Reliable Hybridization Approach for Estimation of The Heating Load of Residential Buildings

Huanhuan Li
School of Civil Engineering, Xi'an Traffic Engineering Institute, Xi'an Shaanxi 710300, China

***Abstract*—In recent times, the world's growing population, coupled with its ever-increasing energy demands, has led to a significant rise in the consumption of fossil fuels. Consequently, this surge in fossil fuel usage has exacerbated the threat of global warming. Building energy consumption represents a significant portion of global energy usage. Accurately determining the energy consumption of buildings is crucial for effective energy management and preventing excessive usage. In pursuit of this goal, this study introduces a novel and robust machine learning (ML) method based on the K-nearest Neighbor (KNN) algorithm for predicting the heating load of residential buildings. While the KNN model demonstrates satisfactory performance in predicting heating loads, for the attainment of optimal results and accuracy, two novel optimizers, the Snake Optimizer (SO) and the Black Widow Optimizer (BWO), have been incorporated into the hybridization of the KNN model. The results highlight the effectiveness of KNSO in predicting heating load, as evidenced by its impressive R^2 value of 0.986 and the low RMSE value of 1.231. This breakthrough contributes significantly to the ever-pressing pursuit of energy efficiency in the built environment and its pivotal role in addressing global environmental challenges.**

***Keywords*—Heating load; residential buildings; k-nearest neighbor; snake optimizer; black widow optimizers**

I. Introduction

Buildings' energy consumption has substantial consequences for both the economic and the environmental health of a country [1]. The buildings sector is responsible for roughly 40% of the total energy consumption [2]. For instance, in the United States, the building sector represents 39% of the total energy consumption, while residential buildings in the European Union account for approximately 40% of the energy consumption within the building sector [3]. This significant energy usage in the building sector has positioned carbon emissions (CO_2) as a primary driver of climate change, global warming, and air pollution [4], [5]. Consequently, many architects, researchers, and engineers have taken up the task of investigating models that centre on building envelopes and design features with the goal of reducing the negative effects associated with energy consumption in buildings [6], [7].

The primary elements within a building's envelope and design features that influence its energy consumption encompass the U-value of the envelope (comprising materials such as wall materials, roof materials, and glazing properties), the window-to-wall ratio, and the orientation of the façade [8], [9]. Hence, to promote sustainability and mitigate negative impacts on both the natural and built environments, it is essential to consider and comprehend these variables in terms of their thermal efficiency and energy consumption [10]. Energy prediction tools are essential for enabling well-informed decision-making aimed at reducing energy consumption in buildings [11]. These tools possess the capacity to evaluate a broad spectrum of building designs and strategies, thereby enhancing energy demand and management [12]. It is crucial to acknowledge, however, that factors other than a building's envelope and characteristics impact energy consumption, such as external weather conditions, occupant behaviour, the adoption of technologies, and equipment [13]. The task of energy prediction is a complex research challenge.

Nevertheless, progress has been achieved in the quest for sustainable buildings concerning energy demand [14]. However, energy forecasting continues to fall behind the rapid urbanization and advancements in building design and features [15]. The energy efficiency of buildings has garnered substantial research attention, with numerous studies concentrating on prediction models based on data analysis [16], [17]. AI models show substantial potential for both forecasting and enhancing building energy usage [18]. These models leverage historical data, real-time sensor information, and ML algorithms to produce precise predictions, offering valuable insights for effective energy management. Over time, the field of predicting energy consumption has witnessed notable progress. Researchers and industry experts have devised a range of methods and strategies to forecast energy utilization consistently [19]. Kim and Cho [20] introduced a neural network in their study, which combined the characteristics of Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) architectures, tailoring them for accurate forecasting of residential energy usage. The fusion of CNN and LSTM skillfully harnessed spatial and temporal features, enabling the capture of complex energy consumption patterns with great proficiency.

Experimental results showcased the exceptional accuracy of the CNN-LSTM approach, especially in the context of electric energy usage, surpassing conventional forecasting techniques. Roy et al. [1] introduced a tailor-made Deep Neural Network (DNN) model that was specifically crafted for forecasting heating and cooling loads in residential buildings. They proceeded to perform a comparative assessment, pitting the DNN model against the gradient-boosted machine (GBM), Gaussian process regression (GPR), and minimax probability models, particularly machine regression (MPMR). The findings revealed that both the DNN and GPR models demonstrated the most substantial variance accounted for (VAF) when it came to predicting both heating and cooling loads. In a study presented by Moradzadeh et al. [15], SVR and MLP models were employed to predict Cooling and Heating

Loads. The MLP technique yielded impressive outcomes, achieving the highest R-value of 0.9993 for Heating Load prediction. In contrast, for the Cooling Load prediction, the SVR method excelled, attaining the highest R-value of 0.9878.

In this research, a fresh ML approach is presented with the objective of attaining accurate and optimal predictive outcomes. The hybridization technique employed in this study is meticulously designed to boost the effectiveness of KNN models, guaranteeing the generation of dependable results. By combining two advanced and efficient optimization methods, the creation of these novel hybrid models surpassed conventional approaches, representing a notable advancement. A thorough assessment of these models was carried out, covering both their individual and hybrid setups, in order to guarantee a fair evaluation of their capabilities. In order to ascertain the robustness of the outcomes, the evaluation of model effectiveness included well-recognized metrics like R^2 and RMSE. Furthermore, the purposeful choice of two separate optimizers, specifically the Snake Optimizer (SO) and the Black Widow Optimizer (BWO), for building the hybrid models was guided by the objective of harnessing the distinct advantages of each optimizer, with the ultimate aim of enhancing performance.

This study significantly contributes to the field of building energy efficiency by comparing various predictive models for heating load in residential buildings. Furthermore, the study elucidates the factors influencing predictive accuracy and provides clear visualizations of error distribution patterns, aiding researchers and practitioners in selecting appropriate modeling approaches. Overall, these contributions advance knowledge in building energy efficiency, offering pathways for future research and the adoption of more effective predictive modeling techniques in practice.

The rest of the article is organized as follows:

In section two, the explanations of the materials and the methodology which utilized in this study are provided, the methodology incudes the descriptions of the fundamental framework KNN, and the optimization algorithms SO and BWO. Then, in the third section, the performance evaluation metrics are defined, along with their formulas Furthermore in the section three, the results of the predictive models based on the results of evaluators presented. At the end of this section, a comparative analysis based on the results of the present study and the previous studies is illustrated. In section four, the potential future works are identified. Finally, the last section includes the conclusion of the study.

II. Materials and Methodology

A. Materials

In the realm of predicting heating load for residential buildings, the utilization of various input features plays a crucial role in model accuracy and performance. As observed in Table I, which provides a comprehensive overview of the statistical parameters associated with these inputs, these features encompass a range of factors, including Relative Compactness (RCE), Surface Area (SA), Wall Area (WA), Roof Area (RA), Overall-Height (OVH), Orientation (OR), Glazing Area (GA), and Glazing Area Distribution (GAD).

The statistical insights presented in Table I allow researchers and practitioners to assess the distribution, variability, and characteristics of these features, providing a foundation for model development and evaluation. These input features serve as the basis for data-driven approaches, where ML models, neural networks, and optimization techniques are employed to forecast and optimize heating load in residential buildings.

B. KNN for Regression

1) Theory: The exact process applies to regression, attributing the entity's value as the average of its nearest neighbours. In reversal, the goal is to predict dependent variables from independent ones. The $1-$closest neighbour technique, illustrated using KNN, finds the nearest neighbour to predict outcomes [21].

2) Distance metric: KNN predicts based on the K nearest neighbours' outcomes, utilizing distance metrics like Euclidean, Euclidean squared, City-block, or Chebyshev:

$$D(x,p)=\begin{cases}\sqrt{(x-p)^2} & Euclidean\ squared\\ (x-p)^2 & Euclidean\ squared\\ |x-p| & Cityblock\\ Max(|x-p|) & Chebyshev\end{cases} \quad (1)$$

In which x and p represent the inquiry spot and an instance from the selection of the illustrations, correspondingly [22], [23].

TABLE I. STATISTICAL PROPERTIES OF THE VARIABLES

Indicator	Input								Output
	RCE	SA	WA	RA	OVH	OR	GA	GAD	Heating
Max	0.98	808.5	416.5	220.5	7	5	0.4	5	43.1
Min	0.62	514.5	245	110.25	3.5	2	0	0	6.01
Median	0.75	673.75	318.5	183.75	5.25	3.5	0.25	3	18.95
Avg	0.764	671.708	318.500	176.604	5.250	3.500	0.234	2.813	22.307
Skew	0.496	-0.125	0.533	-0.163	0.000	0.000	-0.060	-0.089	0.360
St. Dev	0.106	88.086	43.626	45.166	1.751	1.119	0.133	1.551	10.090

3) K-Nearest neighbor predictions: Once the K value has been determined, forecasts can be generated using the KNN instances. In the context of regression, KNN forecasting is equivalent to the average of the results from the K nearest neighbours:

$$y = \frac{1}{K}\sum_{i=1}^{k} y_i \tag{2}$$

In which y_i represents the ith instance of the sample examples, and y is the forecast (result) of the inquiry spot. Unlike regression, regarding classification issues, KNN prognostications rely on a balloting system where the victor is employed to tag the inquiry. KNN analysis, thus far, overlooks relative proximity, giving equal influence to K neighbours. An alternative uses large K values with distance weighting for nearby instances [24].

C. Snake Optimization (SO)

Snake reproduction is influenced by temperature and food availability. Mating in cooler regions happens in late spring and summer. Female choice, male competition, and egg-laying are part of the process [25].

1) Inspiration source: SO is inspired by snake mating behaviour. Mating occurs in cold conditions with food; otherwise, snakes explore for food. During exploitation, stages optimize global efficiency. High-temperature prompts feeding, while cold environments and food lead to mating, with fighting and mating modes, possibly resulting in new snakes [26].

2) Initialization: Like all metaheuristic algorithms, SO commences by creating a uniformly distributed random population to initiate the optimization algorithm. The original population is acquired using the following equation:

$$X_i = X_{min} + r \times (X_{max} - X_{min}) \tag{3}$$

Here, X_i denotes the location of the ith individual, r represents a random value falling within the range of 0 to 1, while X_{min} and X_{max} correspond to the inferior and higher problem limits.

3) Diving the swarm into two equal groups, males and females Within this research, it is presumed that there is an equal distribution, with 50% males and 50% females in the population. The population is then categorized into two groups: males and females. To divide the swarm, the subsequent two Eq. (4) and (5) are employed:

$$N_m \approx N/2 \tag{4}$$

$$N_f = N - N_m \tag{5}$$

Here, N stands for the total number of individuals, N_m represents the count of male individuals and N_f signifies the count of female individuals.

4) Evaluate each group defining temperature and food quantity

- Identify the top individual within each cluster and determine the best female ($f_{best,f}$), and the best male ($f_{best,m}$) along with the food position (f_{food}).
- The temperature, Temp, may be characterized utilizing the subsequent equation:

$$Temp = \exp(\frac{-t}{T}) \tag{6}$$

In this case, t alludes to the present repetition, and T represents the utmost count of repetition.

- Describing the quantity of food (Q) involves determining it with the subsequent equation:

$$Q = c_1 * \exp(\frac{t-T}{T}) \tag{7}$$

The constant c_1 is fixed at a value of 0.5.

5) Exploration phase (no food): When Q is less than the threshold ($Threshold = 0.25$), the snakes forage for nourishment by picking a random location and adjusting their position accordingly. To simulate the exploration phase, the subsequent steps are taken:

$$X_{i,m}(t+1) = X_{rand,m}(t) \pm c_2 \times A_m \times ((X_{max} - X_{min}) \times rand + X_{min}) \tag{8}$$

Here, $X_{i,m}$ designates the position of the ith male, $X_{rand,m}$ signifies the random male's position, $rand$ represents a random value ranging from 0 to 1 , and A_m denotes the male's capability to locate food, which can be computed using the subsequent equation:

$$A_m = \exp(\frac{-f_{rand,m}}{f_{i,m}}) \tag{9}$$

Here, $-f_{rand,m}$ stands for the fitness of $X_{rand,m}$, while $f_{i,m}$ represents the fitness of the ith individual within the group of males, and C2 is an unchanging continuous set at 0.05:

$$X_{i,f} = X_{rand,f}(t+1) \pm c_2 \times A_f \times ((X_{max} - X_{min}) \times rand + X_{min}) \tag{10}$$

Here, the position of the ith female is denoted by $X_{i,f}$, while the location of a random female is indicated by $X_{rand,f}$. A random value between 0 and 1 is represented by rand, and the female's capacity to locate food is signified by A_f which can be computed as follows:

$$A_f = \exp(\frac{-f_{rand,f}}{f_{i,f}}) \tag{11}$$

In this case, $-f_{rand,f}$ represents the fitness of $X_{rand,f}$ and $f_{i,f}$ denotes the fitness of the ith individual within the group of females.

6) Exploitation phase (food exists)

If $Q > $ Boundary

If the Heat > Boundary (0.6)% (hot)

The snakes will exclusively relocate toward the sustenance:

$$X_{i,j}(t+1) = X_{food} \pm c_3 \times Temp \times rand \times (X_{food} X_{i,j}(t)) \tag{12}$$

If the Heat < Boundary (0.6) %cold

The snake will either engage in combat or enter the mating phase.

- Combat Mode:

$$X_{i,m}(t+1) = X_{i,m}(t) + c_3 \times FM \times rand \times (Q \times X_{best,f} - X_{i,m}(t)) \tag{13}$$

Here, $X_{i,m}$ pertains to the mode of the male in the ith mode, $X_{best,f}$ signifies the location of the superior individual within the female group, and FM represents the male agent's combat proficiency:

$$X_{i,f}(t+1) = X_{i,f}(t+1) + c_3 \times FF \times rand \times (Q \times X_{best,m} - X_{i,F}(t)) \tag{14}$$

In this scenario, $X_{i,f}$ designates the location of the female at the ith position, $X_{best,m}$ points to the location of the top individual within the male group, and FF signifies the female agent's combat aptitude.

FM and FF are derivable from the subsequent equation:

$$FM = \exp(\frac{f_{best,f}}{f_i}) \tag{15}$$

$$FF = \exp(\frac{-f_{best,m}}{f_i}) \tag{16}$$

In this context, $f_{best,f}$ represents the fitness of the top agent in the female group, $-f_{best,m}$ signifies the fitness of the foremost agent in the male group, and fi denotes the fitness of an individual agent.

- Coupling Mode:

$$X_{i,m}(t+1) = X_{i,m}(t) c_3 \times M_m \times rand \times (Q \times X_{i,f}(t) - X_{i,m}(t)) \tag{17}$$

$$X_{i,f}(t+1) = X_{i,f}(t) + c_3 \times M_f \times rand \times (Q \times X_{i,m}(t) - X_{i,f}(t)) \tag{18}$$

$X_{i,f}$ represents the ith agent's location in the female group and $X_{i,m}$ denotes the ith agent's position in the male group. Mm and Mf indicate the reproductive capacity of males and females, respectively, which can be computed as follows:

$$M_m = \exp(\frac{-f_{i,f}}{f_{i,m}}) \tag{19}$$

$$M_f = \exp(\frac{-f_{i,m}}{f_{i,f}}) \tag{20}$$

If the eggs hatch, pick the least competent male and female and substitute them:

$$X_{worst,m} = X_{min} + rand \times (X_{max} - X_{min}) \tag{21}$$

$$X_{worst,f} = X_{min} + rand \times (X_{max} - X_{\mathrm{min}}) \tag{22}$$

The operator $\pm$, or diversity factor, influences solution positions, enhancing exploration in various directions. It is a common element in metaheuristic algorithms, as seen in Hunger Games Search and others.

D. Black Widow Optimization Algorithm

The black widow spider in Mediterranean Europe uses Cannibalism in its lifecycle. In 2020, researchers developed the Black Widow Optimization (BWO) algorithm inspired by this behaviour, which has four key stages [27].

1) Initialization: $W_{N,D} = [X_1, X_2, \ldots, X_N]$ represents a group of N black widow spiders $X_1, X_2, \ldots, X_N$. D signifies the dimension relevant to an optimization problem within the given population $X_i = [x_{i,1}, x_{i,2}, \ldots, x_{i,D}] (1 \le i \le N)$ denotes the $i - th$ widow. Every component within an individual $X_i = [x_{i,1}, x_{i,2}, \ldots, x_{i,D}] (1 \le i \le N)$ each element is initialized using the formula provided in Eq. (23):

$$x_{i,j} = l_j + rand(0,1). (u_j - l_j), 1 \le j \le D \tag{23}$$

Here $L = [l_1, l_2, \ldots, l_D], U = [u_1, u_2, \ldots, u_D]$, which do the minimum and maximum limits of the parameters in the optimization model.

2) Procreate: The new generation is created through unique breeding behaviour in black spiders, randomly selecting maternal and paternal spiders to reproduce based on a specified proportion (Pp) using Eq. (24):

$$\begin{cases} Y_i = aX_i + (1-a)X_i \\ Y_j = aX_j + (1-a)X_i \end{cases} \tag{24}$$

In which Xi and Xj stand as maternal and paternal arachnids correspondingly. Yi and Yj signify the descendants resulting from mating. Additionally, α comprises a D −dimensional assortment featuring chance figures.

E. Performance Evaluator

In this segment, a series of metrics (R^2, RMSE, MSE, n10-index, and MRAE) has been developed to assess the hybrid models. These metrics gauge both the degree of error and correlation, providing valuable insights into the models' performance.

1) R^2 (Coefficient of Determination):

- R^2 measures the proportion of the variance in the dependent variable (target) that can be explained by the independent variables (features) in the model.
- It ranges from 0 to 1, where 1 indicates a perfect fit and 0 indicates no correlation.
- A higher R^2 value suggests better model performance.

2) RMSE (Root Mean Squared Error):

- RMSE quantifies the average magnitude of errors between predicted values and actual values.
- It is calculated as the square root of the average of squared differences between predicted and actual values.

- Smaller RMSE values indicate better model accuracy.

3) MSE (Mean Squared Error):

- MSE is similar to RMSE but without the square root operation.
- It represents the average of squared errors.
- Like RMSE, smaller MSE values indicate better model performance.

4) n10-index:

- The n10-index assesses the model's ability to predict extreme values.
- It focuses on the top 10% of predictions (highest or lowest).
- A higher n10-index indicates better performance in capturing extreme events.

5) MRAE (Mean Relative Absolute Error):

- MRAE measures the average relative difference between predicted and actual values.
- It considers the magnitude of errors relative to the actual values.
- Smaller MRAE values imply better model accuracy.

These metrics collectively provide valuable insights into the hybrid models' performance, considering both error and correlation aspects.

The equations for these metrics, which were employed in this study, can be found in Table II [28].

TABLE II. MATHEMATIC EQUATIONS OF THE PERFORMANCE METRICS

Coefficient Correlation (R^2):	$\boldsymbol{R^2} = \left(\frac{\sum_{i=1}^{n}(\boldsymbol{b_i} - \overline{\boldsymbol{b}})(\boldsymbol{m_i} - \overline{\boldsymbol{m}})}{\sqrt{\left[\sum_{i=1}^{n}(\boldsymbol{b_i} - \overline{\boldsymbol{b}})^2\right]\left[\sum_{i=1}^{n}(\boldsymbol{m_i} - \overline{\boldsymbol{m}})^2\right]}}\right)^2$	(25)
Root Mean Square Error (RMSE):	$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(m_i - b_i)^2}$	(26)
Mean Square Error (MSE):	$\text{MSE} = \frac{1}{n}\sum_{j=1}^{n}(m_i - b_i)^2$	(27)
$\boldsymbol{n10-index}$:	$n10-index = \frac{n10}{n}$	(28)
Mean Relative Absolute Error (MRAE)	$MRAE = \frac{1}{n}\sum_{i=1}^{n}\frac{\lvert m_i - b_i\rvert}{\lvert b_i - \overline{b}\rvert}$	(29)

Where:

- The measured value is indicated by m_i.
- Predicted values are expressed as b_i.
- The n denotes the sample size.
- The means of the measured and predicted values are represented as $\overline{m}$ and $\overline{b}$, respectively.
- The mean of the predictor variable in the dataset is symbolized as $\bar{x}$.

III. RESULT AND DISCUSSION

A. Results of the Evaluation Metrics

The results presented in Table III illustrate the effectiveness of the developed models for predicting heating load in residential buildings. Specifically, the KNSO model, which integrates the SO, emerges as the frontrunner across various performance metrics. Its low RMSE values in the training, validation, and test phases, along with high R^2 values, reflect its capacity to provide accurate forecasts. The consistency of the KNSO model in maintaining low MARE values across all phases underscores its reliability in capturing the actual heating load values. Additionally, the high n10_index observed in the training phase suggests that a significant portion of the predicted values falls within a tolerance band of actual heating load values. This reflects the KNSO model's ability to match real-world heating load data closely.

Comparatively, the traditional KNN and KNBW models exhibit commendable performance, but the KNSO model stands out as the superior choice, particularly in terms of precision and accuracy in heating load prediction. The integration of the SO Optimizer effectively refines the KNN model, providing a valuable tool for optimizing energy management and enhancing sustainability in residential buildings. These findings emphasize the significance of optimization techniques in enhancing the predictive capabilities of ML models for energy consumption. The results have practical implications for energy-efficient building design and the reduction of heating load, contributing to both economic and environmental sustainability. In conclusion, the KNSO model, when applied to heating load prediction in residential buildings, demonstrates outstanding performance and offers substantial promise for improving energy efficiency in the built environment.

In order to delve deeper into the distinctions and levels of accuracy exhibited by the various models, it is crucial to turn attention to Fig. 1. This figure offers a comprehensive comparative analysis of critical evaluation metrics, namely R^2 values, RMSE, and MSE, which are indispensable for assessing the precision of these models in predicting heating load. As previously mentioned, the KNSO model emerges as the star performer among the models. This distinction is exceptionally evident in Fig. 1, where its results consistently demonstrate the lowest values across these metrics. The exceptionally low values of RMSE and MSE and the high R^2

score underscore the remarkable precision of the KNSO model in predicting heating load, making it the standout choice. In contrast, the KNN base models exhibit comparatively weaker results when scrutinized through the lens of these metrics. They demonstrate higher RMSE and MSE values and lower R^2 scores, signifying a lower level of accuracy in their predictions compared to the KNSO model.

TABLE III. THE RESULT OF DEVELOPED MODELS FOR KNN

Model	Phase	Index values				
		RMSE	R^2	MSE	n10_index	MARE
KNN	Train	1.878	0.966	3.527	0.747	0.080
	Validation	2.254	0.952	5.080	0.635	0.101
	Test	2.145	0.956	4.603	0.696	0.085
	All	1.980	0.963	3.921	0.723	0.084
KNSO	Train	1.231	0.986	1.515	0.796	0.059
	Validation	1.422	0.979	2.021	0.896	0.059
	Test	1.500	0.977	2.251	0.922	0.058
	All	1.304	0.984	1.701	0.829	0.059
KNBW	Train	1.549	0.977	2.399	0.725	0.074
	Validation	1.803	0.967	3.251	0.661	0.078
	Test	1.729	0.970	2.988	0.774	0.064
	All	1.617	0.975	2.615	0.723	0.073

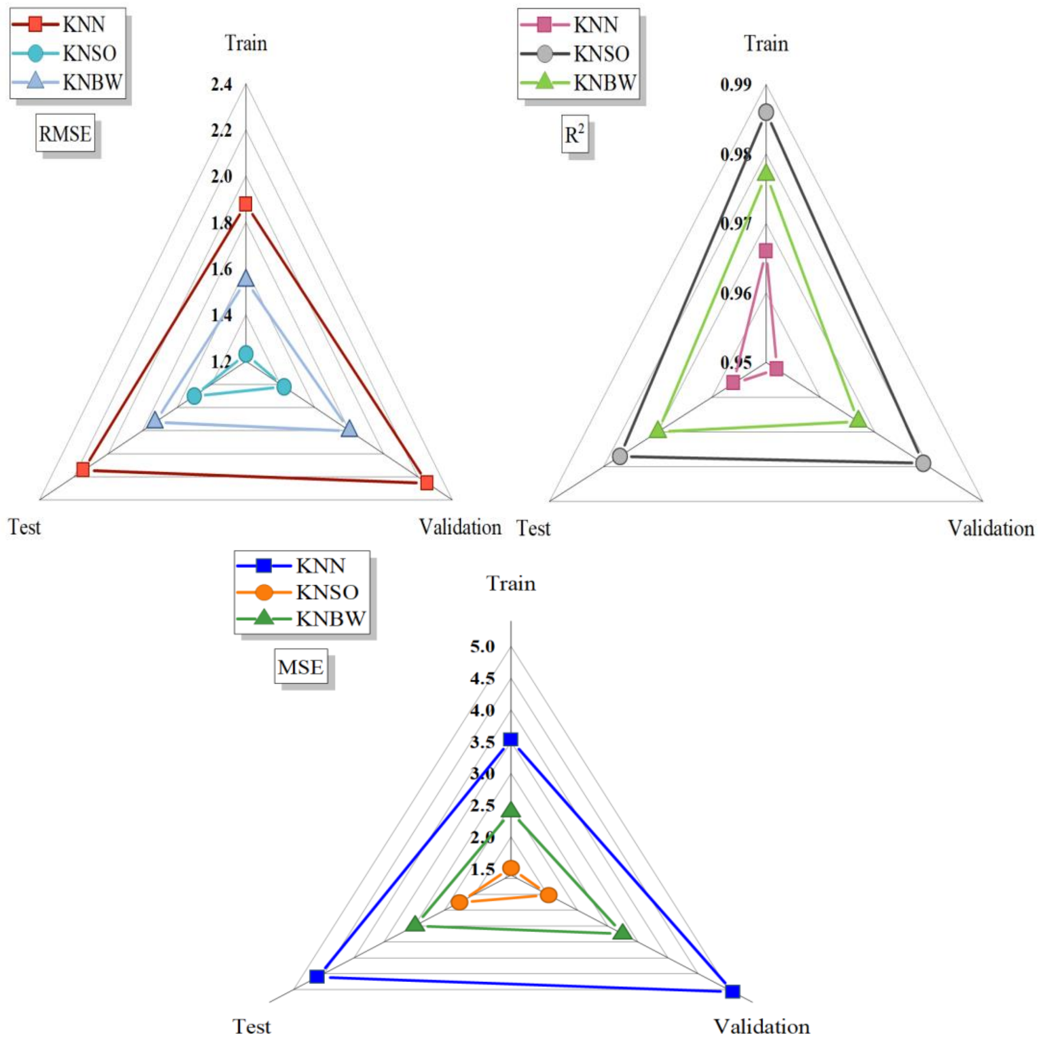


Fig. 1. Comparison between models based on RMSE, R^2, and MSE.

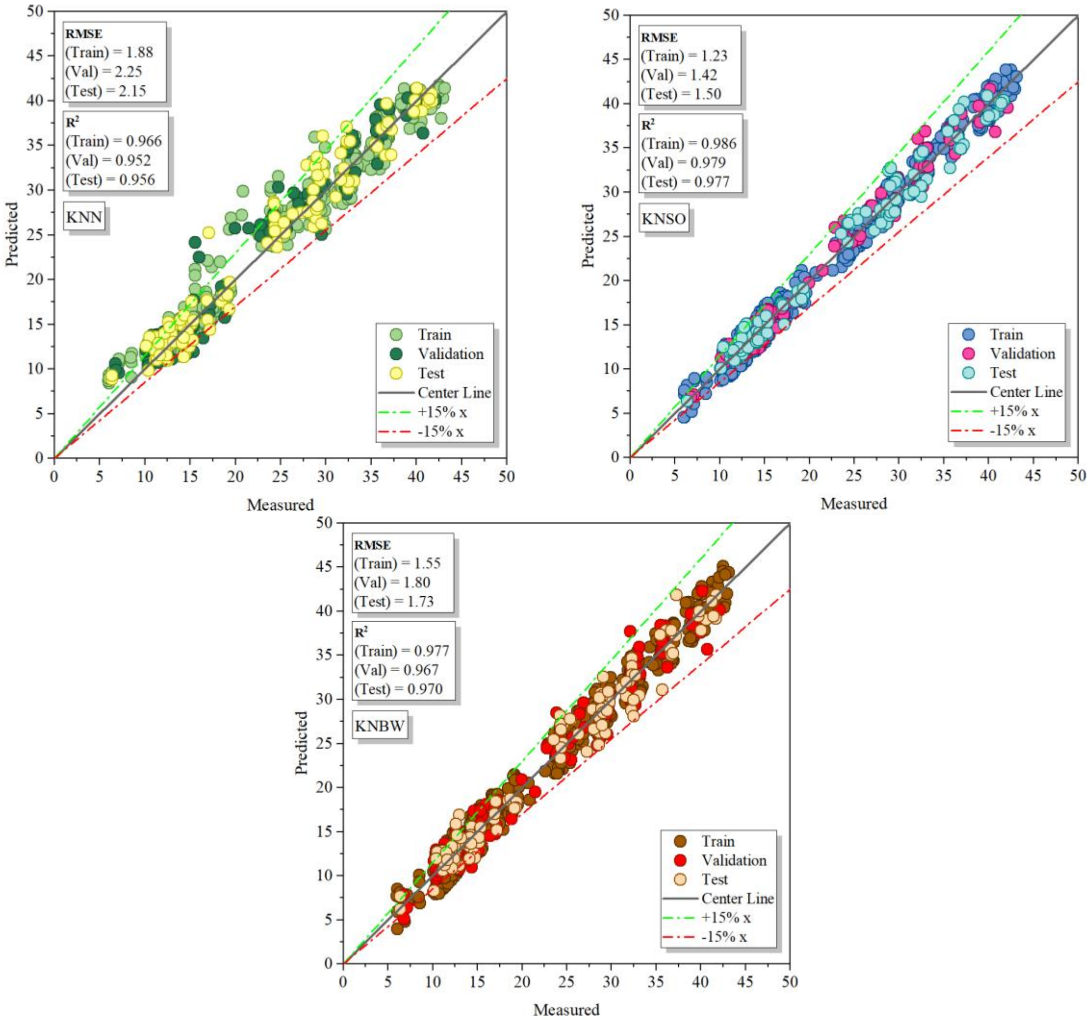


Fig. 2. Scatter plot for developed models.

In Fig. 2, a scatter plot is presented to visually illustrate the performance of the models concerning their R^2 and RMSE values. Each model, in both the training and validation phases, is represented by distinct circular markers distinguished by various colours. These markers converge towards a central line, symbolizing the ideal R^2 value of 1, signifying a perfect alignment between the predicted and actual values. A more in-depth examination of the data points associated with the KNSO model within the scatter plot reveals a closely-knit cluster positioned near the central line. The tight clustering of data points around this central line serves as compelling evidence of the KNSO model's precision in prediction, consistently remaining in proximity to the ideal R^2 value. In contrast, the KNBW and KNN models exhibit scattered data points, indicative of a broader spread of values. This dispersion within the scatter plot implies that these models show less consistency and accuracy in predicting heating load, as their data points deviate more widely from the ideal R^2 value of 1.

Carrying out a comprehensive error analysis is essential to gain a more profound understanding of the distinctive attributes and accuracy of the models under scrutiny. Such an analysis allows us to delve into the complexities of their performance. In this endeavour, Fig. 3 plays a pivotal role, offering valuable insights into the models' performance in terms of errors. Of particular note, the graph underscores the noteworthy error rate associated with the KNN model, which was particularly prominent during the testing phase. This observation serves as a crucial reference point for evaluating the model's performance in real-world scenarios, shedding light on areas where improvements may be necessary. The maximum error rate reached as high as 30%, highlighting the challenges faced by the KNN model, particularly when it comes to accurately predicting heating load (HL) values within this specific range of samples. In contrast, a more detailed examination of the KNSO model reveals an exceptional level of precision in the training phase, where the majority of data points exhibit nearly negligible errors, staying close to 0%.

This demonstrates the KNSO model's proficiency in accurately forecasting HL values during the training phase. However, the testing phase presents a slightly different scenario, with some errors emerging, although they remain relatively lower than those observed in the KNN model. Conversely, the performance of the KNBW model displays distinct characteristics. During the training phase, it registered a peak error of 50%, signifying a certain degree of inconsistency in its predictive accuracy. Remarkably, these errors persist across all three phases: training, testing, and validation, further emphasizing its unique behaviour.

In Fig. 4, the distribution characteristics of the proposed models are visually represented through a scatter interval plot, encompassing the three distinct phases: training, validation, and testing. Particularly noteworthy is the scattering of data points that correspond to the KNN model, which spans a broad range of error percentages, extending from 60 to -20. This dispersion is most conspicuous during the training phase. To effectively identify outlier data points for comparative analysis among the models, a range equivalent to 1.5 times the Interquartile Range (IQR) is employed. In contrast, the data points associated with the KNSO model are notably concentrated within a relatively narrow range of error percentages, which extends from 20 to -20. This concentration signifies a higher degree of consistency in the predictions generated by the KNGO model. On the other hand, the data points for KNBW have contained a range of -40 to 40 per cent errors, indicating a distinct distribution pattern when compared to both the KNN and KNSO models.

B. Comparison between the Outcomes of Present Study and the Existing Studies

Heating Load prediction has been the subject of numerous studies, including those conducted by Afzal et al. [29], utilizing the MLP model, and Gong et al. [30], employing the GBM technique. Notably, among the various studies referenced in Table IV, superior performance was demonstrated by the GPR model, achieving an R^2 value of 0.99 and an RMSE value of 0.059 in research conducted by Roy et al. [31]. In the current study, the foundational framework adopted was the KNN model, which was enhanced through hybridization with BWO and SO algorithms. Upon evaluating the results obtained, it was found that the integration of SO into the KNN model demonstrated exceptional applicability, yielding an R^2 value of 0.986 and an RMSE of 1.231, surpassing the performance of the other two models in this study.

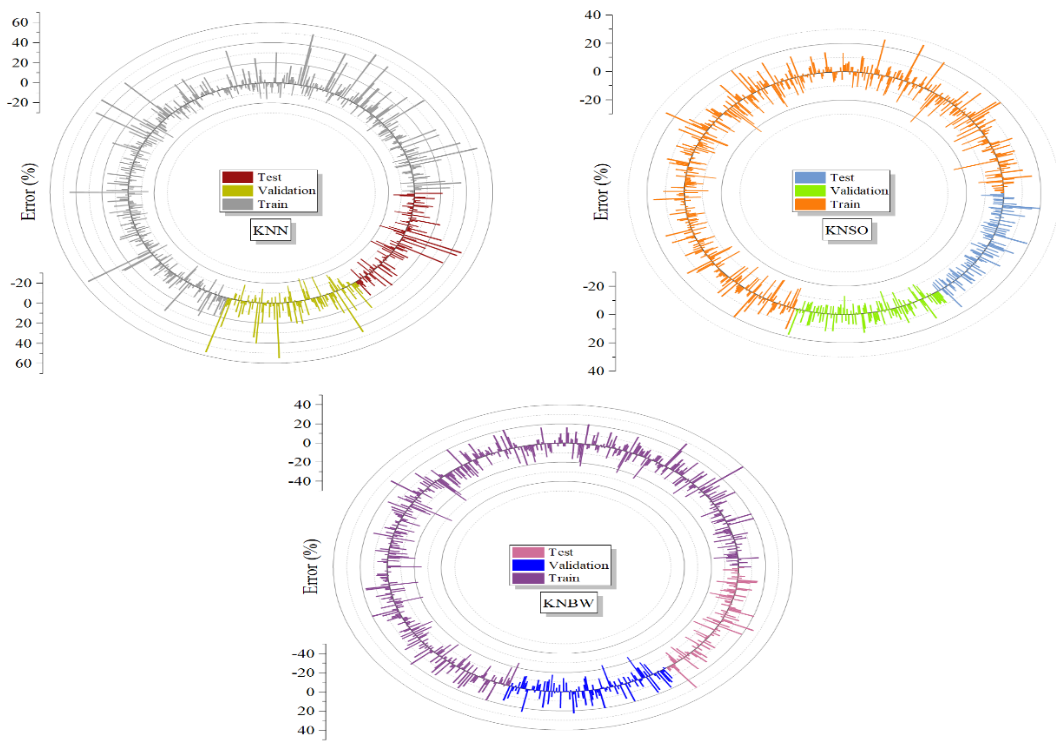


Fig. 3. The models' error percentage based on the Radial Staked Bar plot.

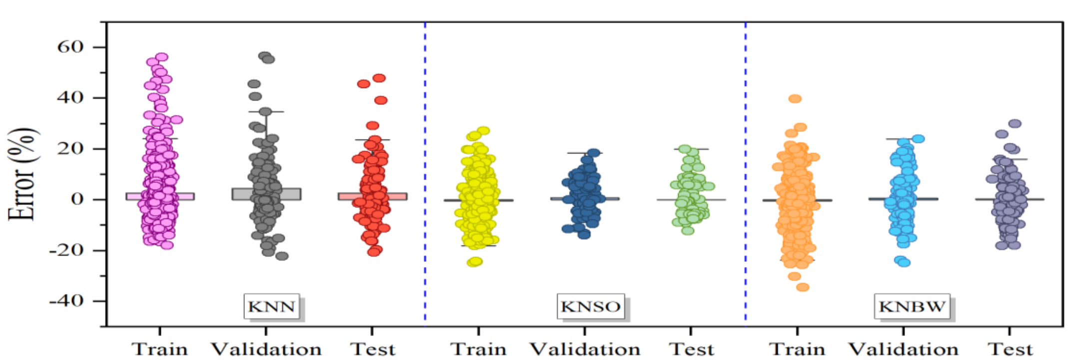


Fig. 4. The Bar Overlap of errors among the developed models.

TABLE IV. THE COMPARATIVE ANALYSIS BETWEEN EXISTING PUBLICATIONS AND CURRENT STUDY

Name	Model	Results	
		RMSE	R^2
Roy et al. [31]	GPR	0.059	0.99
Gong et al. [30]	GBM	0.1929	0.9882
Afzal et al. [29]	MLP	1.4122	0.9806
Present study	KNSO	1.231	0.986

IV. CONCLUSION

The article discussed herein delves into the realm of predictive modelling for heating load in residential buildings, focusing on the performance of various models. This exploration of the models' precision and characteristics has revealed several key insights. One of the most prominent findings is the substantial variation in accuracy across the models. The KNSO model, enhanced by the Snake Optimizer, emerges as the star performer. This model consistently exhibited the lowest RMSE and MARE values and the highest R^2 scores. These results indicate the remarkable precision of the KNSO model in predicting heating load, which holds significant promise for improving energy efficiency and sustainability in building design and management. Conversely, the KNN model, serving as the baseline, demonstrated weaker performance, with notably higher RMSE and MARE values and lower R^2 scores. This performance divergence emphasizes the significance of optimization techniques, such as the Snake Optimizer, in enhancing predictive capabilities. The KNBW model, while not reaching the same level of accuracy as the KNSO model, displayed moderate performance. Its performance characteristics, including errors and consistency, were distinct from both the KNN and KNSO models. This suggests that the optimization techniques applied in each model have a significant impact on their predictive accuracy. Furthermore, the distribution patterns of error percentages among the models, visualized in the scatter interval plot, underline the consistency and accuracy disparities. The KNSO model exhibited a notably concentrated distribution within a narrow range of error percentages, reflecting its consistent and precise predictions. In contrast, the KNN model showed a wide scattering of data points with a broader range of errors, particularly during the training phase. KNBW had its distinct distribution pattern, encompassing a specific range of errors. In conclusion, this study underscores the pivotal role of optimization techniques in refining predictive models for heating load. The KNSO model, with the Snake Optimizer, stands out as a powerful tool for accurate heating load prediction, offering valuable insights for sustainable building design. These findings hold significant implications for energy efficiency and environmental sustainability in the construction and management of residential buildings. By harnessing the capabilities of advanced optimization techniques, substantial strides could be made toward more energy-efficient and environmentally friendly building practices, contributing to a greener and more sustainable future.

V. FUTURE WORK

To enhance the effectiveness of predictive modeling for heating load in residential buildings, a multifaceted approach is warranted. Firstly, an in-depth exploration into the integration of additional variables, such as occupancy patterns, weather forecasts, and building materials, holds promise for refining prediction accuracy and capturing the intricacies of real-world scenarios more comprehensively. Moreover, delving into the application of a broader spectrum of ML algorithms, beyond those examined in this study, could yield fresh perspectives and potentially unveil more efficient models. Concurrently, conducting rigorous field studies to validate predictive model performance in authentic settings would furnish invaluable practical insights, substantiating the conclusions drawn from this research. Furthermore, a longitudinal analysis of optimized models, assessing their adaptability to evolving environmental conditions and shifting building usage patterns, stands to offer crucial data for informing sustainable building management strategies. Lastly, integrating cutting-edge advancements in optimization techniques and data analytics methodologies holds the potential to usher in a new era of even more robust and precise predictive models, thereby advancing the overarching objective of fostering energy-efficient and environmentally sustainable residential constructions.

REFERENCES

[1] S. S. Roy, P. Samui, I. Nagtode, H. Jain, V. Shivaramakrishnan, and B. Mohammadi-Ivatloo, "Forecasting heating and cooling loads of buildings: A comparative performance analysis," J Ambient Intell Humaniz Comput, vol. 11, pp. 1253–1264, 2020.

[2] Y. Ding, Q. Zhang, T. Yuan, and K. Yang, "Model input selection for building heating load prediction: A case study for an office building in Tianjin," Energy Build, vol. 159, pp. 254–270, 2018.

[3] Y. Lu, Z. Tian, Q. Zhang, R. Zhou, and C. Chu, "Data augmentation strategy for short-term heating load prediction model of residential building," Energy, vol. 235, p. 121328, 2021.

[4] Q. Zhang, Z. Tian, Z. Ma, G. Li, Y. Lu, and J. Niu, "Development of the heating load prediction model for the residential building of district heating based on model calibration," Energy, vol. 205, p. 117949, 2020.

[5] F. Dalipi, S. Yildirim Yayilgan, and A. Gebremedhin, "Data-driven machine-learning model in district heating system for heat load prediction: A comparison study," Applied Computational Intelligence and Soft Computing, vol. 2016, 2016.

[6] C. Wang et al., "Research on thermal load prediction of district heating station based on transfer learning," Energy, vol. 239, p. 122309, 2022.

[7] B. S. A. J. khiavi; B. N. E. K. A. R. T. K. hadi Sadaghat;, "The Utilization of a Naïve Bayes Model for Predicting the Energy Consumption of Buildings," Journal of artificial intelligence and system modelling, vol. 01, no. 01, 2023, doi: 10.22034/JAISM.2023.422292.1003.

[8] R. Chaganti et al., "Building heating and cooling load prediction using ensemble machine learning model," Sensors, vol. 22, no. 19, p. 7692, 2022.

[9] M. Protić et al., "Appraisal of soft computing methods for short term consumers' heat load prediction in district heating systems," Energy, vol. 82, pp. 697–704, 2015.

[10] E. Guelpa, L. Marincioni, M. Capone, S. Deputato, and V. Verda, "Thermal load prediction in district heating systems," Energy, vol. 176, pp. 693–703, 2019.

[11] G. Xue, Y. Pan, T. Lin, J. Song, C. Qi, and Z. Wang, "District heating load prediction algorithm based on feature fusion LSTM model," Energies (Basel), vol. 12, no. 11, p. 2122, 2019.

[12] J. Yuan et al., "Identification heat user behavior for improving the accuracy of heating load prediction model based on wireless on-off control system," Energy, vol. 199, p. 117454, 2020.

[13] Y. Zhang, Z. Zhou, J. Liu, and J. Yuan, "Data augmentation for improving heating load prediction of heating substation based on TimeGAN," Energy, vol. 260, p. 124919, 2022.

[14] J. Ling, N. Dai, J. Xing, and H. Tong, "An improved input variable selection method of the data-driven model for building heating load prediction," Journal of Building Engineering, vol. 44, p. 103255, 2021.

[15] A. Moradzadeh, A. Mansour-Saatloo, B. Mohammadi-Ivatloo, and A. Anvari-Moghaddam, "Performance evaluation of two machine learning techniques in heating and cooling loads forecasting of residential buildings," Applied Sciences, vol. 10, no. 11, p. 3829, 2020.

[16] G. Xue, C. Qi, H. Li, X. Kong, and J. Song, "Heating load prediction based on attention long short term memory: A case study of Xingtai," Energy, vol. 203, p. 117846, 2020.

[17] M. Sajjad et al., "Towards efficient building designing: Heating and cooling load prediction via multi-output model," Sensors, vol. 20, no. 22, p. 6419, 2020.

[18] K. Kato, M. Sakawa, K. Ishimaru, S. Ushiro, and T. Shibano, "Heat load prediction through recurrent neural network in district heating and cooling systems," in 2008 IEEE international conference on systems, man and cybernetics, IEEE, 2008, pp. 1401–1406.

[19] Z. Wang, T. Hong, and M. A. Piette, "Building thermal load prediction through shallow machine learning and deep learning," Appl Energy, vol. 263, p. 114683, 2020.

[20] T.-Y. Kim and S.-B. Cho, "Predicting residential energy consumption using CNN-LSTM neural networks," Energy, vol. 182, pp. 72–81, 2019.

[21] S. B. Imandoust and M. Bolandraftar, "Application of k-nearest neighbor (knn) approach for predicting economic events: Theoretical background," Int J Eng Res Appl, vol. 3, no. 5, pp. 605–610, 2013.

[22] L. Xiong and Y. Yao, "Study on an adaptive thermal comfort model with K-nearest-neighbors (KNN) algorithm," Build Environ, vol. 202, p. 108026, 2021.

[23] H. A. Abu Alfeilat et al., "Effects of distance measure choice on k-nearest neighbor classifier performance: a review," Big Data, vol. 7, no. 4, pp. 221–248, 2019.

[24] S. Uddin, I. Haque, H. Lu, M. A. Moni, and E. Gide, "Comparative performance analysis of K-nearest neighbour (KNN) algorithm and its different variants for disease prediction," Sci Rep, vol. 12, no. 1, p. 6256, 2022.

[25] F. A. Hashim and A. G. Hussien, "Snake Optimizer: A novel meta-heuristic optimization algorithm," Knowl Based Syst, vol. 242, p. 108320, 2022.

[26] P. V Klimov, J. Kelly, J. M. Martinis, and H. Neven, "The snake optimizer for learning quantum processor control parameters," arXiv preprint arXiv:2006.04594, 2020.

[27] G. Hu, B. Du, X. Wang, and G. Wei, "An enhanced black widow optimization algorithm for feature selection," Knowl Based Syst, vol. 235, p. 107638, 2022.

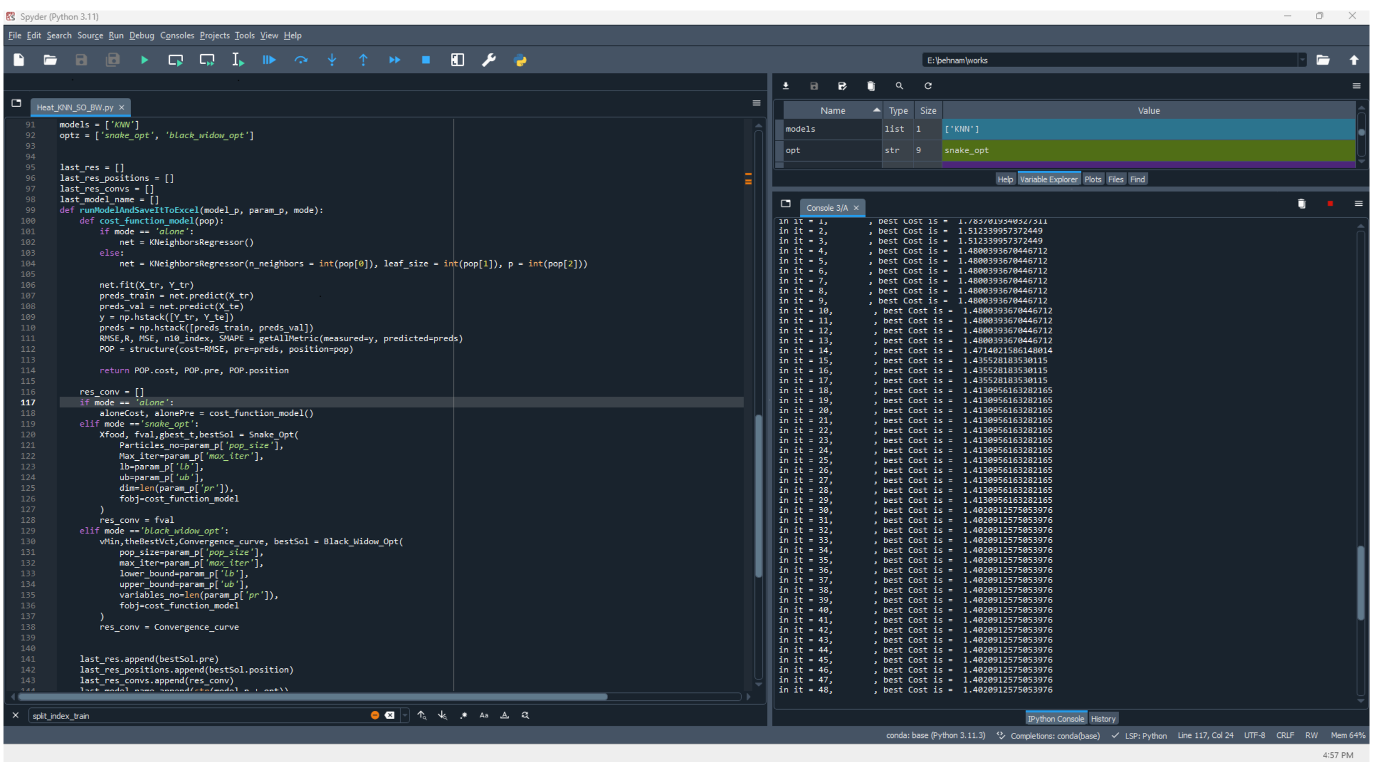
[28] A. Botchkarev, "Performance metrics (error measures) in machine learning regression, forecasting and prognostics: Properties and typology," arXiv preprint arXiv:1809.03006, 2018.

[29] S. Afzal, B. M. Ziapour, A. Shokri, H. Shakibi, and B. Sobhani, "Building energy consumption prediction using multilayer perceptron neural network-assisted models; comparison of different optimization algorithms," Energy, p. 128446, Jul. 2023, doi: 10.1016/j.energy.2023.128446.

[30] M. Gong, Y. Bai, J. Qin, J. Wang, P. Yang, and S. Wang, "Gradient boosting machine for predicting return temperature of district heating system: A case study for residential buildings in Tianjin," Journal of Building Engineering, vol. 27, p. 100950, 2020.

[31] S. S. Roy, P. Samui, I. Nagtode, H. Jain, V. Shivaramakrishnan, and B. Mohammadi-Ivatloo, "Forecasting heating and cooling loads of buildings: A comparative performance analysis," J Ambient Intell Humaniz Comput, vol. 11, pp. 1253–1264, 2020.

APPENDIX I



Screenshot of the Simulation