.90764	313.87	3
ANN-DT	50.176	96.947	0.91295	313.23	2	49.592	103.35	0.90264	313.85	4
ANN-ELM	49.45	97.014	0.91327	314.81	3	49.228	96.293	0.91448	313.73	2

Table 3. The statistical metrics for all hybrid ML models for heating energy consumption.

Method	MAE	RMSE	R^2	σ	RANK	MAE	RMSE	R^2	σ	RANK
ANN-LR	35.611	92.625	0.8142	193.9	1	35.999	94.566	0.80463	192.83	3
ANN-RR	64.054	104.86	0.76068	184.72	6	64.302	106.14	0.75678	183.81	6
ANN-RF	36.378	94.123	0.80808	193.71	3	35.176	100.9	0.77775	195.54	5
ANN-GB	37.993	93.816	0.80987	193.23	5	32.739	92.398	0.81237	193.22	1
ANN-DT	37.612	94.298	0.80636	192.29	2	34.527	96.094	0.80087	192.94	4
ANN-ELM	37.683	94.364	0.80696	192.94	4	37.578	93.309	0.81025	191.26	2

Table 4. The statistical metrics for all hybrid ML models for total energy consumption**.**

Method	MAE	RMSE	R^2	σ	RANK	MAE	RMSE	R^2	σ	RANK
ANN-LR	83.745	136.69	0.8560	333.29	1	84.186	136.66	0.8586	332.66	2
ANN-RR	120.34	164.03	0.7939	320.14	6	120.57	163.97	0.7936	320.16	6
ANN-RF	84.267	136.44	0.8566	334.08	2	86.858	148.24	0.8334	337.42	5
ANN-GB	86.171	136.64	0.8564	334.18	4	79.863	136.86	0.8574	334.64	1
ANN-DT	86.383	138.24	0.8533	333.69	5	83.277	139.02	0.8524	332.33	3
ANN-ELM	83.852	137.38	0.8553	334.09	3	84.046	137.56	0.8549	333.68	4

5. Conclusion

In this research paper, six hybrid models were developed by combining the artificial neural network with several other machine learning techniques. The process consisted of first training the ANN on the input data, obtained from the results of the dynamic simulation using TRNSYS, to get initial cooling, heating, and total energy consumption predictions. These predictions were then used as input variables for other learning models. The performance of the hybrid machine learning models, namely, ANN-LR, ANN-RR, ANN-RF, ANN-GB, ANN-DT, and ANN-ELM is evaluated using different statistical performance metrics. The results showed that the hybrid machine learning models surpass the standalone model (ANN) in terms of performance. Some hybrid machine learning methods outperformed others depending on the phase and data. The ANN-LR and ANN-GB models emerged as the most reliable hybrid approaches for accurate prediction, while ANN-RR model proved to be the least effective. Additionally, total energy consumption was more complex and challenging to predict.

Conflict of interest

The authors declare that there is no conflict of interest.

References

1. Chen, X.; Vand, B.; Baldi, S. Challenges and Strategies for Achieving High Energy Efficiency in Building Districts. Buildings 2024, 14, 1839, doi:10.3390/buildings14061839.
2. Chen, X.; Vand, B.; Baldi, S. Challenges and Strategies for Achieving High Energy Efficiency in Building Districts. Buildings 2024, 14, 1839, doi:10.3390/buildings14061839.
3. Li, L.; Wang, Y.; Wang, M.; Hu, W.; Sun, Y. Impacts of Multiple Factors on Energy Consumption of Aging Residential Buildings Based on a System Dynamics Model--Taking Northwest China as an Example. J. Build. Eng. 2021, 44, 102595, doi:10.1016/j.jobe.2021.102595.
4. Jraida, K.; Farchi, A.; Mounir, B.; Mounir, I. A Study on the Optimum Insulation Thicknesses of Building Walls with Respect to Different Zones in Morocco. Int. J. Ambient Energy 2017, 38, 550–555, doi:10.1080/01430750.2016.1155490.
5. Abdou, N.; El Mghouchi, Y.; Jraida, K.; Hamdaoui, S.; Hajou, A.; Mouqallid, M. Prediction and Optimization of Heating and Cooling Loads for Low Energy Buildings in Morocco: An Application of Hybrid Machine Learning Methods. J. Build. Eng. 2022, 61, 105332, doi:10.1016/j.jobe.2022.105332.
6. United Nations Environment Programme 2023 Global Status Report for Buildings and Construction: Beyond Foundations - Mainstreaming Sustainable Solutions to Cut Emissions from the Buildings Sector; United Nations Environment Programme, 2024; ISBN 978-92-807-4131-5.
7. Chen, L.; Zhang, Y.; Chen, Z.; Dong, Y.; Jiang, Y.; Hua, J.; Liu, Y.; Osman, A.I.; Farghali, M.; Huang, L.; et al. Biomaterials Technology and Policies in the Building Sector: A Review. Environ. Chem. Lett. 2024, 22, 715–750, doi:10.1007/s10311-023-01689-w.
8. Pathway to Sustainability: An Overview of Renewable Energy Integration in Building Systems Available online: https://www.mdpi.com/2071-1050/16/2/638 (accessed on 26 November 2024).
9. Chen, L.; Hu, Y.; Wang, R.; Li, X.; Chen, Z.; Hua, J.; Osman, A.I.; Farghali, M.; Huang, L.; Li, J.; et al. Green Building Practices to Integrate Renewable Energy in the Construction Sector: A Review. Environ. Chem. Lett. 2024, 22, 751–784, doi:10.1007/s10311-023-01675-2.
10. Chua, K.J.; Chou, S.K.; Yang, W.M.; Yan, J. Achieving Better Energy-Efficient Air Conditioning – A Review of Technologies and Strategies. Appl. Energy 2013, 104, 87–104, doi:10.1016/j.apenergy.2012.10.037.
11. Yan, B.; Hao, F.; Meng, X. When Artificial Intelligence Meets Building Energy Efficiency, a Review Focusing on Zero Energy Building. Artif. Intell. Rev. 2021, 54, 2193–2220, doi:10.1007/s10462-020-09902-w.
12. Kośny, J. PCM-Enhanced Building Components: An Application of Phase Change Materials in Building Envelopes and Internal Structures; Springer, 2015; ISBN 978-3-319-14286-9.
13. Frigione, M.; Lettieri, M.; Sarcinella, A.; Barroso de Aguiar, J.L. Applications of Sustainable Polymer-Based Phase Change Materials in Mortars Composed by Different Binders. Materials 2019, 12, 3502, doi:10.3390/ma12213502.
14. Nazir, K.; Memon, S.A.; Saurbayeva, A. A Novel Framework for Developing a Machine Learning-Based Forecasting Model Using Multi-Stage Sensitivity Analysis to Predict the Energy Consumption of PCM-Integrated Building. Appl. Energy 2024, 376, 124180, doi:10.1016/j.apenergy.2024.124180.
15. Jraida, K.; EL Mghouchi, Y.; Mourid, A.; Haidar, C.; EL Alami, M. Machine Learning-Based Predicting of PCM-Integrated Building Thermal Performance: An Application under Various Weather Conditions in Morocco. J. Build. Eng. 2024, 96, 110395, doi:10.1016/j.jobe.2024.110395.
16. Machine Learning for Estimation of Building Energy Consumption and Performance: A Review | Visualization in Engineering Available online: https://link.springer.com/article/10.1186/s40327-018-0064-7 (accessed on 23 November 2024).
17. Yan, H.; Ji, G.; Cao, S.; Zhang, B. Developing an Integrated Prediction Model for Daylighting, Thermal Comfort, and Energy Consumption in Residential Buildings Based on the Stacking Ensemble Learning Algorithm. Build. Simul. 2024, 17, 2125–2143, doi:10.1007/s12273-024-1181-y.
18. Shrestha, A.; Shimizu, T. Thermal Performance Assessment of Traditional Japanese Wooden Houses with Short-Term Measurements: Machine Learning Approaches. J. Build. Eng. 2024, 97, 110954, doi:10.1016/j.jobe.2024.110954.
19. Liu, Y.; Li, T.; Xu, W.; Wang, Q.; Huang, H.; He, B.-J. Building Information Modelling-Enabled Multi-Objective Optimization for Energy Consumption Parametric Analysis in Green Buildings Design Using Hybrid Machine Learning Algorithms. Energy Build. 2023, 300, 113665, doi:10.1016/j.enbuild.2023.113665.

20. Mohebbi, M.; Afzal, S. Enhancing Residential Heating Load Prediction with Advanced Machine Learning and Optimization Techniques. J. Build. Eng. 2024, 95, 110199, doi:10.1016/j.jobe.2024.110199.
21. Hussien, A.; Khan, W.; Hussain, A.; Liatsis, P.; Al-Shamma'a, A.; Al-Jumeily, D. Predicting Energy Performances of Buildings' Envelope Wall Materials via the Random Forest Algorithm. J. Build. Eng. 2023, 69, 106263, doi:10.1016/j.jobe.2023.106263.
22. Roodkoly, S.H.; Fard, Z.Q.; Tahsildoost, M.; Zomorodian, Z.; Karami, M. Development of a Simulation-Based ANN Framework for Predicting Energy Consumption Metrics: A Case Study of an Office Building. Energy Effic. 2024, 17, 5, doi:10.1007/s12053-024-10185-1.
23. Qin, H.; Yu, Z.; Li, Z.; Li, H.; Zhang, Y. Nearly Zero-Energy Building Load Forecasts through the Competition of Four Machine Learning Techniques. Buildings 2024, 14, 147, doi:10.3390/buildings14010147.
24. Shen, Y. Load Estimation Models for the Heat Demand of Buildings: Application of Optimized Gaussian Process Regression. J. Appl. Sci. Eng. 2024, 28, 527–541, doi:10.6180/jase.202503_28(3).0010.
25. Abbasian-Naghneh, S.; Kalbasi, R. Implementation of ANN and GA on Building with PCM at Various Setpoints, PCM Types, and Installation Locations to Boost Energy Saving and CO2 Saving. Eng. Anal. Bound. Elem. 2022, 144, 110–126, doi:10.1016/j.enganabound.2022.08.006.
26. Urresti, A.; Sala, J.M.; García-Romero, A.; Diarce, G.; Escudero, C. Validation of Heat Transfer Models for PCMs with a Conductivimeter. Energy Procedia 2012, 30, 395–403, doi:10.1016/j.egypro.2012.11.047.
27. Urresti, A.; Campos-Celador, A.; Sala, J.M. Dynamic Neural Networks to Analyze the Behavior of Phase Change Materials Embedded in Building Envelopes. Appl. Therm. Eng. 2019, 158, 113783, doi:10.1016/j.applthermaleng.2019.113783.
28. Yang, H.; Xu, Z.; Shi, Y.; Tang, W.; Liu, C.; Yunusa-Kaltungo, A.; Cui, H. Multi-Objective Optimization Designs of Phase Change Material-Enhanced Building Using the Integration of the Stacking Model and NSGA-III Algorithm. J. Energy Storage 2023, 68, 107807, doi:10.1016/j.est.2023.107807.
29. Zhussupbekov, M.; Memon, S.A.; Khawaja, S.A.; Nazir, K.; Kim, J. Forecasting Energy Demand of PCM Integrated Residential Buildings: A Machine Learning Approach. J. Build. Eng. 2023, 70, 106335, doi:10.1016/j.jobe.2023.106335.
30. EnergyPlus Available online: https://energyplus.net/weather (accessed on 10 February 2025).
31. Renganathan, V. Machine Learning Algorithms for Data Scientists: An Overview; Vinaitheerthan Renganathan, 2021; ISBN 978-93-5473-769-5.
32. Neto, A.H.; Fiorelli, F.A.S. Comparison between Detailed Model Simulation and Artificial Neural Network for Forecasting Building Energy Consumption. Energy Build. 2008, 40, 2169–2176, doi:10.1016/j.enbuild.2008.06.013.
33. El Mghouchi, Y.; Chham, E.; Zemmouri, E.M.; El Bouardi, A. Assessment of Different Combinations of Meteorological Parameters for Predicting Daily Global Solar Radiation Using Artificial Neural Networks. Build. Environ. 2019, 149, 607–622, doi:10.1016/j.buildenv.2018.12.055.
34. Adedeji, B.P. Electric Vehicles Survey and a Multifunctional Artificial Neural Network for Predicting Energy Consumption in All-Electric Vehicles. Results Eng. 2023, 19, 101283, doi:10.1016/j.rineng.2023.101283.
35. Suthaharan, S. Decision Tree Learning. In Machine Learning Models and Algorithms for Big Data Classification: Thinking with Examples for Effective Learning; Suthaharan, S., Ed.; Springer US: Boston, MA, 2016; pp. 237–269 ISBN 978-1-4899-7641-3.
36. Huang, G.-B. What Are Extreme Learning Machines? Filling the Gap Between Frank Rosenblatt's Dream and John von Neumann's Puzzle. Cogn. Comput. 2015, 7, 263–278, doi:10.1007/s12559-015-9333-0.
37. Liu, C.; Sun, B.; Zhang, C.; Li, F. A Hybrid Prediction Model for Residential Electricity Consumption Using Holt-Winters and Extreme Learning Machine. Appl. Energy 2020, 275, 115383, doi:10.1016/j.apenergy.2020.115383.
38. Metaheuristic Extreme Learning Machine for Improving Performance of Electric Energy Demand Forecasting Available online: https://www.mdpi.com/2073-431X/11/5/66 (accessed on 30 November 2024).
39. Di Persio, L.; Fraccarolo, N. Energy Consumption Forecasts by Gradient Boosting Regression Trees. Mathematics 2023, 11, 1068, doi:10.3390/math11051068.
40. Touzani, S.; Granderson, J.; Fernandes, S. Gradient Boosting Machine for Modeling the Energy Consumption of Commercial Buildings. Energy Build. 2018, 158, 1533–1543, doi:10.1016/j.enbuild.2017.11.039.
41. Bassi, A.; Shenoy, A.; Sharma, A.; Sigurdson, H.; Glossop, C.; Chan, J.H. Building Energy Consumption Forecasting: A Comparison of Gradient Boosting Models. In Proceedings of the Proceedings of the 12th International Conference on Advances in Information Technology; Association for Computing Machinery: New York, NY, USA, July 20 2021; pp. 1–9.
42. Chen, Y.; Guo, M.; Chen, Z.; Chen, Z.; Ji, Y. Physical Energy and Data-Driven Models in Building Energy Prediction: A Review. Energy Rep. 2022, 8, 2656–2671, doi:10.1016/j.egyr.2022.01.162.
43. Masini, R.P.; Medeiros, M.C.; Mendes, E.F. Machine Learning Advances for Time Series Forecasting. J. Econ. Surv. 2023, 37, 76–111, doi:10.1111/joes.12429.
44. Xiao, C. Using Machine Learning for Exploratory Data Analysis and Predictive Models on Large Datasets. Master thesis, University of Stavanger, Norway, 2015.