International Journal of Multiphysics
Volume 18, No. 4, 2024
ISSN: 1750-9548

Development of the Artificial Intelligence Models for Estimating the Building Heating Performance with the Implementing of the Design-Architecture Parameters

Amir Hossein Hojati*[1], Ali Ashghar Amir Kardoost[2], Davood sedaghat shayegan[3]

[1]Department of civil engineering, Roudehen Branch, Islamic Azad University, Roudehen, Iran

[2]Department of civil engineering, Roudehen Branch, Islamic Azad University, Roudehen, Iran

[3]Department of civil engineering, Roudehen Branch, Islamic Azad University, Roudehen, Iran

*Corresponding author: AmirHossein Hojati, a.hojati@iau.ir

Abstract

This study investigates the application of machine learning techniques, specifically Multivariate Adaptive Regression Splines (MARS) and Model Trees (MT), in estimating building heating performance. Accurate heating performance estimation is crucial for improving energy efficiency, reducing operational costs, and achieving sustainability goals. By leveraging real-world datasets that incorporate variables such as weather conditions, building characteristics, and energy consumption patterns, the study aims to evaluate the effectiveness of these two advanced modeling approaches.

The results indicate that both MARS and MT models provide reliable and accurate predictions of building heating performance. However, the MARS model (RMSE=0.247, R=0.993) demonstrates superior performance compared to the MT approach (RMSE=9, R=0.947). The MARS model's flexibility in capturing nonlinear relationships and interactions among variables contributes to its enhanced predictive accuracy. In contrast, the MT model, which relies on classification-based and formula-driven methods, exhibits limitations in handling complex variable interactions.

This study highlights the advantages of using MARS for heating performance estimation, emphasizing its potential as a robust and adaptable tool for energy management in buildings. The findings underscore the importance of selecting appropriate machine learning methods tailored to specific predictive tasks, ultimately advancing the state-of-the-art in energy modeling and building performance optimization.

Keywords: Artificial intelligence models, building heating performance, architecture parameters, MARS method, MT method

1. Introduction

Artificial intelligence (AI) models are rapidly transforming the way we analyze and optimize building performance. In the context of heating systems, the integration of AI offers significant advantages over traditional approaches [1]. One of the most compelling benefits of AI is its ability to process vast amounts of data with remarkable speed and accuracy [2]. Modern buildings generate substantial datasets from sensors, energy consumption records, weather patterns, and occupant behavior [3]. AI models can seamlessly analyze these complex datasets to identify patterns, make predictions, and provide actionable insights. AI-driven approaches also excel in precision and adaptability [4]. Unlike conventional methods that rely on static models or generalized

International Journal of Multiphysics
Volume 18, No. 4, 2024
ISSN: 1750-9548

assumptions, AI can adapt to specific building conditions and learn from real-time data [5]. This dynamic learning capability enables AI to deliver highly accurate heating performance estimations that account for variables such as fluctuating weather conditions, changes in occupancy, and the thermal properties of building materials. Consequently, building managers and engineers can make informed decisions to enhance energy efficiency, reduce operational costs, and minimize environmental impact [6-8].

Another key advantage of AI is its capacity for optimization. Machine learning algorithms can be employed to predict energy demand, optimize system settings, and even propose maintenance schedules to prevent system failures [9]. These proactive measures not only ensure consistent heating performance but also extend the lifespan of heating equipment. Additionally, the ability to simulate various scenarios through AI-powered digital twins allows stakeholders to explore energy-saving strategies without disrupting actual operations [10]. Fig. 1 presented the visualization includes architectural parameters such as building orientation, glass facades, skylights, and thermal walls, all annotated to illustrate their impact on heating load.

This paper proposed an artificial methods (AI) named multivariate adaptive regression splines (MARS) and model tree (MT) techniques utilizing several functions to model the energy performance of residential structures. The MARS and model tree can autonomously identify the most significant predictor input variables in the model, establish the model structure, and determine the unknown parameters of the regression equation, all while optimizing for accuracy and complexity. It is thought that loads exhibit a linear and non-linear relationship concerning functions. Consequently, an equation is formulated using the MARS and MT approaches to forecast the heating load.

2. Materials and methods

2.1. MARS method

Friedman originally elucidated the use of MARS, a prominent machine learning technique, to identify a function that delineates the link between several predictors (Xs) and an output (Y) [11]. Nonparametric data modeling is a machine learning approach that does not need any assumptions on the distribution of the relevant variables. MARS is particularly advantageous when the mapping function varies for each subset of the acquired data. Predictive models utilizing MARS have been effectively created for intricate engineering datasets previously. MARS divides the compressive strength learning area into smaller segments throughout the learning phase [12]. (Refer to Fig. 2).

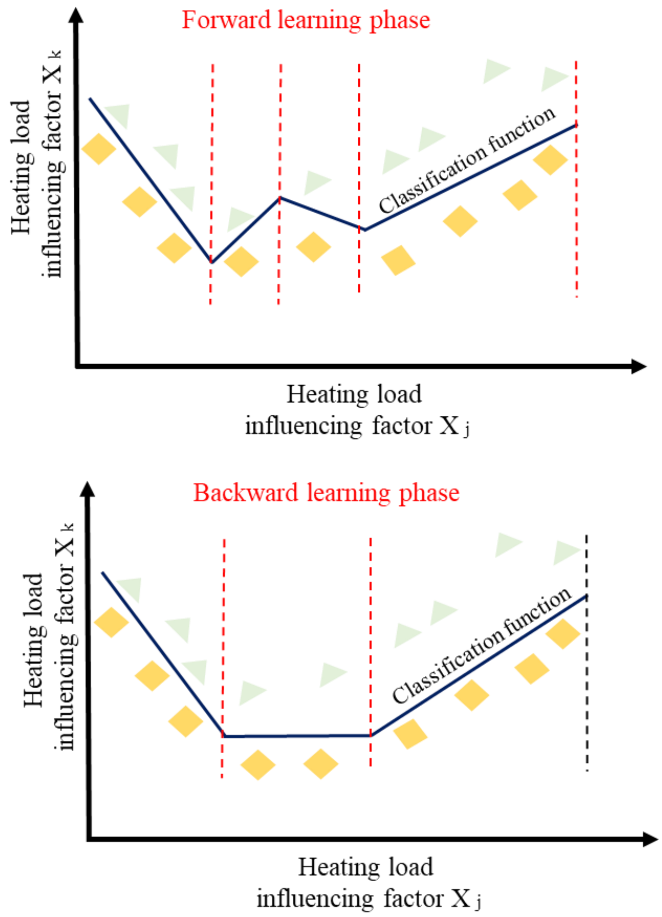

Fig. 2: MARS training phases

MARS constructs a linear classification model for each zone to achieve optimal data matching [13]. Figure 2 illustrates that MARS constructs a data-driven piecewise linear function to represent a global model. This unique characteristic has enabled MARS to be utilized across several sectors. An additional crucial element of the final prediction model's learning phase is the generation of Basis Functions (BFs) throughout this process. These BFs articulated the correlation between experimental variables and compressive strength. A standard BF is represented in the subsequent equations.

$$b_m(x) = \max(0, C - x) \; or \; b_m(x) = \max(0, X - C) \tag{1}$$

In this context, b_m is a BF, x signifies a predictor of compressive strength, and C is a threshold parameter identified during the learning phase. The ultimate classification model created by MARS for predicting heating load may be encapsulated as follows:

$$f(x) = sign(\propto_0 + \sum_{m=1}^{k} \propto_m b_m(x)) \tag{2}$$

2.2. MT method

Quinlan introduced the MT model, which employs a binary decision tree combined with a sequence of linear regression functions at the terminal (leaf) nodes. The standard deviation of class values at a node serves as an indicator of the node's error level in the first phase. The anticipated decrease in error for each feature is then ascertained [14]. The term "*Standard Deviation Reduction*" (SDR) is utilized to denote the reduction of mistakes.

$$SDR = sd(T) - \sum \frac{|T_i|}{T} sd(T_i) \tag{3}$$

sd signifies the standard deviation, while Ti and T denote a subset of examples corresponding to the i_{th} potential outcome and a collection of cases that reach the node, respectively. The standard deviation of a child node is less than that of its parent node due to the splitting process. The optimal split is determined after evaluating all possible splits to minimize expected error (Refer to Fig. 3).

[15].

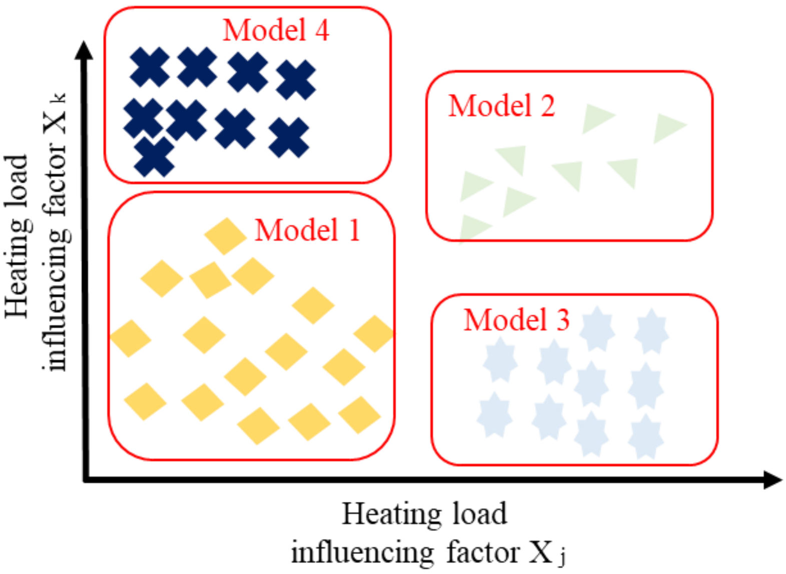

Fig. 3: MT data classification for regression analysis

2.3. Modeling database

A database of 768 building samples, derived from 12 building shapes with a volume of 771.75 m^3, utilizing Autodesk Ecotect Analysis, has been examined by Tsanas and Xifara [16]. Each building sample is defined by eight predictor input parameters: relative compactness (RC), surface area (SA), wall area (WA), roof area (RA), overall height (OH), orientation (OR), glazing area (GA), and glazing area distribution (GAD). Additionally, two outputs for heating load (HL) have been documented for each building sample throughout the AI simulation procedure.

In this database, 12 buildings with a total volume of 771.75 cubic meters were examined, and measurement data were extracted to perform the modeling process. The levels of reflective glass in this research are 0-40%, and also the distribution of these levels in the direction of the building with 5 placement scenarios, including 25% in 4 directions of the building, 55% in the south and the rest equal to 15% in other directions, 55% in the north and the rest in an equal amount of 15% in other directions, 55% in the east and the rest in an equal amount of 15% in other directions, and 55% in the west and the rest in an equal amount of 15% in the directions, has been Therefore, according to the relationship number (4), the heating load in the building is a function of this relationship, and the models of this research were formed using this relationship of predictor variables.

$$\mathrm{HL} = f\ (RC, SA, WA, RA, OH, O, GA, GAD) \quad (4)$$

2.4. Performance metrics

This article employed the assessment criteria (Eqs. (5-7)), which include the correlation coefficient (r), Root Mean Square Error (RMSE), and Mean Absolute Error (MAE) [17].

$$R = \frac{\sum_{i=1}^{M}(O_i-\overline{O}).(P_i-\overline{P})}{\sqrt{\sum_{i=1}^{M}(O_i-\overline{O})^2 \sum_{i=1}^{M}(P_i-\overline{P})^2}} \quad (5)$$

$$RMSE = \frac{\sum_{i=1}^{M}(P_i-O_i)^2}{\sum_{i=1}^{M}(\overline{O}-O_i)^2} \quad (6)$$

$$MAE = \frac{\sum_{i=1}^{M}|P_i-O_i|}{M} \quad (7)$$

3. Result and discussion

3.1. MARS model development

This study evaluates the developed models using the MARS method. A total of 15 models were designed and tested by varying the parameters of this method. Based in the Fig. 4, among these models, Model 8 was identified as the best-performing model, as it achieved the lowest error rate along with the highest accuracy and solution convergence. In this research, the penalty parameter was initially set to 1 to 8. Ultimately, the optimal penalty parameter value, C_{best}=2, was determined through the process of error minimization using the least squares error method. These findings highlight the high capability of the MARS method in optimizing models and reducing prediction errors. Table 1 reported the MARS training procedure for the hyper-parameters tuning.

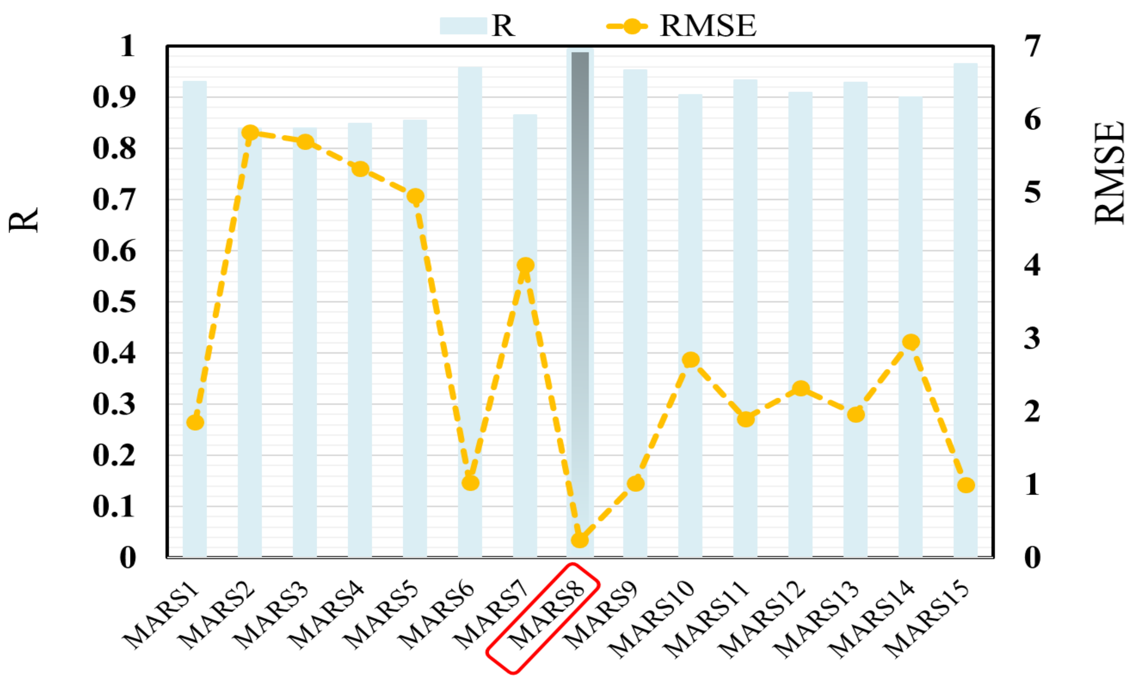

Fig. 4: Evaluation of the MARS models development

Table 1: MARS training procedure

Output	HL
Function	Piece-wise linear
Max function	30
Max interaction	3

The MARS model was modified to optimize the prediction of the heating load of the building performance with a value of 18, while the greatest degree of interaction of the BFs was established at the second order. Furthermore, as suggested by Jekabsons, the penalty factor c was determined by trial and error within the range of 1 to 8, with the optimal value discovered as c_{best}=2. Ultimately, the subsequent non-linear equation was put out for the modeling of the CS. The functionalities for each BF were displayed in Table 2.

Table 2: MARS functions for prediction of the HL.

Func. NO.	Functions	Func. NO.	Functions
BF1	max(0, x4 -122.5)	BF10	BF1 × max(0,0.1 -x7)
BF2	max(0,122.5 -x4)	BF11	BF7 × max(0, x3 -343)
BF3	max(0, x7 -0.1)	BF12	BF7 × max(0,343 -x3)
BF4	max(0,0.1 -x7)	BF13	BF3 × max(0, x4 -147)
BF5	max(0, x2 -637)	BF14	BF3 × max(0,147 -x4)
BF6	max(0,637 -x2)	BF15	BF7 × max(0, x1 -0.71)
BF7	BF1 × max(0, x2 -612.5)	BF16	BF7 × max(0,0.71 -x1)
BF8	BF1 × max(0,612.5 -x2)	BF17	BF3 × max(0, x8 -4)
BF9	BF7 × max(0, x1 -0.69)	BF18	BF3 × max(0,4 -x8)

Note: X1-X8 refers to relative compactness (RC), surface area (SA), wall area (WA), roof area (RA), overall height (OH), orientation (OR), glazing area (GA), and glazing area distribution (GAD) respectively.

$$HL = 45.816 - 0.682 \times BF1 + 0.572 \times BF2 + 18.285 \times BF3 - 82.045 \times BF4 - 0.533 \times BF5 - 0.23255 \times BF6 + 0.0087 * BF7 + 0.014 \times BF8 + 0.019 \times BF9 + 0.396 \times BF10 + 9.8753e - 06 \times BF11 - 2.3198e - 05 \times BF12 - 0.09 \times BF13 + 0.24 \times BF14 + 0.02 \times BF15 - 0.026 \times BF16 - 1.080 \times BF17 + 0.305 \times BF18 \quad (8)$$

3.2. MT model development

After separating the data into two phases, training and testing, the MT model was fitted on the training data. At this stage, the data were classified into groups, and a linear multivariate regression model (LM) was presented for each group at the following LMs. The regression equations and the governing rules for estimating the heating load are presented.

LM num: 1

HL = 52.7347 × Relative compactness + 0.0064 × Wall area + 12.8684 × Glazing area - 21.759

LM num: 2

HL = 52.7347 × Relative compactness + 0.0064 × Wall area + 8.4506 × Glazing area - 21.1103

International Journal of Multiphysics
Volume 18, No. 4, 2024
ISSN: 1750-9548

LM num: 3

HL = 52.7347 × Relative compactness + 0.0064 × Wall area + 7.4641 × Glazing area - 20.4313

LM num: 4

HL = 44.9747 × Relative compactness + 0.0064 × Wall area + 15.4286 × Glazing area - 0.0295 × Glazing area distribution - 13.7995

LM num: 5

HL = 44.9747 × Relative compactness + 0.0064 × Wall area + 9.3712 × Glazing area - 0.1633 × Glazing area distribution - 12.8929

LM num: 6

HL = -10.0707 × Relative compactness + 0.0064 × Wall area+ 20.6848 × Glazing are+ 17.1893

LM num: 7

HL = -10.6277 × Relative compactness + 0.0064 × Wall area + 10.5852 × Glazing area - 0.1234 × Glazing area distribution + 18.7905

LM num: 8

HL = 105.5609 × Relative compactness + 0.0119 × Wall area + 17.9549 × Glazing area - 52.6766

LM num: 9

HL = 42.8508 × Relative compactness + 0.0225 × Wall area + 33.2161 × Glazing area - 18.6578

LM num: 10

HL = 4.6956 × Relative compactness + 0.0225 × Wall area + 15.024 × Glazing area + 0.106 × Glazing area distribution + 14.3216

LM num: 11

HL = 9.1387 × Relative compactness + 0.0371 × Wall area + 17.3941 × Glazing area + 5.6521

LM num: 12

HL = 6.1985 × Relative compactness + 0.1301 × Wall area + 0.2605 × Orientation + 18.1584 × Glazing area - 18.8164

The HL variable divides the two-branched tree pattern associated with cooling load estimation into two categories. Next, each category undergoes another binary branching process. We repeatedly branch each node until we reach the final node (leaf), where the sum of the squared deviations from the average of the data is approximately zero. After pruning the extra branches, the optimal tree emerges. Finally, we built a tree model for the output variable based on the relevant equations and calculated the values predicted by the model using the training and test data. The proposed MT technique includes 7 input parameters and 1 output parameter (heating load), which was developed using 12 conditional rules in the form of linear equations. Therefore, Fig. 5 displays the outline diagram of the MT method's tree formation in the form of rules for heating load estimation.

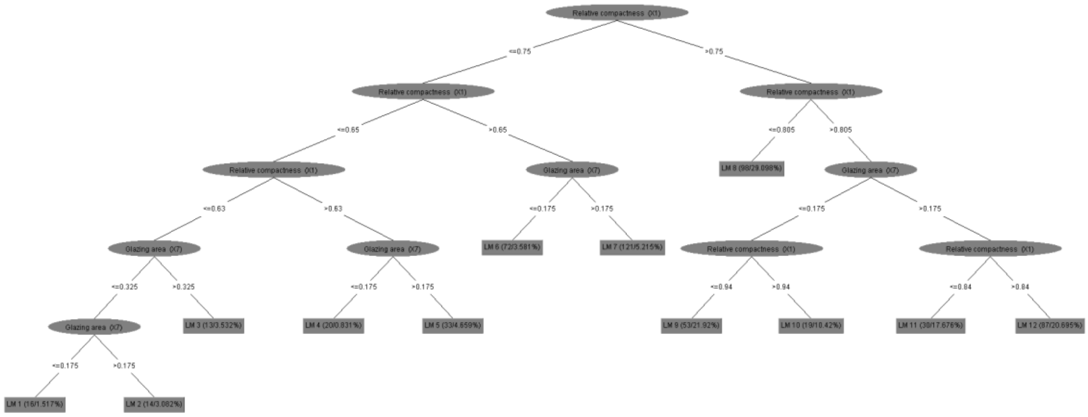

Fig. 5: MT structure for HL estimation

Based on the tree structure, the branching variable to provide LM1 is the HL parameter, and its corresponding values are 0.75, 0.65, and 0.63. In addition, the value of the GA variable was considered to be 0.325 and 0.175 to produce the LM1. In LM2, the branching parameter in the tree structure was GA, with a value of 0.175. Using all variables as input variables, the model showed that selecting these parameters based on architectural features had a significant impact on the structure's heating load. It shows how to develop the tree model and settings of effective parameters in Weka 3.7 software. Additionally, the LMs also display the models that were developed using the decision tree method. To select the optimal model, 50 models have been developed in this section, and 20 models are presented in Fig. 6. The MT4 model with a correlation coefficient of 0.979 has performed better than other models.

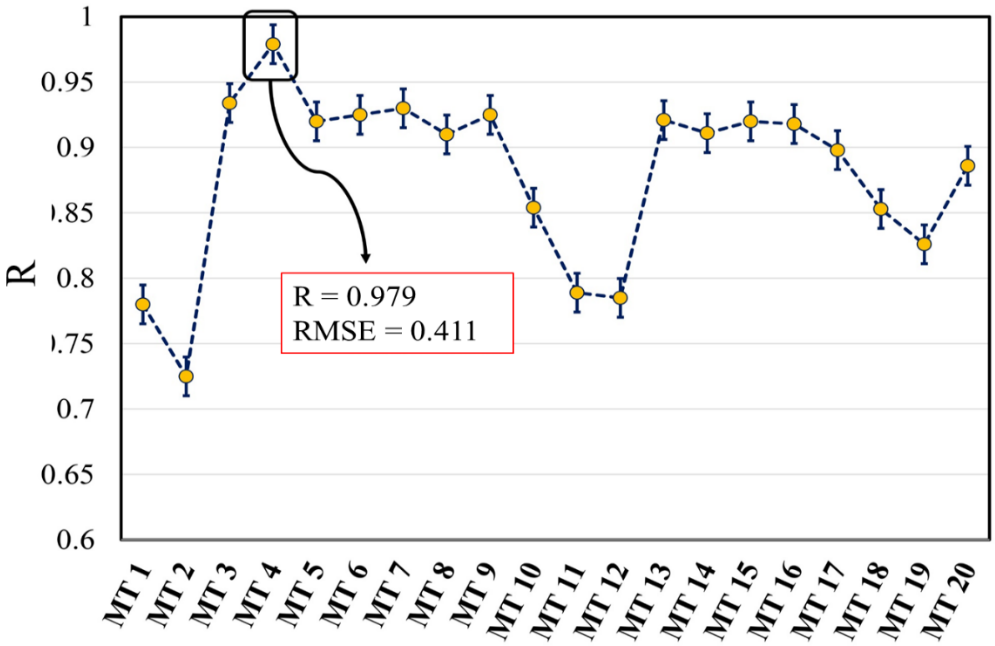

Fig. 6: Evaluation of the MT models development

3.3. Performance assessment

As shown in Table 3, the correlation coefficient in the training phase for the March model and the tree model is 0.995 and 0.979, respectively. is Also, the value of the RMSE statistical index for the proposed model for the March model and the tree model has been obtained as 0.216 and 0.315, respectively. The results showed that the models had excellent results in the training phase of convergence conditions (R≥0.8). Also, the MAE statistical index in the testing phase for March (0.376) was better than the other two proposed models and reported lower error values. So, the numbers show that MARS's non-parametric and non-linear model that can come up with values works well and is very accurate during the training phase compared to other models. The use of a non-linear and non-parametric approach in model training played an essential role in model accuracy.

Table 3: MARS functions for prediction of the HL.

Performance	Models	Metrics		
		r	RMSE	MAE
Training	MARS	0.995	0.216	0.338
	MT	0.979	0.315	0.405
Testing	MARS	0.993	0.247	0.376
	MT	0.956	0.749	0.607

In this research, Figs. 7 and 8 illustrated the observed and predicted HL values by the proposed model for the training stages. In the training phase, the MARS model has less error compared to the MT model in predicting the HL of the building. In a qualitative comparison, most of the HL values are concentrated on the regression line (ideal line), and only a small number of these points are outside the concentration area. In the evaluation of the models, a deviation of more than 20% has been observed in the prediction of the estimated points in the range of 20-30% (over estimation) in the models and 35-45% in the models (under estimation). In the MARS model, when the estimation values increase, they typically align with the software's measurement values. Overall, this study's methods have undergone proper training for evaluation. During the testing phase, as shown in Fig. 8, the models accurately predicted the HL values. Predicting these values is mostly less than 20%, and the correlation between actual and predicted laboratory values is at least 90%. Also, the percentage of absolute error in the testing phase has shown that the MARS model has decreased by 4 and 5% compared to the neural network and regression model. The red line in the figure indicates that the indicated values overestimate or underestimate data with weaker prediction (less fitting). In summing up the evaluations, this error rate has been reduced in the MARS model, and a more efficient and accurate model has been presented.

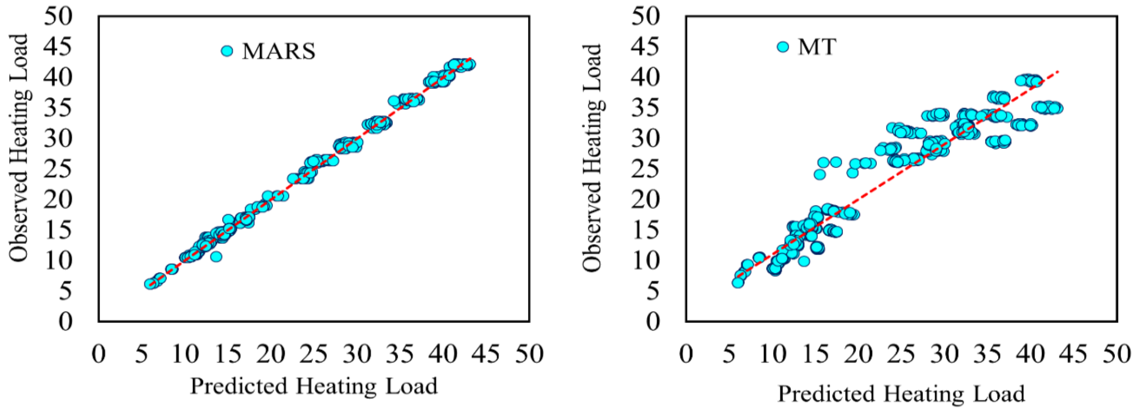

Fig. 7: Scatterplots of the training performance of the HL models

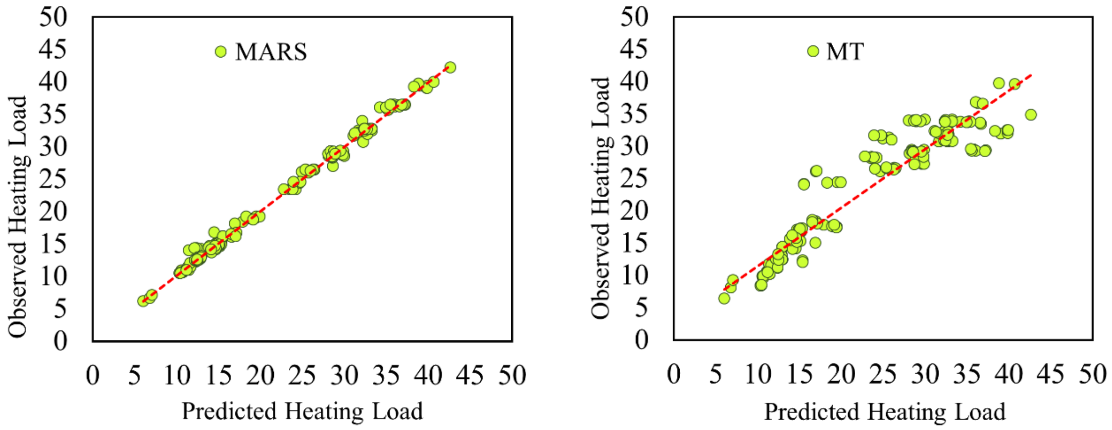

Fig. 8: Scatterplots of the testing performance of the HL models

4. Summary and Conclusion

Nowadays, life requires energy at a rapid pace, a time when energy resources are limited. In recent years, this need has significantly increased globally. The importance of using energy to achieve thermal comfort in buildings has always been the focus of researchers. Among the cases of thermal comfort in the cold season is the heating load of the building, which gives satisfaction to people in terms of the feeling of heat, air conditioning and proper heat transfer due to the temperature difference. In the cold season, the temperature outside is lower than the temperature of the comfort conditions inside. This temperature difference causes heat to be transferred from inside the space to outside it in different ways. We refer to this heat transfer as a "heating load" because, in order to maintain the comfort temperature of our desired space, we must add heat to the space using heat-generating equipment (such as radiators), equal to the amount of heat that has already left the space.

The heating load of a building is a vital determinant of its energy performance, affecting both comfort and energy efficiency. It denotes the quantity of thermal energy necessary to sustain a building's interior at a specified temperature during colder seasons. Principal elements influencing heating demand encompass: The design, dimensions, and orientation of the building significantly influence heat retention and dissipation. South-facing windows can capture passive solar energy. In this attempt, AI models used to evaluate the simulation and analysis of building performance. To this end, MARS and MT method were developed to fast and accurate formula-based models to estimate the heating load of the buildings. The following result were explored for this paper:

- This study's statistical analysis approach offers critical insights into the subject at hand, which is often overlooked in the literature within this area. The density and scatter plots provide substantial evidence that non-linear approaches are unsuitable for the data in this application.
- The results of this study align with the AI literature that robustly supports the implementation of the MARS and MT in intricate scenarios.
- The MARS model (RMSE=0.247, R=0.993) does a better performance of predicting heating loads than the MT (RMSE=9, R=0.947), as presented by the performance measures used during the training phase.
- The MAE statistics indicator for MARS (0.376) indicates a reduced average error. Based on the statistical indicators, the models created from the MARS formula perform better and more accurately during the training phase compared to other models. This study indicates that non-linear models yield greater accuracy in training compared to linear correlations.

Refrences

1- Merabet, G. H., Essaaidi, M., Haddou, M. B., Qolomany, B., Qadir, J., Anan, M., ... & 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, 110969.

2- He, Z., Guo, W., & Zhang, P. (2022). Performance prediction, optimal design and operational control of thermal energy storage using artificial intelligence methods. Renewable and Sustainable Energy Reviews, 156, 111977.

3- Ngarambe, J., Yun, G. Y., & Santamouris, M. (2020). The use of artificial intelligence (AI) methods in the prediction of thermal comfort in buildings: Energy implications of AI-based thermal comfort controls. *Energy and Buildings*, *211*, 109807.

4- Roy, Sanjiban Sekhar, Pijush Samui, Ishan Nagtode, Hemant Jain, Vishal Shivaramakrishnan, and Behnam Mohammadi-Ivatloo. "Forecasting heating and cooling loads of buildings: A comparative performance analysis." *Journal of Ambient Intelligence and Humanized Computing* 11 (2020): 1253-1264.

5- Makhsusy, M., Amirkardoust, A., & Sedaghat Shayegan, D. (2024). Integrating Building Information Modelling (BIM) and Artificial Intelligence (AI) Technologies to Improve the Planning and Control Process of Construction Projects. Civil and Project.

6- Sedaghat Shayegan, D., Fazlikhani, E., & Amirkardoust, Ů. A. (2024). Development of innovative intelligent model to estimate the strength properties of hemp bio composite mixture using the combination of water cycle algorithm and MARS method. *Amirkabir Journal of Civil Engineering*.

7- Sedaghat Shayegan, D., & Roosta, M. R. (2024). Numerical Investigation of Block Form Effect in Amount of Dissipated Energy of Baffled Apron Spillway Drop by Flow3D. *Journal of Civil and Environmental Engineering*, *54*(114), 56-62.

8- Doulabi, R. Z. (2024). The Identification and Prioritization of Critical Success Factors in Healthcare Projects Through the Application of the Analytic Hierarchy Process (AHP). *Power System Technology*, *48*(1), 1643-1653.

9- Saberi, A., Ahmadi, H., Shayegan, D. S., & Amirkardoust, A. (2023). A study on the impact of water storage and regulation projects on sustainable development goals using mouth brooding fish (MBF) algorithm. *Sustainable Water Resources Management*, *9*(6), 196.

10- Saberi, A. A., Ahmadi, H., Shayegan, D. S., & Amirkardoust, A. (2023). Prediction of electricity consumption using three meta-heuristic algorithms. *Int. J. Optim. Civil Eng*, *13*(1), 111-125.

11- Taheri Amiri, M. J., Ashrafian, A., Haghighi, F. R., & Javaheri Barforooshi, M. (2019). Prediction of the compressive strength of self-compacting concrete containing rice husk ash using data driven models. *Modares Civil Engineering journal*, *19*(1), 209-221.

12- Ashrafian, A., Taheri Amiri, M. J., & Haghighi, F. (2019). Modeling the slump flow of self-compacting concrete incorporating Metakaolin using soft computing techniques. *Journal of Structural and Construction Engineering*, *6*(2), 5-20.

13- Ashrafian, A., Behnood, A., Golafshani, E. M., Panahi, E., & Berenjian, J. (2024). Toward presenting an ensemble meta-model for evaluation of pozzolanic mixtures incorporating industrial by-products. *Structural Concrete*, *25*(2), 1305-1323.

14- Rezaie-balf, M., Naganna, S. R., Ghaemi, A., & Deka, P. C. (2017). Wavelet coupled MARS and M5 Model Tree approaches for groundwater level forecasting. *Journal of hydrology*, *553*, 356-373.

15- Ghaemi, Alireza, Mohammad Rezaie-Balf, Jan Adamowski, Ozgur Kisi, and John Quilty. "On the applicability of maximum overlap discrete wavelet transform integrated with MARS and M5 model tree for monthly pan evaporation prediction." *Agricultural and Forest Meteorology* 278 (2019): 107647.

16- Tsanas, A., & Xifara, A. (2012). Accurate quantitative estimation of energy performance of residential buildings using statistical machine learning tools. *Energy and buildings*, *49*, 560-567.

17- Asuero, A. G., Sayago, A., & González, A. G. (2006). The correlation coefficient: An overview. *Critical reviews in analytical chemistry*, *36*(1), 41-59.