Estimation of Heating Load Consumption in Residual Buildings using Optimized Regression Models Based on Support Vector Machine

Chao WANG[1], Xuehui QIU[2]
Qinhuangdao Vocational and Technical College, Qinhuangdao, 066100, China[1]
School of Urban Geology and Engineering, Hebei GEO University, Shijiazhuang 050000, China[2]

***Abstract*—Accurate energy consumption forecasting and assessing retrofit options are vital for energy conservation and emissions reduction. Predicting building energy usage is complex due to factors like building attributes, energy systems, weather conditions, and occupant behavior. Extensive research has led to diverse methods and tools for estimating building energy performance, including physics-based simulations. However, accurate simulations often require detailed data and vary based on modeling sophistication. The growing availability of public building energy data offers opportunities for applying machine learning to predict building energy performance. This study evaluates Support Vector Regression (SVR) models for estimating building heating load consumption. These models encompass a single model, one optimized with the Transit Search Optimization Algorithm (TSO) and another optimized with the Coot optimization algorithm (COA). The training dataset consists of 70% of the data, which incorporates eight input variables related to the geometric and glazing characteristics of the buildings. Following the validation of 15% of the dataset, the performance of the remaining 15% is evaluated using five different assessment metrics. Among the three candidate models, Support Vector Regression optimized with the Coot optimization algorithm (SVCO) demonstrates remarkable accuracy and stability, reducing prediction errors by an average of 20% to over 50% compared to the other two models and achieving a maximum R^2 value of 0.992 for heating load prediction.**

Keywords*—*Heating load demand; prediction models; building energy consumption; support vector machine; metaheuristic optimization algorithms

I. Introduction

Residential energy consumption accounts for approximately 30% of the total energy used [1,2]. Consequently, the precise forecasting of energy consumption during the design phase and the assessment of retrofit options emerge as critical endeavors in the adventure for energy conservation and emissions reduction. The prediction of energy usage in buildings presents a formidable challenge, given its reliance on numerous factors, including building attributes, control, characteristics and maintenance, meteorological parameters, and occupants' behavior, among other sociological variables [3,4]. In response to this challenge, significant efforts from the scientific community, governmental entities, and industry stakeholders have spurred numerous research initiatives, resulting in various approaches, methodologies, and tools for estimating building energy performance. Building energy simulation tools, notably those based on physics, such as Energy Plus [5], have gained widespread adoption for investigating and evaluating building energy efficiency. However, achieving precise simulations necessitates detailed building information, including specific space characteristics, which can be challenging to obtain [6]. Moreover, research has demonstrated substantial variations in outcomes based on the level of sophistication, both in terms of physical modeling and mathematical complexity, applied in the energy of building models [7].

In order to forecast total energy consumption or particular end uses, researchers have complemented physics-based models with a variety of statistical techniques [8]. By taking into account the characteristics of the building and its occupants and comparing the outcomes with simulations, regression-based approaches have become popular for forecasting energy consumption. Catalina et al. [9] used polynomial regression with a model displaying a maximum variance of 5% from simulated data across scenarios for estimating heating demand. Due to their ability to handle complicated interactions, more current advanced machine learning algorithms like Artificial Neural Networks (ANN) and Support Vector Machines (SVM) have been deployed. By taking into account the features of the building and its occupants and comparing the outcomes with simulations, regression-based approaches have become popular for estimating energy use. Xifara and Tsanas [10] demonstrated the superiority of Random Forest (RF) over regression in estimating heating load (HL) and cooling load (CL). The study employed a statistical machine learning framework to analyze how eight input variables affect HL and CLs in residential buildings. It systematically investigated the association strength between each input variable and the output variable star using classical and non-parametric statistical tools. Comparisons were made between classical linear regression and RF for estimating HL and CL. Simulations on 768 residential buildings demonstrated the capability to predict HL and CL with low mean absolute error deviations. Overall, the study supported the use of machine learning for accurate building parameter estimation in the context of energy-efficient design and operation. Li et al. [11] employed SVM and ANN to predict cooling demand, with SVM-based predictions having roughly half the errors compared to ANN predictions when matched against simulation data. They established an hourly building CL prediction model using SVM and applied it to an office building located in Guangzhou, China. The simulation results indicated that the SVM method outperformed the traditional back-propagation (BP) neural network model in terms of accuracy and generalization. Neto and Fiorelli [12] found similar performance between Energy Plus and an ANN model for energy consumption estimation. These studies collectively indicate the effectiveness of empirical methods in capturing complex achieving and relationship accuracy similar to or better than regression-based models compared to simulation results.

Instead of relying exclusively on simulated results, it is crucial to evaluate how well these sophisticated statistical models anticipate the precise energy performance of buildings in the future [13]. Significant differences between original design simulations and actual energy calculations have been found in prior studies. These differences have mostly been related to modeling assumptions, building quality, weather variations, operating practices, and occupant behavior [14]. A variety of opportunities exist to use cutting-edge approaches for examining the complicated relationship between building and occupant features and real energy performance through the analysis of large datasets as a result of the increasing availability of data regarding actual energy usage [15].

Numerous case studies have used sophisticated algorithms and previous data to predict the energy performance of buildings. For instance, Gonzalez and Zamarreno [16] presented a novel approach for short-term load prediction in buildings. The method relied on a specialized artificial neural network (ANN) that incorporated feedback from a portion of its outputs. The training of this ANN utilized a hybrid algorithm. The new system incorporated current and forecasted values of temperature, the current load, and the hour and day as inputs. The performance of this predictor underwent evaluation using real data and results from international contests. The obtained results demonstrated the high precision achieved with this system. Dong et al. [17] investigated the use of SVM for predicting building energy consumption in the tropical region, which is crucial for baseline model development and measurement and verification protocols. Four commercial buildings in Singapore were studied, employing weather data and monthly utility bills. SVM's performance, influenced by parameters C and ε, was analyzed using a radial basis function (RBF) kernel. Results indicated SVM's effectiveness, producing predictions with coefficients of variance under 3% and percentage errors within 4%. The study demonstrated SVM's feasibility and applicability in building load forecasting, offering valuable insights for accurate energy consumption predictions in tropical climates. Tso and Yau [18] conducted an empirical study comparing regression, decision tree, and ANN and decision tree models to beat regression techniques in forecasting power use in residential buildings. Collectively, these case studies demonstrate that machine learning algorithms offer dependable outcomes. They possess the ability to model non-linear relationships, and many are non-parametric, obviating the need for specific probability distribution assumptions [19]. It is important to remember that earlier examinations sometimes concentrated on a single structure or a small group of buildings in a particular place. Because of this, there is still a gap in the development of generalized prediction models based on Machine Learning (ML) algorithms, making it difficult to use these techniques to fully examine the relevance of different architectural and occupant features for building energy efficiency [20].

Bashir and Alotaibi [21] underscored the crucial role of implementing effective building cooling and heating load prediction models for enhanced energy efficiency. In recent years, several research studies addressed challenges in determining efficient input parameters and developing accurate prediction models. Various data-driven approaches were proposed to optimize energy consumption systems and ensure indoor comfort. Despite existing reviews on prediction models, gaps remained in assessing cooling and heating load predictions. This study critically reviewed recent models, focusing on performance and accuracy. Comparative analysis revealed specific advantages for each model, yet shortcomings persisted in input parameters and implementation techniques. The review aimed to highlight and compare existing models' disadvantages in cooling and heating load predictions. Gong et al. [22] investigated the prediction of heating energy consumption in residential structures in Tianjin. They used a variety of methods, such as Support Vector Regression (SVR), Multilayer Perceptron (MLP), RF, and Light Gradient Boosted Machine (LGBM). The results showed that the LGBM model beat its competitors in a variety of assessment measures, demonstrating its potential for exact energy consumption predictions. Nebot and Mugica the [23] investigated prediction of heating and cooling loads in residential constructions utilizing fuzzy logic approaches such as fuzzy inductive reasoning (FIR) and adaptive neural fuzzy inference system (ANFIS). In their study, thirteen machine learning algorithms were compared to various fuzzy approaches, and SVR, ANFIS, and FIR performed better. Moradzadeh et al. [24] concentrated on estimating cooling and heating loads using SVR and MLP models. The MLP model had an incredible R^2-value of 0.9993 for heating load prediction, while the SVR model excelled with an R-value of 0.9878 for cooling load prediction, yielding outstanding results for their study. These findings illustrate the level of precision that may be achieved using machine learning algorithms. Karijadi and Chou [25] addressed the challenge of accurately predicting building energy consumption, which is crucial for effective building energy management systems. Due to the non-linear and nonstationary nature of energy consumption data, conventional prediction methods faced difficulties. The research introduced a novel hybrid approach, combining RF and Long Short-Term Memory (LSTM) based on Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN). The method transformed the original energy consumption data into components using CEEMDAN, where RF predicted the highest frequency component and LSTM predicted the rest. Combining the predictions yielded superior results compared to benchmark methods, as demonstrated in experiments using real-world building energy consumption data.

In the current study, inspired by prior successful results demonstrating the superior performance of SVM over other models, support vector regression-based models were developed to predict heating loads (HL) in buildings. Another advantage of this study is the utilization of numerous datasets, including various input variables related to building geometry and glazing status, which were collected from previous literature for training predictive models. The predictive performance of a single SVR model was assessed, and in optimizing the training process, two distinct optimizers, namely the Transit Search Optimization Algorithm ($TSOA$) and the Coot optimization algorithm (COA), were employed. The predicted results of the three models were compared using performance metrics, including R^2, $RMSE$, MAE, RSR, and $MRAE$. Subsequently, the most optimal hybrid model for predicting HL in buildings was determined.

The novelty of this study lies in its application of SVR models for predicting HL in buildings, driven by prior evidence showcasing the superior performance of SVR over other modeling techniques. Additionally, the study introduces a unique aspect by incorporating a diverse range of datasets that encompass various input variables associated with building geometry and glazing status. These datasets, sourced from existing literature, contribute to the comprehensive training of predictive models.

Moreover, the study distinguishes itself by evaluating the predictive performance of a singular SVR model and introducing innovation in the optimization of the training process. Specifically, the study employs two distinct optimizers, the TSOA and the COA, to enhance and fine-tune the efficiency of the SVR model training. This dual-pronged approach toward model optimization adds a novel dimension to the study, contributing to its overall uniqueness in addressing the prediction of heating loads in buildings.

In Section II, data, models, and optimizers will be introduced. In Section III, the evaluation models are developed, and the metrics used for evaluation are discussed. Finally, in Section IV, the conclusion of the study is mentioned along with the limitations and future study.

II. Materials and Methods

A. Data Collection

To guarantee the validity and efficacy of the approaches described in this study, the availability of reliable and substantial data is crucial. The dataset created to train the intelligent models from earlier research was utilized in this investigation. This dataset provides the crucial data needed to put the suggested strategies into practice and evaluate how well they anticipate building heating needs. Eight significant factors, including relative compactness (which represents the building's surface area-to-volume ratio), roof area, surface area, wall area, overall height, orientation, glazing area (which includes glazing, frame, and sash components), and glazing area distribution, have an impact on the analysis of the input parameters in this study. The key criteria used for the statistical analysis of the dataset are shown in Table I, together with metrics such as data averages (Avg), standard deviation ($St.Dev.$), minimum (Min), and maximum (Max) values.

B. Overview of Machine Learning (ML) Methods and Optimizers

1) Support Vector Regression (SVR): In the early steps of pattern recognition research, support vector machines (SVM) were used to identify patterns. This approach was initially proposed by Vapnik [26] and later advocated utilizing the SVM to address issues about function approximation. The SVR methodology involves a dataset comprising $\overline{\mathrm{N}}$ elements $\{(\mathrm{X_i}, \mathrm{y_i}), \mathrm{i} = 1,2, \dots, \overline{\mathrm{N}}\}$, where N represents the amount of training examples. The variable $\mathrm{X_i}$ denotes the $\mathrm{i} - \mathrm{th}$ section of an N $-$dimensional vector, where $\mathrm{X_i} = \{\mathrm{x_1}, \mathrm{x_2}, \dots, \mathrm{x_n}\} \in \mathrm{R^n}$, and $\mathrm{y_i} \in \mathrm{R}$ presents the actual value related to $\mathrm{X_i}$.

The underlying principle of the SVR involves utilizing a ML technique to chart train data opinions, denoted as precisly X_i, onto a nose space that typically has l dimensions.

TABLE I. The Statistical Properties of the Input Mutable of Heating

Variables		**Indicators**			
	Category	**Min**	**Max**	**Avg**	**St. Dev.**
Relative Compactness	Input	0.62	0.98	0.76	0.11
Surface Area (m^2)	Input	514.50	808.50	671.71	88.09
Wall Area (m^2)	Input	245.00	416.50	318.50	43.63
Roof Area (m^2)	Input	110.25	220.50	176.60	45.17
Overall height (m)	Input	3.50	7.00	5.25	1.75
Orientation	Input	2.00	5.00	3.50	1.12
Glazing Area (%)	Input	0.00	0.40	0.23	0.13
Glazing Area Distribution	Input	0.00	5.00	2.81	1.55
Heating (KW)	Output	6.01	43.10	22.31	10.09

An optimized hyperplane that precisely depicts the non-linear relationship between the output and the current input independent variables is created by carefully designing the feature space. One formal way to represent the expression for SVR is as shown in Eq. (1):

$$f(x) = V^T \emptyset(x) + a \quad (1)$$

Here, a is the variable factor, $f(x)$ is the predicted ideals, and V is the $l - dimensional$ weight factor. $\emptyset(x)$ represents the arrangement of every component (X_i) into a feature space with in height dimensions. Eq. (2) represents the way the $\varepsilon -$insensitive loss function is expressed.

$$|y - f(x)|_\varepsilon = \max(0, |y - f(x)| - \varepsilon) \quad (2)$$

The residual is denoted by Eq. (3) as the disparity among the definite value, y, and the estimated cost, $f(x)$.

$$R(x, y) = y - f(x) \quad (3)$$

The ideal model is determined by incorporating the entire remaining element within a predefined variety of ε, as follows:

$$-\varepsilon \le R(x, y) \le \varepsilon \quad (4)$$

Eq. (4) provides the hypothesis for the complete training data. Hence, the data displays the highest disparity from the hyperplane when the remaining adheres to the condition $R(x, y) = \pm\varepsilon$. The physical distance among a specific data point (x, y) and the hyperplane $R(x, y) = 0$ is calculated as $|R(x, y)|/\|V^*\|$ obtained in the following manner:

$$V^* = (1, -V^T)^T \quad (5)$$

The hypothesis of this study suggests that the most significant movement among the dataset (x, y) and the hyperplane $R(x, y) = 0$ can be expressed as the adjustable δ. Hence, it can be inferred that the complete train dataset meets the criteria specified in Eq. (6). The attainment of the maximum value of δ implies that the SVR model can demonstrate the best performance of generalization.

$$|R(x, y)| \le \delta\|V^*\| \quad (6)$$

The highest distance is reached where the value of the $R(x, y)$ generations a predefined ε value. Following this, Eq. (6) can be formulated again and expressed as Eq. (7). In order to reach the maximum value of δ, it is crucial to minimize $\|V^*\|$, and as $\|V^*\|^2 = \|V\|^2 + 1$, the problem of optimization is transformed into minimizing $\|V\|$.

$$\varepsilon = \delta\|V^*\| \quad (7)$$

Despite efforts made over the train phase to minimize errors in the variety of $(-\varepsilon, \varepsilon)$, there is still a possibility that specific errors may exceed this limit. Errors occurring during training that are less than $-\varepsilon$ are denoted as ζ_i, whereas training errors greater than ε are represented as ${\zeta^*}_i$. The notations ζ_i and ${\zeta^*}_i$ are clarified based on following equations:

$$\zeta_i = \begin{cases} 0 & R(x_i, y_i) - \varepsilon \le 0 \\ R(x_i, y_i) - \varepsilon & others \end{cases} \quad (8)$$

$${\zeta^*}_i = \begin{cases} 0 & \varepsilon - R(x_i, y_i) \le 0 \\ \varepsilon - R(x_i, y_i) & others \end{cases} \quad (9)$$

The primary aim of SVR algorithm is to identify the hyperplane that yields the optimal result though reducing the disparity among the error of training and the hyperplane. This is accomplished by utilizing the ε insensitive loss purpose. Eq. (10) presents the objective function for optimizing SVR.

$$minF(W, b, \zeta_i, {\zeta^*}_i) = \frac{1}{2}\|W\|^2 + c\sum_{i=1}^{N}(\zeta_i + {\zeta^*}_i) \quad (10)$$

With the restrictions:

$$y_i - W^T\varphi(x_i) - b \le \varepsilon + \zeta_i \quad i = 1, 2, \dots, \overline{N}$$

$$W^T\varphi(x_i) + b - y_i \le \varepsilon + {\zeta^*}_i \quad i = 1, 2, \dots, \overline{N}$$

$$\zeta_i \ge 0, {\zeta^*}_i \ge 0 \quad i = 1, 2, \dots, \overline{N}$$

Parameter C plays an essential role in achieving a equility between minimizing training errors and an optimal separation among the hyperplane space and the data points in SVR involves preserving an ideal margin.

In Eq. (10), the initial part penalizes excessive weight values to maintain a flat regression function. The second part balances error margins and experience risk using the ε-insensitive loss function.

Upon effectively resolving the quadratic optimization, which involves disparity constraints, the factor W is derived using the guidelines explained in Eq. (1) and Eq. (11).

$$W = \sum_{i=1}^{N}\left({\beta^*}_i - \beta_i\right)\varphi(x_i) \quad (11)$$

To calculate ${\beta^*}_i and\ \beta_i$, it is necessary to solve a quadratic programming problem that identifies the Lagrangian multipliers.

Eq. (12) represents the SVR function:

$$f(x) = \sum_{i=1}^{N}\left({\beta^*}_i - \beta_i\right)K(x_i - x) + b \quad (12)$$

The kernel function, denoted as $K(x_i - x)$, possesses the ability to nonlinearly project the train data onto a various characterized by a high−dimensional space with l dimensions.

The Kernel function $K(x_i - x)$ possesses the proficiency to non-linearly project the train data onto a various with many dimensions (l-dimensions), rendering it well-suited for tackling issues associated with non-linear relationships. This is particularly valuable within the context of electrical forecasting. Fig. 1 demonstrates the schematic representation of the SVR's workflow.

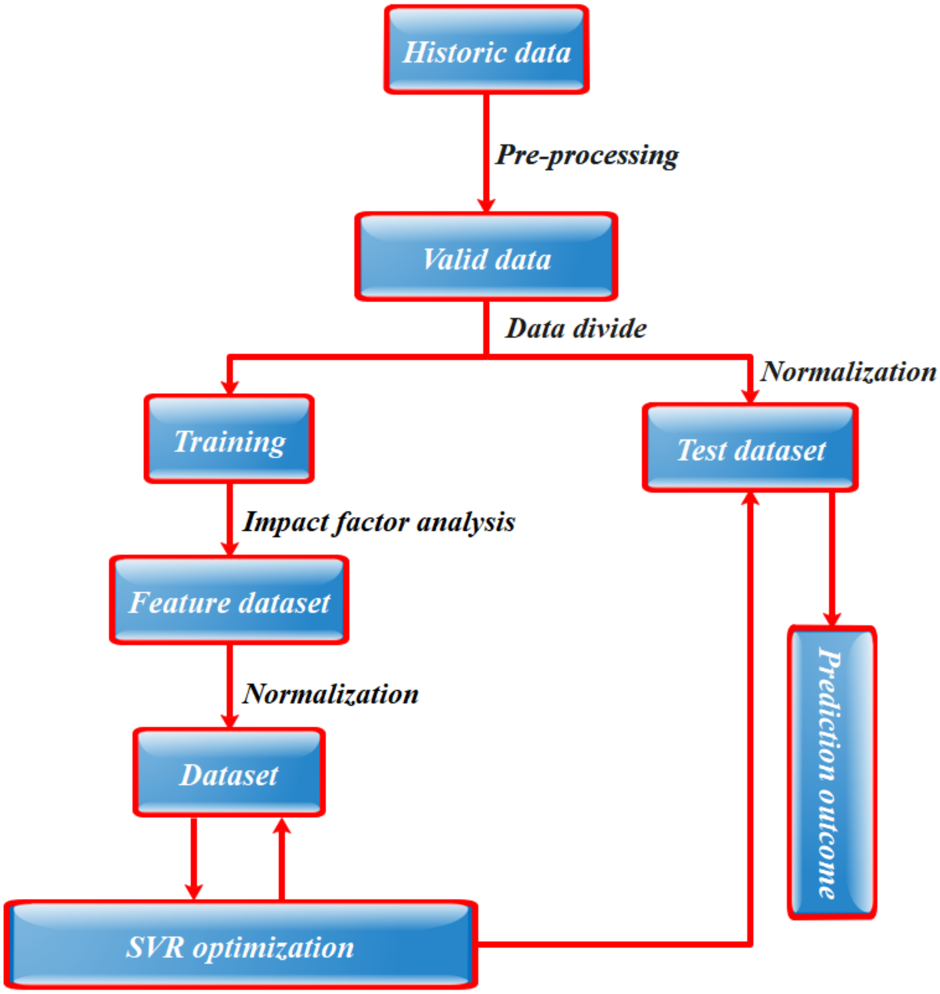


Fig. 1. Flowchart of SVR model.

2) Coot Optimization Algorithm (COA): The COA uses a metaheuristic optimization approach inspired by Coots' collective behaviors. Coots display various movements in water, like chain, random, leader-driven, and leader-adjusted, to reach food sources or specific locations. The COOT algorithm incorporates these behaviors and initiates by choosing a population using Eq. (13) [27]:

$$CootPos(i) = rand\ (1, N) \times (UB - LB) + LB \tag{13}$$

$CootPos(i)$ is the spatial organizes of an individual coot, while N represents the problem's dimensionality or the quatity of involved variables. UB and LB are the $upper$ and $lower$ limits of the exploration space, individually.

$$UB = [UB_1, UB_2, \dots, UB_N]\ , LB = [LB_1, LB_2, \dots, LB_N] \tag{14}$$

Following the initial population setup, the positions of the coots are subsequently modified according to four patterns of movement.

a) Random-Movement: The position Q for this particular movement is firstly randomized:

$$Q = rand(1, N) \times (UB - LB) + LB \tag{15}$$

Then, the position is updated to prevent getting stuck in local optima:

$$CootPos(i) = CootPos(i) + A \times R_2 \times (Q - CootPos(i)) \tag{16}$$

The value R_2 is number within the range $[0, 1]$, and A is defined as follows:

$$A = 1 - L \times (\frac{1}{Iter}) \tag{17}$$

Here, $Iter$ is the maximum acceptable number of iterations, while L represents the present iteration number.

b) Chain-Movement: To perform the chain program, one can determine the mean position of 2 coot birds utilizing the formula outlined in Eq. (18).

$$CootPos(i) = \frac{CootPos(i-1) + CootPos(i)}{2} \tag{18}$$

Here, $CootPos(i - 1)$ presents the position of the next coot in the sequence.

c) Adjusting position following the leader: In each group, a coot bird adjusts its location according to that of the leader, bringing the follower closer to the leader. The equation given in Eq. (19) is used to determine the leader selection procedure.

$$K = 1 + (i\ MOD\ NL) \quad (19)$$

K represents the leader's index, i denotes the follower coot's number, and NL is the amount of group's leaders.

A coot's current location is getting updated applying Eq. (20):

$$CootPos(i) = LeaderPos(K) + 2 \times R_1 \times Cos(2R\pi) \times (LeaderPos(K) - CootPos(i)) \quad (20)$$

$CootPos(i)$ denotes the position of the coot bird, while $LeaderPos(K)$ is the position of the chosen leader. R_1 represents a random number within $[0, 1]$, and R is a random number within the range $[-1, 1]$.

d) Leader-Movement

Leader locations are updated with Eq. (21) aimed at shifting from local optimal locations to global optimal positions.

$$LeaderPos(i) = \begin{cases} B \times B_3 \times Cos(2\pi R) \times \left(gBest - LeaderPos(i)\right) + gBest & B_4 < 0.5 \\ B \times B_3 \times Cos(2\pi R) \times \left(gBest - LeaderPos(i)\right) - gBest & B_4 \geq 0.5 \end{cases} \quad (21)$$

The symbol $gBest$ represents the optimal attainable position, while B_3 and B_4 are randomly chosen numbers selected from the interval [0, 1]. B is determined using the Eq. (22).

$$B = 2 - L \times \left(\frac{1}{Iter}\right) \quad (22)$$

3) Transit Search Optimization Algorithm (TSOA): In the TSOA algorithm, there are two key parameters: the number of host stars (ns) and the signal-to-noise ratio (SN), determined based on the transit model. Noise is estimated using standard deviation from observations outside the transit phase. The product of ns and SN sets the initial population size for TS [28].

This section discusses the five crucial phases of the TSOA as follows:

a) Galaxy phase: The algorithm begins by opting a galaxy and a random center within the search space. It then identifies habitable zones (life belts) within the galaxy by Eq. evaluating ns*SN random regions. L_R using Eq. (23) to Eq. (25). The top ns regions with the best fitness, indicating a high probability of hosting life, are selected for further algorithmic steps.

$$L_{R,I} = L_{Galaxy} + D - Noise \qquad I = 1, \ldots, (ns * SN) \quad (23)$$

$$D = \begin{cases} c_1 L_{Galaxy} - L_r & if\ \ z = 1\ (negative\ region) \\ c_1 L_{Galaxy} + L_r & if\ \ z = 2\ (positive\ region) \end{cases} \quad (24)$$

$$Noise = (c_2)^3 L_r \quad (25)$$

In the equations mentioned above, L_{Galaxy} denotes the central position of the galaxy, while L_r represents a randomly selected location within the exploration space. Additionally, there are *two* coefficients, both ranging from 0 to 1, which denote a random number (c_1) and a random vector (c_2) with a dimension equal to the number of variables in the optimization. Parameter D quantifies the difference between the study's context and the galaxy's center, whether in the front (positive) or back (negative) region. The zone parameter (z) is a random number (1 or 2) for precise positioning. To enhance accuracy, the noise parameter is applied to filter signal-related noise. A power of 3 is applied to the coefficient c_2 to minimize its computational impact, as noise levels are expected to be relatively close to the intended scenarios.

In the subsequent step, the algorithm selects one star from each previously identified region, corresponding to a stellar system, using Eq. (26) to Eq. (28). Consequently, at this stage, the algorithm has ns stars to explore. The positions of these stars are represented as L_s in Eq. (26). Notably, coefficients c_3 and c_4 in these Eqs. are random numbers ranging from 0 to 1, while the coefficient c_5 is a random vector with values in [0,1] interval.

$$L_{S,I} = L_{R,I} + D - Noise \qquad I = 1, \ldots, ns \quad (26)$$

$$D = \begin{cases} c_4 L_{R,I} - c_3 L_r & if\ \ z = 1\ (negative\ region) \\ c_4 L_{R,I} + c_3 L_r & if\ \ z = 2\ (positive\ region) \end{cases} \quad (27)$$

$$Noise = (c_5)^3 L_r \quad (28)$$

b) Transit phase: In the TSOA, categorizing stars by class is essential. Therefore, the algorithm approximates the star's luminosity using Eq. (29):

$$L_I = \frac{R_I/ns}{(d_I)^2} \qquad I = 1, \ldots, ns \quad and \quad R_I \in \{1, \ldots, ns\} \quad (29)$$

$$d_I = \sqrt{(L_s - L_T)^2} \qquad I = 1, \ldots, ns \quad (30)$$

Here, L_I represents star $I's$ luminosity while R_I denotes its rank. d_I signifies the distance among the telescope and star I. The telescope's location, L_T, is randomly selected at the outset of the algorithm and remains constant throughout optimization. To update the received light from a star, the algorithm adjusts L_s by applying Eq. (31) to Eq. (33). In these equations, coefficients c_6 and c_7 are assigned random values: c_6 ranges from -1 to 1, and c_7 is a random vector with values between 0 and 1.

$$L_{S,new} = L_{s,I} + D - Noise \qquad I = 1, \ldots, ns \quad (31)$$

$$D = c_6 L_{s,I} \quad (32)$$

$$Noise = (c_7)^3 L_s \quad (33)$$

Ultimately, the star's brightness is computed based on the newly obtained f_s using the updated $L_{S,new}$. Subsequently, the new luminosity, $L_{I,new}$, is determined according to Eq. (34).

$$L_{I,new} = \frac{R_{I,new}/ns}{(d_{I,new})^2} \qquad I = 1, \ldots, ns \quad and \quad R_I \in \{1, \ldots, ns\} \quad (34)$$

The potential for a transit event can be ascertained by comparing L_I with $L_{I,new}$. The transit probability, denoted as P_T, is determined using Eq. (35), where it takes on values of 1 (indicating a probability of transit) or 0 (indicating no transit).

If $P_T = 1$, the algorithm proceeds with the planet phase; otherwise, it executes the neighbor phase in the current iteration.

$$\begin{array}{ll} if\ L_{I,new} < L_I & P_T = 1(Transit) \\ if\ L_{I,new} \geq L_I & P_T = 0(No\ Transit) \end{array} \quad (35)$$

c) Exploitation phase: In the Exploitation phase of the TSOA, the focus shifts to the planet's characteristics and potential for hosting life. Here, L_P (L_E) pertains to the planet's attributes, including density, composition, and atmosphere. New knowledge (K) is incorporated to modify the planet's characteristics SN times (where $j = 1, \ldots, SN$) using Eq. (36) and Eq. (37). These equations involve random coefficients, such as c_{15}, c_{16}, and c_{17}, and parameter P, which specifies a random exponent between 1 and ($ns * SN$). Additionally, c_K signifies the knowledge index with 1, 2, 3, or 4 values.

The algorithm's global solution is determined by selecting the best planet among all ns detected planets.

$$L_{E,j} = \begin{cases} c_{16}L_P + c_{15}K & if c_K = 1 \quad (state1) \\ c_{16}L_P - c_{15}K & if c_K = 2 \quad (state2) \\ L_P - c_{15}K & if c_K = 3 \quad (state3) \\ L_P + c_{15}K & if c_K = 4 \quad (state4) \end{cases} \quad (36)$$

$$K = (c_{17})^P L_r \quad (37)$$

III. Results and Discussion

A. Prediction Performance Analysis

In this study, the SVR machine learning model was developed to predict HL. Furthermore, the study utilized two efficient optimization algorithms, TSOA and COA, to create hybrid SVR models, enhancing the capacity for fine-tuning model parameters. The dataset was divided into *three* subsets: train, validation, and test, with 70% of the data used for train, 15% for validation, and 15% for test [29]. The performance of these models was comprehensively assessed in Table II by comparing various metrics, including R^2 (coefficient of determination), $RMSE$ (Root Mean Square Error), MAE (Mean Absolute Error), RSR (Root Standard Ratio), and $MRAE$ (Mean Relative Absolute Error), as defined in Eq. (38) to Eq. (42) [30]:

$$R^2 = \left(\frac{\sum_{i=1}^{n}(T_i - \bar{T})(P_i - \bar{P})}{\sqrt{[\sum_{i=1}^{n}(T_i - \bar{P})^2][\sum_{i=1}^{n}(P_i - \bar{P})^2]}} \right)^2 \quad (38)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(P_i - T_i)^2}{n}} \quad (39)$$

$$RSR = \frac{RMSE}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(T_i - \bar{T})^2}} \quad (40)$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\|P_i - T_i\| \quad (41)$$

$$MRAE = \frac{1}{n}\sum_{i=1}^{n}\frac{|T_i - P_i|}{|T_i - \bar{T}|} \quad (42)$$

where, n is the number of samples, P_i and T_i are the predicted and test results, respectively. $\bar{T}$ and $\bar{P}$ are the average of the test and prediction result values.

B. Evaluation of Developed Models

The subsequent discussion provides a thorough examination of the model's effectiveness in predicting HL:

- The SVCO hybrid model demonstrated remarkable performance with maximum R^2 values of $R^2_{train} = 0.994$, $R^2_{validation} = 0.989$ and $R^2_{test} = 0.984$. These high R^2 values signify a strong fit between the model and the data, underscoring the reliability of the chosen input variables as robust predictors of the expected output. Also, for both hybrid models, R^2 for the testing stage is *lower* than that for the train stage, which indicates inadequate training performance of developed models.
- Regarding error metrics, which encompass RMSE, MAE, and MRAE, it is evident that the SVCO model *exhibits* significantly better accuracy when compared to the other models developed, demonstrating error values that are roughly half as large as those observed for the SVR single model.

TABLE II. THE RESULT OF DEVELOPED MODELS FOR SVR

Model	Phase	Index values				
		RMSE	R^2	MAE	RSR	MRAE
SVR	Train	1.575	0.977	1.363	0.155	0.225
	Validation	1.924	0.967	1.691	0.195	4.753
	Test	1.858	0.966	1.634	0.186	0.337
	All	1.676	0.973	1.453	0.166	0.255
SVCO	Train	0.861	0.994	0.607	0.085	0.098
	Validation	1.045	0.989	0.758	0.106	0.979
	Test	1.285	0.984	0.955	0.129	0.230
	All	0.964	0.992	0.682	0.096	0.117
SVTS	Train	1.201	0.988	0.896	0.118	0.134
	Validation	1.513	0.978	1.134	0.154	5.791
	Test	1.626	0.975	1.268	0.163	0.234
	All	1.322	0.984	0.987	0.131	0.160

A lower RSR value as a standard deviation ratio results in more accuracy of the model, so the least RSR value of 0.085 for $SVCO_{train}$ confirms its accurate prediction performance.

C. Comparison with Published Papers

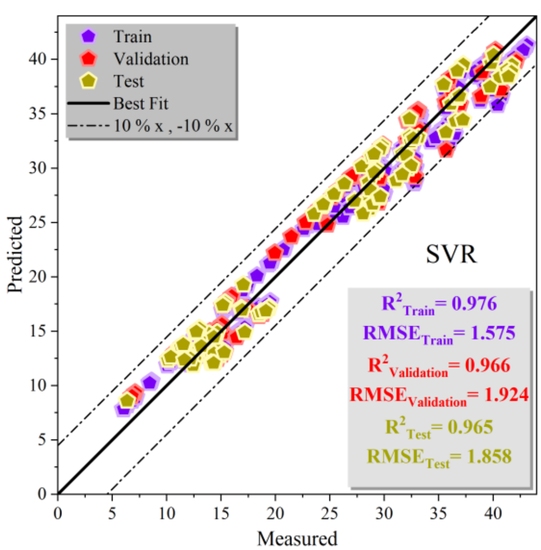
Table III shows the comparison between the presented and published papers. The comparison between the presented model and published articles focuses on key performance metrics, namely RMSE and R^2. The present study exhibits competitive performance with an RMSE of 0.964 and an R^2 of 0.992. While RMSE is higher than in some references, the R^2 aligns closely with high values in the literature. Variability in results across studies underscores the need for future research to explore factors influencing predictive accuracy. The discussion emphasizes the balance between predictive accuracy and generalization, acknowledging differences in dataset characteristics, model complexity, and optimization techniques. The comparative analysis contributes valuable insights for refining and advancing predictive models for building heat demand.

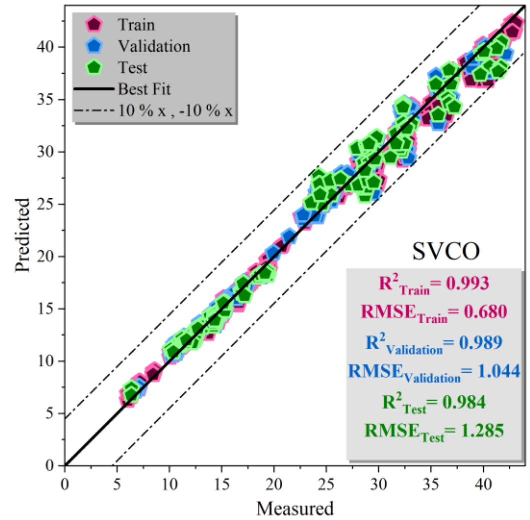
TABLE III. COMPARISON BETWEEN THE PRESENTED AND PUBLISHED ARTICLES

Articles	**Index values**	
	RMSE	R^2
Moradzadeh et al. [31]	0.4832	0.9993
Roy et al. [32]	0.059	0.99
Gong et al. [33]	0.1929	0.9882
Afzal et al. [2]	1.4122	0.9806
Present Study	0.964	0.992

D. Visualizing the Performance of Models

The association between observed and expected values for the three prediction models is shown in a scatter plot in Fig. 2. Additionally, the test, validation, and training datasets' R^2 and RMSE values for each model are supplied individually. The data points in the plot are positioned at a 45 −degree angle to the horizontal axis, about 10% above and below the bold continuous line. This alignment denotes that the models perform well in terms of prediction, which leads to greater R^2 values.





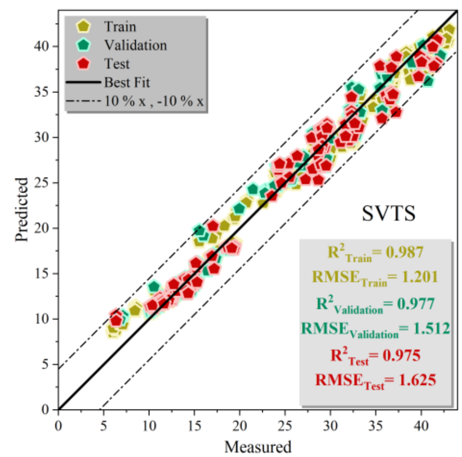


Fig. 2. Dispersion of evolved models.

The R^2 value would be one in a perfect world if all data points on the observation-prediction plot were perfectly aligned with the best-fit line. This would suggest that the model has correctly and error-free estimated all of the data. The SVCO model is shown in the illustration R^2 values of $R^2_{train} = 0.993$, $R^2_{validation} = 0.989$, and $R^2_{test} = 0.984$ for the heating loads. These numbers outperform those of the other models, showing that the SVCO model performs better in this situation than the other hybrid models.

As shown in Fig. 3, a stacked bar plot is employed in this academic study to compare various metrics comprehensively. This visualization technique offers a clear and concise representation of the relationships among different metrics by stacking them atop one another within individual bars. Each metric is assigned a distinct color, enabling a straightforward visual assessment of their contributions to the overall outcome. Fig. 2 displays the different models' calculated RMSE, R^2, and MAE values. On the basis of the measurements of RMSE and MAE, a deeper look indicates that the SVCO model exhibits reduced error rates. The SVTS and SVR models have comparably decreased error rates after SVCO. Additionally, the R^2 measures prediction accuracy, and the SVCO model surpasses the other generated models in the research in this regard.

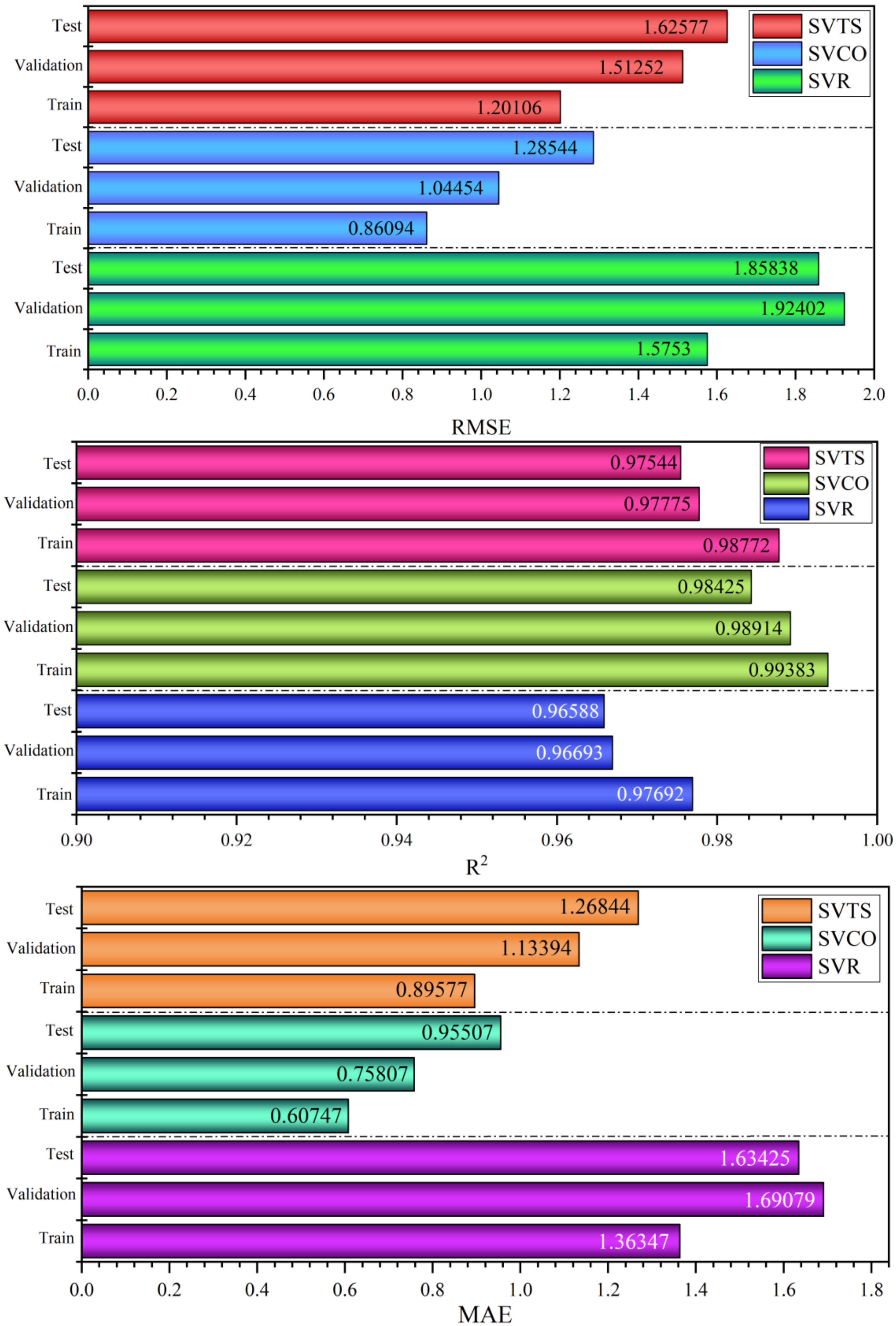


Fig. 3. Stacked bar plot for comparing the metrics.

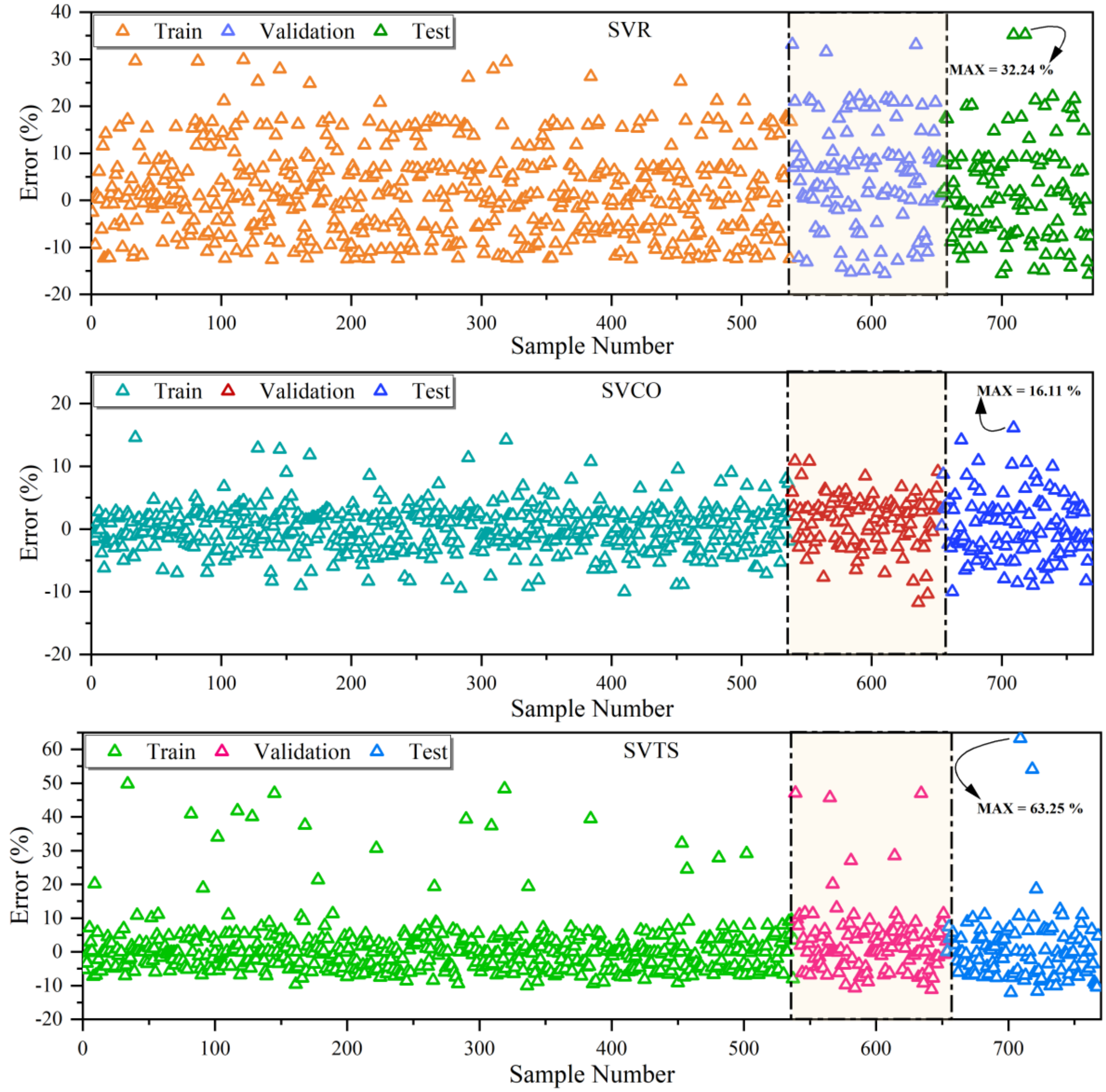


Fig. 4. Error percentage of the models based on the scatter plot.

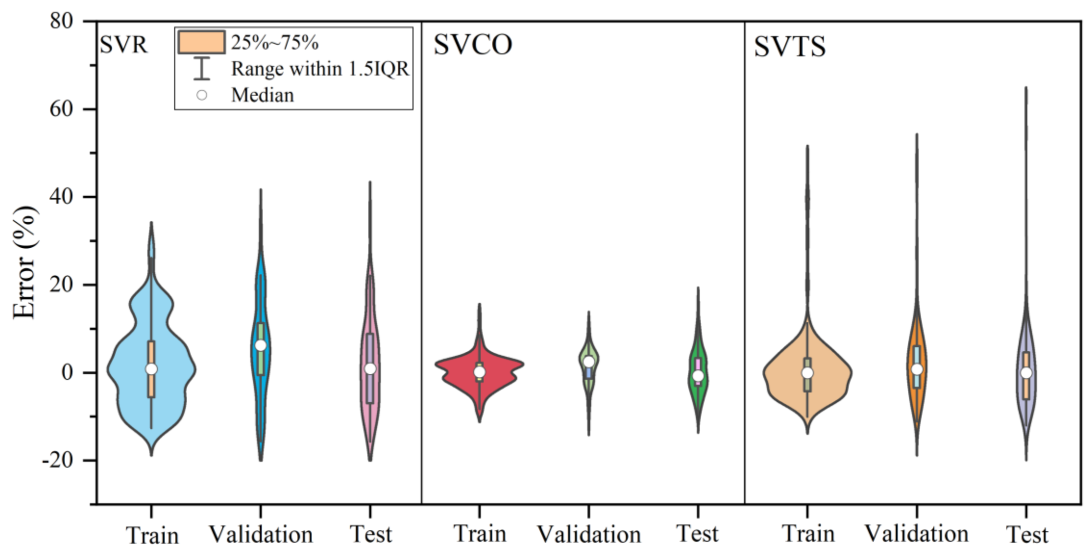


Fig. 5. The violin diagram for error percentage of proposed models.

In Fig. 4 and Fig. 5, the error percentages for the models are displayed through scatter plots and violin diagrams, with categorization into the training, validation, and test datasets. The density of data points close to zero in Fig. 3 shows how effective the strategy is. A greater concentration near zero indicates increased effectiveness. Notably, the training instances exhibit a significant preponderance of values close to zero, predominantly attributed to the SVCO method. Notably, the maximum error values for all three models are observed during the testing phase, with peak values of 32.24%, 16.11%, and 63.25% for SVR, SVCO, and SVTS, respectively. This observation underscores the beneficial fitting ability of the SVCO method.

Furthermore, Fig. 4 confirms the high accuracy of SVCO, with 25-75% of errors confined to approximately (-5 to +5). It is also evident that errors associated with SVR and SVTS are approximately three and six times higher, respectively, compared to those associated with SVCO. This further emphasizes the superior performance of the SVCO method relative to its counterparts.

IV. Conclusion

In summary, this study addresses the critical imperative for precise energy consumption forecasting and the evaluation of retrofit strategies within the framework of building energy management. The complexities in predicting building energy usage, driven by multifaceted variables, including building attributes, energy systems, weather conditions, and occupant behavior, have historically posed formidable challenges. While physics-based simulations have provided valuable insights, their accuracy hinges on data comprehensiveness and modeling intricacy. In response, this research harnesses the expanding wealth of public building energy data to explore the potential of machine learning techniques, explicitly emphasizing Support Vector Regression (SVR) models. The research findings underscore the exceptional performance of the SVR optimized with the Coot optimization algorithm (SVCO) model, consistently outperforming its counterparts by reducing prediction errors by an average of 20% to over 50% and achieving a maximum R^2 value of 0.992 for heating load prediction. This highlights the substantial potential of machine learning, as SVCO exemplifies, to significantly enhance the precision of energy consumption forecasts. Consequently, it empowers decision-makers in energy conservation and retrofit strategies, contributing to the overarching goals of sustainable building operations and reduced environmental impact. The study has several limitations. These include potential challenges in generalizing findings across diverse datasets and real-world scenarios due to a singular focus on SVR models. The reliance on datasets from previous literature introduces concerns about data quality, consistency, and relevance. The sensitivity of the SVR model to hyperparameters and the impact of optimization algorithms may also affect generalizability. The study's limited scope on heating loads may restrict its applicability to broader aspects of building energy performance. Future studies in this field could enhance predictive models by exploring multi-modal predictions, dynamic and adaptive models, and incorporating diverse datasets, including real-time sensor data. The inclusion of human behavior aspects, uncertainty analyses, and the application of models for guiding energy-efficient interventions in buildings are additional avenues for investigation. Furthermore, validating predictive models in real-world settings through field studies would improve practical applicability.

References

[1] N. Fumo, A review on the basics of building energy estimation, Renewable and Sustainable Energy Reviews 31 (2014) 53–60.

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

[3] S. Page, S. Krumdieck, System-level energy efficiency is the greatest barrier to development of the hydrogen economy, Energy Policy 37 (2009) 3325–3335.

[4] X.-N. Bui, H. Moayedi, A.S.A. Rashid, Developing a predictive method based on optimized M5Rules–GA predicting heating load of an energy-efficient building system, Eng Comput 36 (2020) 931–940.

[5] A. Yezioro, B. Dong, F. Leite, An applied artificial intelligence approach towards assessing building performance simulation tools, Energy Build 40 (2008) 612–620.

[6] A.-T. Nguyen, S. Reiter, P. Rigo, A review on simulation-based optimization methods applied to building performance analysis, Appl Energy 113 (2014) 1043–1058.

[7] D.B. Crawley, J.W. Hand, M. Kummert, B.T. Griffith, Contrasting the capabilities of building energy performance simulation programs, Build Environ 43 (2008) 661–673.

[8] H. Zhao, F. Magoulès, A review on the prediction of building energy consumption, Renewable and Sustainable Energy Reviews 16 (2012) 3586–3592.

[9] T. Catalina, J. Virgone, E. Blanco, Development and validation of regression models to predict monthly heating demand for residential buildings, Energy Build 40 (2008) 1825–1832.

[10] A. Tsanas, A. Xifara, Accurate quantitative estimation of energy performance of residential buildings using statistical machine learning tools, Energy Build 49 (2012) 560–567.

[11] Q. Li, Q. Meng, J. Cai, H. Yoshino, A. Mochida, Applying support vector machine to predict hourly cooling load in the building, Appl Energy 86 (2009) 2249–2256.

[12] A.H. Neto, F.A.S. Fiorelli, Comparison between detailed model simulation and artificial neural network for forecasting building energy consumption, Energy Build 40 (2008) 2169–2176.

[13] A. Gebremedhin, Optimal utilisation of heat demand in district heating system—A case study, Renewable and Sustainable Energy Reviews 30 (2014) 230–236.

[14] P. De Wilde, The gap between predicted and measured energy performance of buildings: A framework for investigation, Autom Constr 41 (2014) 40–49.

[15] F. Masoumi, S. Najjar-Ghabel, A. Safarzadeh, B. Sadaghat, Automatic calibration of the groundwater simulation model with high parameter dimensionality using sequential uncertainty fitting approach, Water Supply 20 (2020) 3487–3501. https://doi.org/10.2166/ws.2020.241.

[16] P.A. Gonzalez, J.M. Zamarreno, Prediction of hourly energy consumption in buildings based on a feedback artificial neural network, Energy Build 37 (2005) 595–601.

[17] B. Dong, C. Cao, S.E. Lee, Applying support vector machines to predict building energy consumption in tropical region, Energy Build 37 (2005) 545–553.

[18] G.K.F. Tso, K.K.W. Yau, Predicting electricity energy consumption: A comparison of regression analysis, decision tree and neural networks, Energy 32 (2007) 1761–1768.

[19] M. Hollander, D.A. Wolfe, E. Chicken, Nonparametric statistical methods, John Wiley & Sons, 2013.

[20] 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 01 (2023). https://doi.org/10.22034/JAISM.2023.422292.1003.

[21] M.B. Bashir, A.A. Alotaibi, Smart buildings Cooling and Heating Load Forecasting Models, IJCSNS 20 (2020) 79.

[22] M. Gong, Y. Bai, J. Qin, J. Wang, P. Yang, 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 27 (2020) 100950.

[23] À. Nebot, F. Mugica, Energy performance forecasting of residential buildings using fuzzy approaches, Applied Sciences 10 (2020) 720.

[24] A. Moradzadeh, A. Mansour-Saatloo, B. Mohammadi-Ivatloo, A. Anvari-Moghaddam, Performance evaluation of two machine learning techniques in heating and cooling loads forecasting of residential buildings, Applied Sciences 10 (2020) 3829.

[25] I. Karijadi, S.-Y. Chou, A hybrid RF-LSTM based on CEEMDAN for improving the accuracy of building energy consumption prediction, Energy Build 259 (2022) 111908.

[26] V.N. Vapnik, The nature of statistical learning, Theory (1995).

[27] I. Naruei, F. Keynia, A new optimization method based on COOT bird natural life model, Expert Syst Appl 183 (2021) 115352.

[28] M. Mirrashid, H. Naderpour, Transit search: An optimization algorithm based on exoplanet exploration, Results in Control and Optimization 7 (2022) 100127.

[29] Q.H. Nguyen, H.-B. Ly, L.S. Ho, N. Al-Ansari, H. Van Le, V.Q. Tran, I. Prakash, B.T. Pham, Influence of Data Splitting on Performance of Machine Learning Models in Prediction of Shear Strength of Soil, Math Probl Eng 2021 (2021) 4832864. https://doi.org/10.1155/2021/4832864.

[30] A. Botchkarev, Performance metrics (error measures) in machine learning regression, forecasting and prognostics: Properties and typology, ArXiv Preprint ArXiv:1809.03006 (2018).

[31] A. Moradzadeh, A. Mansour-Saatloo, B. Mohammadi-Ivatloo, A. Anvari-Moghaddam, Performance evaluation of two machine learning techniques in heating and cooling loads forecasting of residential buildings, Applied Sciences 10 (2020) 3829.

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

[33] M. Gong, Y. Bai, J. Qin, J. Wang, P. Yang, 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 27 (2020) 100950.