Energy Efficiency in Smart Buildings through Prediction modeling and Optimization Using a Modified Whale Optimization Algorithm

Nasima EL ASSRI[1]*, *Mohammed* ENNEJJAR[1], *Mohammed Ali* JALLAL[1, 3], *Samira* CHABAA[1,2], *Abdelouhab* ZEROUAL[1]

[1]I2SP Research Team, Physics department, Faculty of Sciences Semlalia, Cadi Ayyad University, Marrakesh, Morocco
[2]Industrial Engineering Department, National School of Applied Sciences, Ibn Zohr University, Agadir, Morocco
[3]Univ. Grenoble Alpes, CEA, Liten, Campus Ines, 73375, Le Bourget du Lac, France

Abstract. This paper presents a comparative study on the prediction of energy consumption in buildings using machine learning techniques. The dataset encompasses a diverse range of buildings with 8 input features and one output variable, representing the energy consumption. The primary focus is on evaluating the performance of two prominent and widely-used machine learning algorithms: Artificial Neural Networks (ANN) and Random Forest (RF). The results indicate a promising predictive capacity of both models, showcasing their effectiveness in capturing intricate patterns within the dataset. In the case of ANN, the Root Mean Squared Error (RMSE) is reported at 3.806, demonstrating the model's ability to approximate the true energy consumption values. Furthermore, the Random Forest model exhibits enhanced predictive accuracy, as reflected by a lower RMSE of 1.392. In addition to predictive analysis, this study utilizes a Modified Whale Optimization Algorithm (MWOA) to optimize energy consumption. The MWOA helps to identify the associated input values that lead to the lowest possible energy consumption, providing valuable insights for energy-efficient building design. The implications of this research extend to the broader field of sustainable architecture and urban planning, paving the way for more informed decisions aimed at reducing energy consumption and fostering environmental sustainability.

1 Introduction

As global energy consumption continues to rise, there is a growing emphasis on improving energy efficiency, particularly within the built environment [1], [2]. Buildings account for a significant portion of overall energy use, making them a crucial area of focus for achieving sustainability goals. Reducing energy consumption in buildings not only mitigates environmental impacts but also contributes to the long-term economic and operational efficiency of urban spaces[3].

In recent years, the application of machine learning techniques has garnered considerable attention in various fields, including energy prediction [4]. These techniques offer the potential to uncover complex patterns in data and provide more accurate predictions, enabling decision-makers to optimize energy use. In the context of building energy consumption, advanced machine learning models can offer insights into how various factors, such as building characteristics, occupancy, and weather conditions, influence energy use [5].

This study aims to explore the effectiveness of two widely recognized machine learning algorithms: Artificial Neural Networks and Random Forest in predicting energy consumption in buildings. By leveraging a dataset that encompasses a diverse range of building types and energy use profiles, we compare the performance of these models in terms of predictive accuracy. The goal is to determine which model provides better approximations of true energy consumption and to assess their potential for informing energy-efficient design and operation in the building sector.

Furthermore, this study utilizes a Modified Whale Optimization Algorithm (MWOA) to optimize and minimize energy consumption. This approach aims to identify the associated input values that lead to the lowest possible energy consumption, providing valuable insights for energy-efficient building design and operation. The significance of this research lies in its potential contributions to sustainable architecture and urban planning. By offering a data-driven approach to energy management, this work can inform decision-making and contribute to the broader objective of environmental sustainability[6].

The findings from this study hold significant implications for the building industry and the broader pursuit of environmental sustainability. By providing insights into the effectiveness of different machine learning algorithms in predicting energy consumption, this research can inform the design and operation of energy-efficient buildings, ultimately contributing to the

* Corresponding author: nasima.elassri@ced.uca.ma

reduction of overall energy usage and environmental impact [7].

This study is focused on the prediction and the optimization of building energy consumption using machine learning techniques. The structure of this paper is as follows: Section 1 provides an introduction to the study, while Section 2 details the methods employed, including subsections on data description, Artificial Neural Networks (ANN), Random Forest (RF), and the Modified Whale Optimization Algorithm (MWOA). Section 4 presents the results and discussion of the findings, followed by Section 5, which concludes the study and discusses the implications for energy-efficient building design.

2 Methodology

2.1 Data Description

The dataset comprises a comprehensive set of 3840 records, and its primary objective is to investigate the intricate relationships between the energy consumption of residential buildings and eight distinct input variables [8]. These input variables encompass essential features of the buildings, including: Building Size (x1), Floor Height (x2), Glazing Area (x3), Wall Area (x4), window-to-wall ratio (WWR) (x5), Win Glazing U-value (x6), Roof U-value (x7), and External Wall U-value (x8). Each record within the dataset is characterized by these input variables, with the corresponding energy consumption output variable denoted as BEC (MWh). The inclusion of these specific features allows for a comprehensive analysis of the impact of various architectural and structural elements on the energy consumption patterns of residential buildings. The IES simulation software was utilized to create the simulated buildings[9].

2.2 Artificial *Neural Networks (ANNs)*

ANNs stand as dynamic computational structures inspired by the human brain's neural networks[10]. These models comprise layers with interconnected nodes, or neurons, which collectively process and learn complex patterns from input data[11]. In the context of building energy consumption prediction, we harness the formidable capabilities of ANN to decipher intricate relationships among various building parameters. The ANN is particularly adept at capturing nonlinear dependencies between inputs, making it well-suited for the multifaceted nature of building energy systems. The predictive power of the network is expressed mathematically through the following equation [12], where BEC_i represents the output of the ith neuron:

$$BEC_i = f(\sum_{j=1}^{n} w_{ij} * x_j + b_i) \tag{1}$$

Where, w_{ij} denotes the weight between the ith neuron and the jth input, x_j is the jth input, b_i is the bias term, and f is the activation function. Figure 1 illustrates the configuration of a multi-layered feedforward backpropagation network employed in this study.

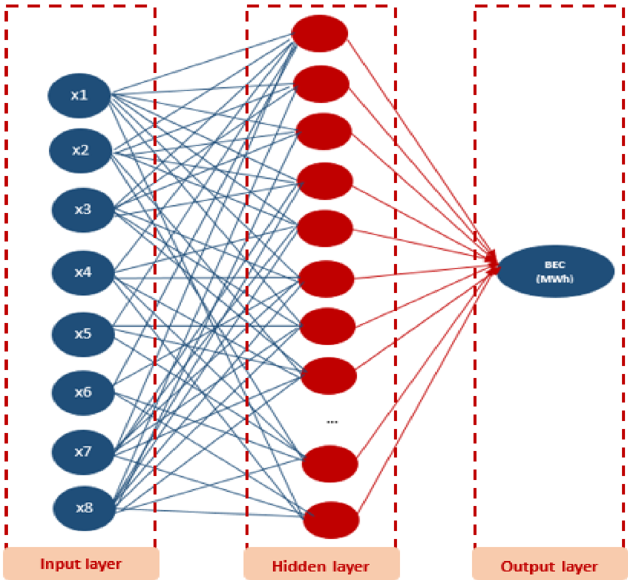


Fig. 1. The used Neural network architecture.

2.3 Random Forest (RF)

Random Forest (RF) represents a versatile and robust ensemble learning technique widely employed in predictive modeling[13]. This method leverages the strength of multiple decision trees to enhance prediction accuracy and mitigate overfitting. In the context of forecasting building energy consumption, Random Forest offers a formidable approach to capture intricate relationships among various input parameters[14]. The ensemble nature of Random Forest involves constructing multiple decision trees during training and aggregating their predictions to arrive at a final output. Mathematically, the prediction BEC for a specific instance can be expressed as:

$$BEC = \frac{1}{N}\sum_{i=1}^{N} f_i(X) \tag{2}$$

Where, N represents the number of trees in the forest, and $f_i(X)$ denotes the prediction of the ith decision tree based on the input features X.

The versatility and efficiency of Random Forest lie in its ability to handle complex, high-dimensional datasets, making it well-suited for modeling the multifaceted relationships inherent in building energy consumption.

2.4 Modified Whale Optimization Algorithm (MWOA)

The Modified Whale Optimization Algorithm (MWOA) is a nature-inspired optimization technique that enhances the original Whale Optimization Algorithm (WOA) by improving its exploration and exploitation capabilities[15], [16]. MWOA mimics the hunting behavior of humpback whales and operates as follows [17]:

1. Initialization: A population of candidate solutions X_i (where i=1,2,…,N) is randomly

initialized within the search space, where N is the number of solutions.

2. Fitness Evaluation: Each solution's performance is assessed using a fitness function $f(X_i)$aimed at minimizing energy consumption, represented as f(X) → min.
3. Encircling Prey: Candidate solutions update their positions based on the best-performing solution X_{best} using the following equation:

$$X_i^{new} = X_{best} - A \,.\, |C. X_{best} - X_i \,| \tag{3}$$

where A and C are the coefficients that control the exploration behavior.

4. Exploration: MWOA enhances exploration by introducing random jumps around the best solution, defined as:

$$X_i^{new} = X_{best} + A \,.\, |R.\text{rand}.\text{D} \,| \tag{4}$$

where R is a random factor, rand is a random number in [0,1], and D is the distance from the current position to X_{best}.

5. Exploitation: The algorithm balances global and local searches to refine solutions effectively by adjusting the coefficients A and C over iterations.
6. Termination: The process continues until a stopping criterion, such as a maximum number of iterations, is met.

By effectively identifying optimal input values for minimizing energy consumption, MWOA is used to enhance energy efficiency in buildings through its mathematical modeling of whale behaviors.

2.5 Evaluation Metrics

In the evaluation of building energy consumption (BEC) prediction models, various metrics are employed to assess the accuracy and performance of the models [18]. These metrics provide valuable insights into the predictive capabilities of the models and their ability to generalize to unseen data. The commonly used evaluation metrics include Root Mean Squared Error (RMSE), Coefficient of Variation (CV), Mean Absolute Percentage Error (MAPE), Mean Absolute Deviation (MAD), and the coefficient of determination (R^2) [19].

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(BEC_{i,true} - BEC_{i,predicted})^2}{n}} \tag{5}$$

$$CV\ (\%) = \frac{RMSE}{\overline{BEC}} \times 100\% \tag{6}$$

$$MAPE\ (\%) = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{BEC_{i,true} - BEC_{i,predicted}}{BEC_{true}}\right| * 100\% \tag{7}$$

$$MAD = \frac{1}{n}\sum_{i=1}^{n}\left|BEC_{i,true} - BEC_{i,predicted}\right| \tag{8}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(BEC_{i,true} - BEC_{i,predicted})}{\sum_{i=1}^{n}(BEC_{i,true} - \overline{BEC})} \tag{9}$$

Where,

- n is the number of observations;
- $BEC_{i,true}$is the actual (observed) value for the ith observation;
- $BEC_{i,predicted}$is the predicted value for the ith observation;
- $\overline{BEC}$ is the mean of the observed values

The application of these metrics enables researchers and practitioners to gauge the effectiveness of building energy consumption prediction models and make informed decisions about model selection and refinement.

In order to enhance the accuracy, the data were normalized to a standardized range between 0 and 1.

3 Results & discussion

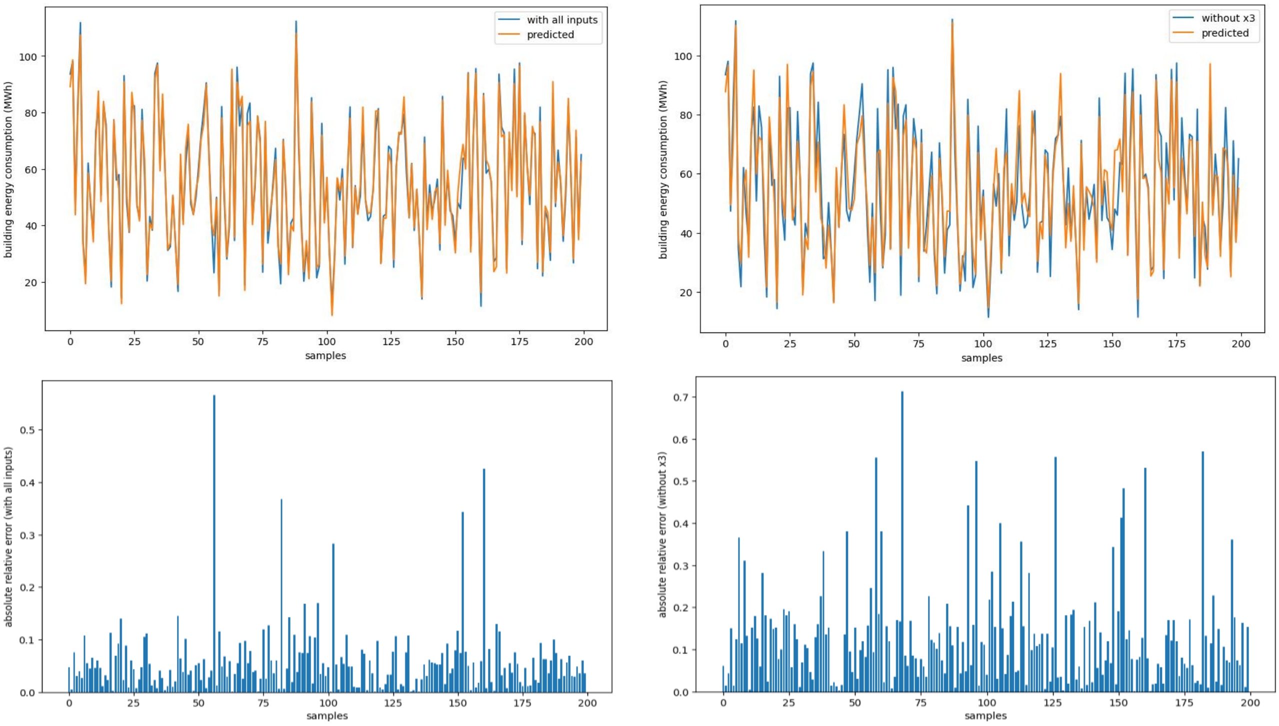
3.1 Comparative *prediction using RF and ANN*

In this study, a comprehensive examination involves eight input variables, specifically Building Size (x1), Floor Height (x2), Glazing Area (x3), Wall Area (x4), window-to-wall ratio (WWR) (x5), Window Glazing U-value (x6), Roof U-value (x7), and External Wall U-value (x8). The performance evaluation encompasses the consideration of all or various combinations of these input variables. Table 1 presents the following results of the RMSE, CV, MAPE, MAD and R2 for artificial neural networks at 1000 epochs, providing insights into how the exclusion or inclusion of specific input variables influences the predictive performance of the ANN model. The configuration of ANNs is: number of input variables: 10 (neurons in the hidden layer): 1 (neuron in the output layer).

Table 1. Performance metrics with different sets of input variables using ANN.

Input variables	RMSE	CV	MAPE	MAD	R^2
x1, x2, x3, x4, x5, x6, x7, x8	3.806	6.899	6.662	2.898	0.970
x3, x6, x8	4.333	7.853	7.461	3.192	0.961
All \ {x3}	7.895	14.310	13.026	6.341	0.873
All \ {x7}	4.363	7.908	7.721	3.547	0.961

The findings indicate that the model, which takes into account all eight input variables, consistently yields superior results across all performance metrics. Furthermore, a difference was observed between the networks that considered all input variables and those that utilized a reduced set of inputs. Figure 2 evaluates the metrics when excluding Glazing Area (x3). The model's performance significantly degrades, especially in terms of RMSE, CV, MAPE, and MAD, while R^2 decreases, indicating that the variable (x3) plays a crucial role in predicting energy consumption.



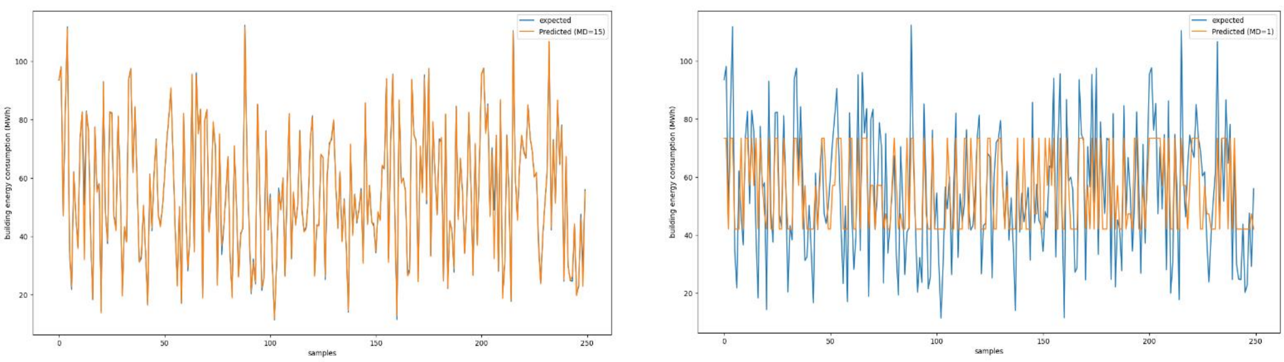
(a) (b)

Fig. 2. Results from ANN models developed (a) with all variables as inputs and (b) without Glazing Area (x3) value.

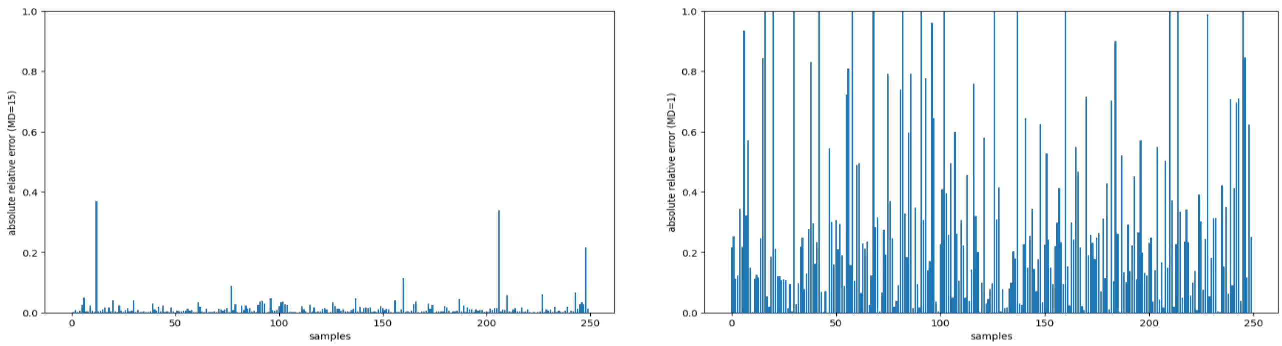
Figure 3 and Table 2 present a thorough evaluation of Random Forest models with varying maximum depths (Max depth). The obtained results in Table 2 demonstrate that deeper trees in the forest lead to improved performance. Specifically, a maximum depth of 1 results in relatively higher errors, characterized by larger values in RMSE, CV, MAPE, MAD, and a lower R^2. As the maximum depth increases, there is a notable enhancement in model performance, with lower errors and a higher R^2. However, the results indicate that the forest's performance starts to deteriorate beyond a maximum depth of 15. This suggests that increasing the maximum depth beyond a certain point may not significantly enhance model performance for the given dataset. Table 2 further illustrates these findings by showcasing the performance of forests with maximum depths. A forest with a maximum depth of 1 results in an under-fit model, characterized by higher values of RMSE (15.589), CV (28.256%), MAPE (30.893%), MAD (12.747), and a lower R^2 (0.505). Conversely, a forest with a maximum depth of 15 demonstrates better performance, featuring lower values of RMSE (1.392), CV (2.523%), MAPE (1.356%), MAD (0.571), and a higher R^2 (0.996).

Table 2. Random forest models with different max depth (MD) parameter.

RF models	RMSE	CV	MAPE	MAD	R^2
MD =1	15.589	28.256	30.893	12.747	0.505
MD = 5	4.239	7.683	7.703	3.338	0.963
MD = 10	1.537	2.786	1.838	0.782	0.995
MD =15	1.392	2.523	1.356	0.571	0.996
MD = 30	1.407	2.551	1.358	0.573	0.9959
MD =50	1.407	2.551	1.358	0.573	0.9959
MD =100	1.407	2.551	1.358	0.573	0.9959



* Corresponding author: nasima.elassri@ced.uca.ma



(a) (b)

Fig. 3. RF models with different maximum depths: (a) MD = 15, (b) MD = 1.

From table 3, the comparison between ANN and RF reveals that RF consistently outperforms ANN in predicting energy consumption, both in full and reduced models. RF achieves a lower RMSE (1.407 for the full model and 2.363 for the reduced model) compared to ANN (4.308 and 4.049, respectively), and it also shows better performance across CV, MAPE, MAD, and R^2. While ANN experiences a slight improvement in accuracy when using important variables, RF sees a decrease in performance with reduced variables. Despite this, RF remains more accurate and robust than ANN in all metrics, making it the superior choice for this dataset.

Table 3. Prediction errors comparison of full and reduced models.

Model	RMSE	CV	MAPE	MAD	R^2
ANN all	4.308	7.808	6.888	3.336	0.962
ANN important variables	4.049	7.340	6.882	2.969	0.966
RF all	1.407	2.551	1.358	0.573	0.9959
RF important variables	2.363	4.283	2.666	1.133	0.9886

In the field of energy consumption prediction for buildings, several studies have highlighted the effectiveness of AI and machine learning models. A previous study compared the performance of an Artificial Neural Network (ANN) with the Random Forest (RF) algorithm for predicting HVAC energy consumption in a hotel in Madrid, Spain[20]. The study concluded that the ANN slightly outperformed RF, with a lower RMSE (4.97 for ANN versus 6.10 for RF). However, it also emphasized the ease of use and parameter tuning of the RF model, especially when dealing with complex categorical variables and multidimensional data, which are common in buildings.

In our study, we observed similar results, with both ANN and RF demonstrating comparable predictive abilities. However, in contrast to the Madrid hotel study, our findings show that the RF model outperformed the ANN in terms of accuracy, with a lower RMSE (1.392 compared to 3.806 for ANN). This difference could be attributed to the nature of the dataset used in our study, which might align better with RF's ability to capture nonlinear and noisy relationships between the input features and energy consumption.

The choice between these two algorithms often depends on the nature of the problem, data availability, and desired performance. RF is commonly favored for its robustness to noisy data, ease of tuning, and ability to handle nonlinear relationships. On the other hand, ANN is highly powerful for modeling complex relationships but typically requires more data and expertise for proper configuration and optimization.

3.2 Optimization using a Modified Whale Optimization Algorithm (MWOA)

Optimizing design and operational parameters is essential for improving energy efficiency in smart buildings. These buildings, equipped with advanced management systems, require the efficient use of resources to minimize energy consumption while maintaining optimal indoor comfort. By adjusting factors such as building size, glazing area, or U-values of walls and windows, energy losses can be reduced while adhering to architectural constraints.

Optimization becomes even more crucial in the context of increasing regulations regarding building energy performance, aiming to reduce carbon footprints while improving energy efficiency.

The goal of this optimization process is to minimize the total energy consumption while adhering to the constraints related to the building design parameters, such as building size, floor height, glazing area, and the thermal performance of windows, roofs, and walls.

The Modified Whale Optimization Algorithm (MWOA) was chosen for this study due to its ability to find optimal solutions in complex, multidimensional search spaces. This algorithm, inspired by the hunting behavior of humpback whales, effectively balances exploration (global search across the solution space) and exploitation (local search around promising solutions). It is particularly well-suited for optimization problems with nonlinear relationships between variables, as seen in building energy efficiency models. The parameters used for the MWOA are summarized in Table 4 below.

Table 4. Parameters used for the modified whale optimization algorithm (MWOA).

Parameter	Description	Value
Population size	Number of candidate solutions (represented as whales)	20

Maximum iterations	Number of iterations to allow the algorithm to converge	20
Number of input variables (dim)	Number of input variables corresponding to the building design parameters	8

The optimization process using the MWOA involved running the algorithm for 20 iterations. The minimal energy consumption achieved during each iteration was also recorded. For instance, in the first iteration, the best solution found was [145.86, 2.8, 165.6353, 178.5094, 14.6358, 4.5807, 0.1823, 0.6899], resulting in a minimum energy consumption value of 33.6525. By the final iteration 20, the best solution adjusted to [145.86, 2.8, 19.6301, 19.6301, 10.0, 0.97, 0.13, 0.26], yielding a minimum energy consumption value of 10.2967.

The key findings from the optimization process demonstrate a substantial reduction in energy consumption, from an initial value of 33.6525 units to 10.2967 units after 20 iterations. This reduction was achieved through iterative adjustments of critical building design parameters, including the glazing area, wall area, window-to-wall ratio, and the U-values of windows, roof, and external walls, which improved the building's thermal performance. These results highlight the efficacy of the Modified Whale Optimization Algorithm (MWOA) in optimizing energy efficiency. For architects, engineers, and building designers, this approach provides a data-driven method for optimizing building designs to achieve lower energy consumption, enabling them to make more informed decisions regarding material selection, building geometry, and thermal insulation.

Conclusion

This study highlights the comparative effectiveness of Artificial Neural Networks (ANN) and Random Forest (RF) algorithms in predicting energy consumption in buildings, revealing that RF significantly outperforms ANN in terms of accuracy, as evidenced by its lower RMSE across various metrics. Both models showcased strong predictive capabilities, yet RF demonstrated greater robustness when handling complex datasets common in energy consumption scenarios, making it a more reliable choice for practitioners in the field.

Moreover, the implementation of the Modified Whale Optimization Algorithm (MWOA) for optimizing energy consumption further enhances the findings of this research. The MWOA successfully identified optimal design parameters, resulting in a notable reduction in energy consumption from an initial value of 33.6525 to a final minimum of 10.2967. These results underscore the potential of combining predictive modeling with advanced optimization techniques to improve energy efficiency in building design, ultimately contributing to more sustainable architectural practices. Future research should focus on refining these methodologies and exploring their applicability in diverse building contexts.

Acknowledgment

This study was supported by National Center for Scientific and Technical Research (CNRST) in Morocco under Grant number: 7UCA2021.

References

[1] H. P. Das *et al.*, « Machine Learning for Smart and Energy-Efficient Buildings », 27 novembre 2022, *arXiv*: arXiv:2211.14889. doi: 10.48550/arXiv.2211.14889.

[2] N. El Assri, M. A. Jallal, S. E. El Aoud, S. Chabaa, et A. Zeroual, « Synergistic Neural Network and Velocity Pausing Particle Swarm Optimization for Enhanced Residential Building Energy Efficiency: A Case Study in Kuwait », *Eng. Technol. Appl. Sci. Res.*, vol. 14, nº 5, p. 17507-17516, oct. 2024, doi: 10.48084/etasr.8278.

[3] S. Fathi, R. Srinivasan, A. Fenner, and S. Fathi, "Machine learning applications in urban building energy performance forecasting: A systematic review," Renewable and Sustainable Energy Reviews, vol. 133, 2020, Art. no. 110287. doi: 10.1016/j.rser.2020.110287.

[4] K. Amasyali et N. M. El-Gohary, « A review of data-driven building energy consumption prediction studies », *Renewable and Sustainable Energy Reviews*, vol. 81, p. 1192-1205, janv. 2018, doi: 10.1016/j.rser.2017.04.095.

[5] S. Ardabili, L. Abdolalizadeh, C. Mako, B. Torok, et A. Mosavi, « Systematic Review of Deep Learning and Machine Learning for Building Energy », *Front. Energy Res.*, vol. 10, mars 2022, doi: 10.3389/fenrg.2022.786027.

[6] T. Hong, Z. Wang, X. Luo, et W. Zhang, « State-of-the-art on research and applications of machine learning in the building life cycle », *Energy and Buildings*, vol. 212, p. 109831, avr. 2020, doi: 10.1016/j.enbuild.2020.109831.

[7] Digital twins based day-ahead integrated energy system scheduling under load and renewable energy uncertainties -

[8] Ibrahim, D. M. (2020). A dataset for residential buildings energy consumption with statistical and machine learning analysis [Data set]. GitHub. https://github.com/Dr-Dina-M-Ibrahim/A-dataset-for-residential-buildings-energy-consumption-with-statistical-and-machine-learning-analysi ».

[9] « IESVE. Integrated Environmental Solutions Virtual Environment (IESVE). httpswww.iesve.com. »

[10] Pittarello, M.; Scarpa, M.; Ruggeri, A.G.; Gabrielli, L.; Schibuola, L. Artificial Neural Networks to Optimize Zero Energy Building (ZEB) Projects from the Early Design Stages. Appl. Sci. 2021, 11, 5377. https://doi.org/10.3390/app11125377 ».

[11] R. Kumar, R. K. Aggarwal, et J. D. Sharma, « Energy analysis of a building using artificial neural network: A review », *Energy and Buildings*, vol. 65, p. 352-358, oct. 2013, doi: 10.1016/j.enbuild.2013.06.007.

[12] T. Khatib, A. Mohamed, K. Sopian, et M. Mahmoud, « Solar Energy Prediction for Malaysia Using Artificial Neural Networks », *International Journal of Photoenergy*, vol. 2012, nº 1, p. 419504, 2012, doi: 10.1155/2012/419504.

[13] Y. Mo, D. Zhao, et M. Syal, « Effective Features to Predict Residential Energy Consumption Using Machine Learning », p. 284-291, juin 2019, doi: 10.1061/9780784482445.036.

[14] L. N. Nyakundi, « PREDICTING ELECTRICITY CONSUMPTION IN NORWAY: A COMPARISON

OF MACHINE LEARNING MODELS », Master thesis, Norwegian University of Life Sciences, 2024.

[15] Y. Sun, X. Wang, Y. Chen, et Z. Liu, « A modified whale optimization algorithm for large-scale global optimization problems », *Expert Systems with Applications*, vol. 114, p. 563-577, déc. 2018, doi: 10.1016/j.eswa.2018.08.027.

[16] S. Mirjalili et A. Lewis, « The Whale Optimization Algorithm », *Advances in Engineering Software*, vol. 95, p. 51-67, mai 2016, doi: 10.1016/j.advengsoft.2016.01.008.

[17] L. Abualigah *et al.*, « 8 - Whale optimization algorithm: analysis and full survey », in *Metaheuristic Optimization Algorithms*, L. Abualigah, Éd., Morgan Kaufmann, 2024, p. 105-115. doi: 10.1016/B978-0-443-13925-3.00015-7.

[18] R. Wang, S. Lu, et Q. Li, « Multi-criteria comprehensive study on predictive algorithm of hourly heating energy consumption for residential buildings », *Sustainable Cities and Society*, vol. 49, p. 101623, août 2019, doi: 10.1016/j.scs.2019.101623.

[19] A. A. Al-Shargabi, A. Almhafdy, D. M. Ibrahim, M. Alghieth, et F. Chiclana, « Buildings' energy consumption prediction models based on buildings' characteristics: Research trends, taxonomy, and performance measures », *Journal of Building Engineering*, vol. 54, p. 104577, août 2022, doi: 10.1016/j.jobe.2022.104577.

[20] M. W. Ahmad, M. Mourshed, et Y. Rezgui, « Trees vs Neurons: Comparison between random forest and ANN for high-resolution prediction of building energy consumption », *Energy and Buildings*, vol. 147, p. 77-89, juill. 2017, doi: 10.1016/j.enbuild.2017.04.038.