



OPEN

Energy consumption prediction in buildings using LSTM and SVR modified by developed Henry gas solubility optimization

Hailu Wan¹✉, Gengqiang Huang¹✉, Ying Huang^{2,4} & Noradin Ghadimi³✉

Accurately predicting building energy consumption is essential for optimizing energy management, sustainability strategies, and operational efficiency. This study proposes a novel hybrid forecasting model that integrates wavelet decomposition for feature extraction, Long Short-Term Memory (LSTM) networks for capturing temporal dependencies, and Support Vector Regression (SVR) for refined estimates, with all model parameters optimized via a Developed Henry Gas Solubility Optimization (DHGSO) algorithm. The dataset comprises two years of hourly energy consumption data from seven campuses, providing a robust foundation for validation. The proposed method achieves a 20% reduction in RMSE and a 15% reduction in MAPE compared to standalone LSTM and SVR models. This performance demonstrates the benefits of jointly leveraging decomposition-based feature engineering, deep learning, and advanced metaheuristic optimization. The results emphasize the method's potential for supporting proactive demand response, accurate budget planning, renewable energy integration, and efficient equipment maintenance in large-scale building energy management systems.

Keywords Energy consumption prediction, Building management, Long-term short-term memory, Support vector regression, Optimization, Energy management, Budget planning, Demand response, Renewable energy sources, Equipment maintenance, Peak demand control

Accurately predicting building energy consumption is vital for sustainable management, influencing multiple operational aspects¹. Such forecasts enable effective energy management by revealing usage patterns², helping managers allocate resources efficiently, identify savings opportunities, and make informed procurement and utilization decisions³. They also enhance budget planning by allowing accurate estimation of energy costs, ensuring proper resource allocation and adequate funding⁴.

Accurate energy consumption forecasts enhance demand response systems by enabling effective participation in programs that improve grid stability and offer financial incentives^{5,6}. Timely predictions support proactive load management, allowing dynamic adjustments to demand fluctuations while minimizing costs and maintaining stability⁷. They are equally vital for integrating renewable energy, as they help in properly sizing, managing, and optimizing clean energy systems^{8,9}, thereby reducing reliance on conventional sources and advancing sustainability goals¹⁰. Accurate forecasts also support efficient equipment maintenance by enabling proactive scheduling of servicing, repairs, and replacements¹¹. This approach keeps systems operating efficiently, reduces waste from faulty equipment, extends lifespan, and lowers operational costs while advancing sustainability goals¹². Likewise, forecasts improve load balancing, allowing managers to distribute energy across systems optimally, prevent overloads, and limit failures or inefficiencies¹³.

Several studies in the literature have recognized the significance of accurate energy consumption prediction in buildings for sustainable building management. The research conducted by Olu-Ajai et al.¹⁴ presented a study that several Machine learning algorithms, including Linear Regression (LR), Support Vector Machine (SVM), Deep Neural Network (DNN), Random Forest (RF), Stacking, K Nearest Neighbor (KNN), Gradient

¹Guangxi Technological College of Machinery and Electricity, Nanning 530007, Guangxi, China. ²School of Computer, Electronics and Information, Guangxi University, Nanning 530004, Guangxi, China. ³Young Researchers and Elite Club, Ardabil Branch, Islamic Azad University, Ardabil, Iran. ⁴Guangxi Economic and Trade Vocational Institute, Nanning 530021, Guangxi, China. ✉email: 63010662004@msu.ac.th; 64010363006@msu.ac.th; nghadimi1985@gmail.com

Boosting (GB), and Artificial Neural Network (ANN) were utilized in order to forecast early building electricity usage involving vast buildings in the study. The present research studied the influence of the clusters of the building on the efficiency of the model¹⁵. This paper aimed to develop a system that made designers able to accomplish vital characteristics of building strategy and predict mean amount of energy consumed in the initial steps of the extension. The findings illustrated Deep Neural Network the most effective model of prediction of the electricity consumption in the initial phase of design. Furthermore, it encouraged the designers to consume it for handling the design, making the promising decisions, and optimizing the design¹⁶.

Ma et al.¹⁷ proposed a dataset-based optimization technique for reducing energy consumption in heterogeneous mobile networks to stress the significance of using optimization techniques for energy efficiency improvements, which is complementarily tackled in the hybrid model of our work through the objective of promoting prediction accuracy through optimized algorithms.

Sun et al.¹⁸ confirmed the usefulness of deep learning-based data generation methods in the prediction of ice resistance that indicates advanced machine learning algorithms such as LSTM and SVR that can significantly enhance predictive performance in various fields, such as building energy utilization.

Ning et al.¹⁹ adopted similar techniques based on similarity optimization and applied them to calculate manufacturing costs, which is consistent with our approach to strengthen the accuracy of the required energy consumption calculations.

The role of green bonds in promoting innovation in the Chinese energy sector were also considered by Dong and Yu, reaffirming the sustainability and efficient resource allocation premise of our hybrid model for energy consumption prediction²⁰.

Wenninger et al.²¹ proposed a model, called QLattice algorithm, that was made to meet efficiency of the prediction and the possibility of explicable Artificial Intelligence (AI). More than 25,000 German buildings were involved in this study to predict yearly electricity prediction performance of buildings. The explainability, the time of computation, and the efficiency of the prediction of QLattice were juxtaposed with other founded Machine Learning algorithms, including Multiple-Linear Regression, Extreme Gradient Boosting, Support Vector Machine, and Artificial Neural Network. After analyzing the data collected, it was evident that the QLattice technology showed great potential in the field of energy performance certificates. Its impressive performance made it a promising alternative to traditional machine learning algorithms for other energy-related predictive tasks as well. Further investigation was necessary to fully explore its capabilities and potential applications.

Yang et al.²² proposed a model that its purpose was to forecast the building's electricity consumption. By the use of the chosen networks and optimization by shuffled frog-leaping algorithm, the experimental data were trained. Moreover, some criteria were examined to recognize the finest network considering speed and accuracy. In order to optimize the suggested algorithm, the rate of convergence and accomplished outcomes illustrated the significance competence. Based on the results, Support Vector Machine (SVM) and Long Short-Term Memory (LSTM) were represented, in turn, as the finest Neural Network for heating and cooling load prediction. The outcomes indicated that when it came to predicting cooling load, LSTM-SFLA performed the best with an R^2 score of 0.9761. However, SVR-SFLA indicated the best performance for predicting heating load that its R^2 score was 0.9583. the outcomes illustrated that by the use of SFLA, the performance of forecasting could be improved.

Ramos et al.²³ proposed a model that by the use of K-Nearest Neighbors and Artificial Neural Network, the most feasible algorithm of energy utilization forecasting in a building was evaluated in various settings. The mentioned algorithms utilized data patterns of utilization that were combined in various settings maintaining supplementary information from data of sensors at the same time. The various settings were categorized in an order of steps that happened every five minutes. A decision tree was used to determine which of the two forecasting algorithms was appropriate for every five-minute interval. The algorithm that appeared to be the best suited was selected, and a reasonable explanation was given to confirm if it was the optimal choice. Studying updates in parameterization related to depth was important to grasp the effects they had on accuracy of forecasting. Decision trees had the capability to enhance prediction accuracy, as they had crucial role in processes of decision-making.

Moon et al.²⁴ proposed a model named RABOLA (Ranger -Based Online Learning Approach), which was a strong two-step model of prediction. The purpose was to permit patterns of quick and practical learning work for the sake of some data, which were unseen. Energy usage data of two office buildings, which were available in public. Next, we processed the data and configured input variables to create training and test sets. In the beginning step, three STL models were designed in accordance with tree-based ensemble learning methods utilizing the set of training. In the second step, a ranger-based model of prediction was made using size of window sliding during seven days utilizing forecasted values for external elements and three designs, like temperature and timestamp, which were counted as input parameters on the test. By wide comparison analysis, it could be illustrated that the suggested model could outperform forecasting efficiency of the other deep learning and stacking ensemble methods considering variation coefficient and mean absolute percentage error of the root-mean square error. Additionally, the relations between output and input parameters in forecasting building energy usage by STL was presented. Table 1 synthesizes key features, datasets, evaluation metrics, and limitations of representative prior works discussed in the literature review. As shown, none of these approaches successfully integrates multi-resolution feature decomposition (wavelet transform), a deep sequential learner (LSTM), and a nonlinear regressor (SVR) optimized via a metaheuristic specifically tailored for hybrid model hyperparameter tuning. This integration, combined with proven accuracy gains, distinguishes the present study from existing methods.

In spite of the studied works in the literature, there is a pressing need to investigate fresh approaches that, when used correctly, may significantly improve the accuracy of energy consumption forecast in buildings.

Reference	Methodology/model	Dataset characteristics	Evaluation metrics	Main limitations/advantages
²²	SVM and LSTM optimized with Shuffled Frog-Leaping Algorithm (SFLA)	Single building electricity data (heating & cooling loads)	RMSE, R ²	Performance tuned for single-site data; lacks multi-building generalization
²³	K-Nearest Neighbors + ANN with Decision Tree selection	Sensor-based, 5-min interval building usage data	MAE, RMSE	Focused on short interval forecasts only; decision tree adds selection overhead
²⁴	Ranger-Based Online Learning (RABOLA)	Energy usage from two office buildings, public dataset	CV, MAPE	Limited to office buildings; no multi-resolution feature extraction
³²	Model Integration of classic statistical & ML approaches	Mixed building energy datasets	RMSE, MAPE	Less accurate for highly non-stationary datasets
³³	Vector field-based SVR	Hourly energy consumption datasets	RMSE, MAPE	Sensitive to kernel parameter choice; limited handling of long-term dependencies
²⁵	Genetic Algorithm enhanced Adaptive DNN	Public building use datasets with environmental variables	RMSE, R ²	Overfitting risk; computationally expensive
³⁵	Deep Reinforcement Learning	Real-time EMS data streams	RMSE	Needs continuous retraining; high data dependency
³⁶	CNN feature extraction + Bi-LSTM	Energy datasets with multi-seasonal trends	RMSE, MAPE	Requires high computational resources; complex to deploy
³⁷	Gradient Boosting Regression Tree	Historical building energy readings	RMSE, MAE	Lower accuracy for abrupt load changes
³⁸	Improved Extreme Gradient Boosting model	Multi-building datasets	RMSE, MAPE	Less robust for multi-scale temporal data
This Study	Wavelet decomposition + LSTM + SVR optimized via DHGSO	Two years of hourly consumption data from seven campus buildings	RMSE, MAPE, MAE, R ²	Advantages: Captures both short- and long-term dependencies; integrates multi-resolution feature extraction and nonlinear refinement; domain-specific metaheuristic tuning (DHGSO) yields 20% RMSE and 15% MAPE improvement over strong baselines; scalable to multi-building, multi-type datasets

Table 1. Summary of prior studies on Building energy consumption forecasting.

Forecasting the energy consumption in buildings is a complex task, due to the need to capture temporal dependencies and the non-linear interactions in high-dimensional, noisy data. Such intricacies often pose challenges to traditional approaches, resulting in poor predictions that impair robust energy management. To handle these challenges, a hybrid technique is proposed using wavelet decomposition for features extraction based on multiple resolution and adjusting model for global temporal data using long short term memory (LSTM) and non-linear behavior using support vector regression (SVR). Additionally, the DHGSO algorithm designed by the novel Developed Henry Gas Solubility Optimization algorithm ensures optimal hyperparameter tuning by optimizing the hyperparameters related to both LSTM and SVR, thereby improving the prediction accuracy and robustness. The comprehensive results confirm that this novel combination not only enhances forecasting precision but also provides a versatile modeling framework that can adapt to varied energy consumption patterns, differentiating it from traditional techniques and conveying advantages of accuracy, adaptability, and scalability in building energy management.

Research gap and novelty

Despite notable advances in building energy consumption forecasting, several key challenges remain unresolved. Many existing models either excel at short-term temporal pattern recognition (e.g., deep recurrent networks) or at modeling nonlinear relationships (e.g., kernel-based regressors), but few can effectively integrate both capabilities without overfitting or sacrificing computational efficiency. Additionally, hyperparameter tuning in hybrid models is often performed using generic algorithms that do not guarantee optimal search across complex parameter spaces. These limitations hinder the robustness, adaptability, and precision required for real-world deployment in multi-building energy management systems.

To address these gaps, we propose a hybrid framework combining wavelet decomposition for multi-resolution feature engineering, an LSTM network for capturing temporal dependencies, and an SVR model for nonlinear refinement, with all components tuned via a DHGSO algorithm. This integrated approach not only captures both short- and long-term dependencies but also ensures optimized performance through domain-specific parameter tuning, distinguishing it from conventional hybrid modeling strategies. Therefore, the key contributions of this study can be summarized as follows:

- i) Development of a novel integration of wavelet decomposition, LSTM networks, and SVR optimized via the DHGSO algorithm for building energy consumption forecasting.
- ii) Use of wavelet-based multi-resolution decomposition to extract both short- and long-term temporal patterns, enhancing the predictive capabilities of sequential learning models.
- iii) Introduction of DHGSO for domain-specific hyperparameter tuning, achieving significant improvements in accuracy (RMSE and MAPE reduction) over baseline LSTM and SVR approaches.
- iv) Demonstration of the model's adaptability across multiple building types in a multi-campus dataset, highlighting its potential for real-world deployment in advanced energy management systems.

Dataset description

This research comprises of information obtained from on-site visits carried out over a period of two years, spanning from February 2014 to April 2016, at seven distinct campuses. The primary objectives of these expeditions encompassed two fundamental aspects: firstly, the acquisition of unprocessed temporal data from the locations, and secondly, the engagement in discussions concerning the present methodologies employed in constructing energy analysis within these academic establishments. This segment presents a comprehensive summary of the on-site visits, delineating the diverse categories of data amassed during these excursions and presenting significant observations from the endeavor.

Upon accomplishment of location inspections, it shows that educational institutions had made significant investments in power metering and data collection systems over the past decade. Nevertheless, it was observed that the data required adequate utilization. Numerous universities and colleges have expressed concerns regarding the abundance of meter data, coupled with the need for the necessary expertise and resources to effectively employ it for analytical purposes. Notwithstanding several instances of visiting and subsequent attempts to gather information, the data remained unattainable for the purpose of doing study. However, the data about the five educational institutions and their performance on the seven benchmarks was easily accessible and has been evaluated in this study.

The first case study is located in the Midwest area of the United States, characterized by a continental climate. The college consists of 226 buildings that are distributed across two main educational institutions, with a combined floor space that exceeds 2.3 million square meters (25 million square feet). The data gathering process for Case Study 1 began with an initial interview conducted in March 2015. Subsequently, a site visit was conducted in June 2015 to extract raw data spanning a period of one year from all electricity meters. This extraction was facilitated by the use of SQL databases. The meta-data file includes supplementary data, such as location, primary space usage, floor area, and EnergyStar score.

The second case study is located in the Northeastern area of the US and focuses on a university that has a major campus housing 180 entities. The data gathering approach for this case study included an initial meeting in April 2015, which was then followed by a site visit in August 2015. During the site visit, a database query was used to pull a year's worth of electricity meter data from the buildings. The given meta-data file included further details on the floor space and major use type of the buildings.

The third case study is situated within the Midwest region of the United States and has a university campus consisting of twenty-five buildings, covering a total area of $204,000\ m^2$ (2.2 million square feet). The data gathering procedure for this particular case study included an initial examination of the location and subsequent deliberation in March 2015, succeeded by a visit to the site in March 2016 to get unprocessed data from the energy management platform used on the campus. The metadata of the energy management platform facilitated simple retrieval of flat files including each data point.

The fourth case study is situated inside the confines of a tropical international school in Southeast Asia. The school has five buildings, covering an estimated area of 58,000 square meters (625,000 square feet). The data included in this case study was collected through a series of continuous talks and interviews conducted with the operations employees over a span of five years. The operations team played a crucial role in the development of the technique, providing significant contributions.

The fifth case study is located in Switzerland and attention on a university campus consisting of 22 buildings, covering an area of $150,000\ m^2$ (1.6 million square feet). The data for this case study was obtained via email correspondence with the building facilities managers and by leveraging raw data from the campus energy management system. The meta-data spreadsheet yielded more information on the fundamental purposes of the spaces inside all of the structures.

To ensure data quality and model readiness, all input time series underwent a two-stage preprocessing procedure:

Missing data imputation

- Short Gaps (< 2 h): Imputed via weighted linear interpolation using adjacent timestamps, preserving short-term temporal trends.
- Medium Gaps (2–24 h): Filled using daily pattern matching, taking the average of the same hour from the previous and following day.
- Long Gaps (> 24 h): Replaced with the mean profile of the corresponding day type (weekday/weekend) from the same building, ensuring seasonality consistency. All imputation operations were validated against raw EMS logs to avoid synthetic bias.

Normalization

All predictive variables (e.g., consumption, temperature, humidity) were scaled to a [0, 1] range using Min–Max normalization. This scaling facilitates stable convergence of the hybrid LSTM–SVR model and makes features comparable across buildings and seasons. The preprocessing scripts were implemented in MATLAB R2017b with custom validation functions to log all modifications, ensuring reproducibility.

Input feature set The predictive framework is based on a combination of endogenous and exogenous variables:

- Historical load features: Hourly electricity consumption measurements for the preceding 24 h ($t - 24$ to $t - 1$), enabling the model to capture recent autocorrelations and short-term consumption dynamics.

- ii) Meteorological variables: Hourly ambient temperature (°C) and relative humidity (%) data sourced from the nearest official meteorological station. These were synchronized with the load data to match timestamps precisely.
- iii) Temporal indicators: Dummy (0/1) encodings for day-of-week (Monday–Sunday) and time-of-day (hour 0–23), allowing the model to account for weekly seasonality and daily load cycles.

All features were aligned using a unified timestamp index and subjected to the preprocessing procedures described above (imputation and normalization). This multi-source feature set supports the model's ability to leverage both autoregressive patterns and environmental drivers of building energy demand.

Data decomposition based on wavelet decomposition

One of the significant mathematical transformations that is used in a variety of scientific subfields is known as the wavelet transform. The Fourier transform has a number of shortcomings and restrictions, which are the primary reasons for the development of the wavelet transform. This transformation, in contrast to the Fourier transform, is suitable for usage with non-stationary signals and dynamic systems.

The utilization of wavelet decomposition is broad in data feature extraction. Its primary objective is to extract significant information from the data that comprise both high-frequency and low-frequency components. This approach offers considerable compensations for scrutinizing time-series data that exhibit diverse frequencies and capturing specific properties.

Decomposition starts by subjecting the original data through a series of low-pass and high-pass filters, thus facilitating the segregation of the low-pass and high-pass constituents. The low-pass filter is designed to retain low-frequency components, while the high-pass filter is intended to separate high-frequency information. The first decomposition is often known as the first level or approximation.

The approximation coefficients, also known as the low-frequency component, are derived from the original signal and represent the coarsest degree of information. The coefficients encapsulate the general trends and lower-frequency elements of the signal. On the other hand, the detail coefficients, which represent the high-frequency component, are responsible for capturing localized features, abrupt variations, and higher-frequency data.

The process of decomposition involves the additional breakdown of the approximation coefficients acquired from the preceding stage. The extraction of complicated frequency information is achieved through implementing equivalent filtering techniques to the approximation coefficients. At every stage of decomposition, new approximation and detail coefficients are acquired, which signify information at different scales.

In the present scenario, the wavelet decomposition technique is employed to partition the time series of building energy into distinct components, namely approximation coefficients and detail coefficients, which are obtained at different hierarchical levels. The number of decomposition levels is dependent on the desired degree of granularity necessary to effectively depict the data. The dataset denoted as ($x[n]$), is a sequence of values indexed by the variable n , which represents the time index.

Step (1) The process of wavelet decomposition commences with the application of a low-pass filter ($h[n]$) and a high-pass filter ($g[n]$) to the main data.

$$a_0[n] = x[n] * h[n] \text{ (Approximation coefficients at level 0)} \quad (1)$$

$$d_0[n] = x[n] * g[n] \text{ (Detail coefficients at level 0)} \quad (2)$$

where, convolution operator is defined by the (*).

Step (2) Apply down sampling by factor two on the detail and approximation coefficients achieved in the previous step to get coefficients at next step.

$$a_1[n] = a_0[2n] \text{ (Approximation coefficients at level 1)} \quad (3)$$

$$d_1[n] = d_0[2n] \text{ (Detail coefficients at level 1)} \quad (4)$$

The downsampling process is responsible for decreasing the resolution of the coefficients. However, it also has the ability to capture lower-frequency trends in the approximation coefficients and higher-frequency features in the specific coefficients.

3) iterate step (1) and step (2) for several decomposition levels pending the detail desired level

$$a_{i+1}[n] = a_i[2n] \text{ (Approximation coefficients at level } i+1) \quad (5)$$

$$d_{i+1}[n] = d_i[2n] \text{ (Detail coefficients at level } i+1) \quad (6)$$

Such that i specifies the level of decomposition.

At each level, the decomposition procedure yields detail coefficients ($d_i[n]$) and approximation coefficients ($a_i[n]$). The detail coefficients are responsible for representing the high-frequency details and localized characteristics, whereas, the approximation coefficients serve to capture the low-frequency components of the data and provide a rudimentary representation of the original data.

Here, Haar wavelet is considered as mother wavelet function to generate the wavelet filters ($h[n]$ and $g[n]$) employed in the decomposition process. Through the use of the wavelet decomposition framework, it becomes

possible to extract significant characteristics at different scales or resolutions. This facilitates a more accurate and thorough examination in forecasting the energy use of buildings.

Specifically, the Haar wavelet has been chosen since it is simple, efficient to compute, and well-suited for discontinuities in time-series data like energy consumption data. The Haar wavelet is the simplest and oldest wavelet, and it splits data into piecewise constant all over the pieces, hence it suits non-stationary signals very well for separating high-frequency details from low-frequency trends. The Haar wavelet offers competitive performance for feature extraction with minimal computational overhead relative to other widely adopted wavelets, such as Daubechies or Symlets making it suitable for practical diagnostics on large datasets. Furthermore, it has a simple mathematical formulation that allows for a faster implementation while maintaining the quality of features extracted. On the other hand, the choice of decomposition levels can have a significant effect on prediction accuracy. However, more decomposition level means a finer separation between frequency components, helping the model to learn complex temporal dependencies. But too deep decomposition may cause noise or over-fitting, especially where data is small. On the other hand, too few decomposition levels may not catch important patterns, leaving underfitting. Consequently, a balance has to be found such that the level of decomposition selected corresponds to the properties inherent to the data, as well as the accuracy of prediction.

The model of the hybrid LSTM and SVR model

The amalgamation of Long Short-Term Memory (LSTM) and Support Vector Regression (SVR) can be employed for the purpose of modeling and forecasting building energy demand consumption. Here, LSTM as a variant of recurrent neural network (RNN), has demonstrated its efficacy in capturing long-term dependencies in sequential data, whereas SVR is a machine learning algorithm that is utilized for regression tasks.

The combination of LSTM and SVR methodologies has the potential to improve the accuracy of estimation. This is because SVR is effective in capturing linear correlations in the data, while LSTM is capable of capturing non-linear correlations and temporal dependencies. The optimization of the model may be achieved by the use of a modified approach known as the Developed Henry Gas Solubility Optimization algorithm. The use of the Developed Henry Gas Solubility Optimization algorithm in optimizing the LSTM/SVR model has been shown to enhance both the resilience and accuracy of the model, resulting in increased effectiveness in the prediction process.

SVR (support vector regression)

Given a training dataset consisting of input features $X = [x_1, x_2, \dots, x_n]$ and corresponding target values $Y = [y_1, y_2, \dots, y_n]$, SVR aims to find a function $f(x)$ that predicts the target value y for a given input x .

Assuming a linear relationship between the input features and the target variable, the basic formulation of SVR can be defined as follows:

$$y = w \cdot x + b \quad (7)$$

where, the variable w denotes the weights or coefficients assigned to each input feature, whereas b represents the bias term. The objective of Support Vector Regression (SVR) is to identify the ideal values for the weight vector (w) and the bias term (b) that minimize the discrepancy between the predicted values and the actual values.

In order to accommodate non-linear interactions and effectively address intricate patterns, Support Vector Regression (SVR) use a modified version of the linear formulation, which incorporates kernel functions. The kernel technique involves the implicit transformation of input characteristics into a higher-dimensional feature space, enabling the possibility of linear separation.

The revised formulation of Support Vector Regression (SVR) may be expressed in the following manner:

$$y = w \cdot \Phi(x) + b \quad (8)$$

where, the symbol $\Phi(x)$ denotes the feature vector that has been translated into a higher-dimensional space. Here, Radial Basis Function (RBF) Kernel function is used and formulated below:

$$K(x, x') = \exp(-\gamma \cdot \|x - x'\|^2) \quad (9)$$

The aforementioned calculations use hyperparameters γ , r , and d , which are kernel-specific. The model's ability to represent non-linearity and complexity is determined by the extent of control exerted over these factors.

In order to formulate the SVR as an optimization problem, it is necessary to include a margin around the projected outputs. The objective of the SVR is to reduce the error inside a specified margin, while simultaneously increasing the width of the margin. The formulation of the optimization issue for the SVR may be expressed as:

$$\text{minimize} \left(\frac{1}{2} \right) \times \|w\|^2 + C \times \sum (\epsilon_i + \epsilon_i^*) \quad (10)$$

subject to:

$$y_i - f(x_i) \leq \epsilon_i, f(x_i) - y_i \leq \epsilon_i^* \quad (11)$$

In the aforementioned formulation, the term $\|w\|^2$ denotes the L_2 norm of the weight vector w . This particular norm is used to regulate the complexity of the model and mitigate the risk of overfitting. The regularization parameter, denoted as C , plays a crucial role in finding an optimal balance between increasing the margin and

decreasing the training error. The expression $\sum (\epsilon_i + \epsilon_i^*)$ denotes the summation of slack variables, which enable some samples to either reside inside the margin or exceed the margin.

After obtaining the Lagrange multipliers, the identification of support vectors is carried out by considering the data points that possess non-zero multipliers. The support vectors are of utmost importance in establishing the regression model.

Upon successfully solving the optimization problem and acquiring the support vectors, the prediction for a novel input x may be produced by using the following equation:

$$f(x) = \sum (\alpha_i \cdot K(x_i, x)) + b \quad (12)$$

where, α_i represents the Lagrange multipliers associated with the support vectors, and b is the bias term.

In this study, to optimize SVR, the decision variables, includes γ , r , and d that are selected to minimize Eq. (10).

Long short-term memory (LSTM)

The Long Short-Term Memory (LSTM) is a particular design of a recurrent neural network (RNN) that has been specifically developed to effectively address the challenge of handling long-term dependencies and effectively capturing sequential patterns within data. The introduction of this technique may be attributed to Hochreiter and Schmidhuber in 1997, and it has since gained significant popularity in a range of domains, such as natural language processing, voice recognition, and time series analysis.

LSTM networks consist of LSTM cells, which include an internal memory state capable of retaining information through extended sequences. The fundamental concept behind Long Short-Term Memory (LSTM) is the incorporation of gating mechanisms that regulate the information flow inside the cell. These mechanisms enable the cell to choose retain or discard information at various time intervals. The model of the LSTM is given below:

Input gate: the input gate is responsible for determining the appropriate amount of new information that should be stored in the cell state. The process involves the integration of the present input with the preceding hidden state, followed by the application of a sigmoid activation function to produce an update gate (i^t).

$$i^t = \text{sigmoid}(W_{\text{in}}^p \odot x_t + U_{\text{in}} \odot h_{t-1} + b_{\text{in}}) \quad (13)$$

Forget Gate: The forget gate is responsible for determining which information from the preceding cell state should be disregarded. The model integrates the present input with the preceding hidden state and employs a sigmoid activation function to produce a forget gate (f^t).

$$f^t = \text{sigmoid}(W_f \odot x_t + U_f \odot h_t^{-1} + b_f) \quad (14)$$

Cell State Update: The process of updating the cell state (C_t) involves the integration of the new input with the old cell state via the use of the input gate. Additionally, the forget gate is used to eliminate obsolete information.

$$C_t = f^t \odot C_t^{-1} + i^t \odot \tanh(W_c \odot x_t + U_c \odot h_t^{-1} + b_c) \quad (15)$$

Output Gate: The output gate is responsible for determining the extent to which the cell state information is used in generating the present hidden state (h_t). The process involves the integration of the present input with the preceding hidden state, followed by the application of a sigmoid activation function.

$$y_t = \text{sigmoid}(W_o \odot x_t + U_o \odot h_t^{-1} + b_o) \quad h_t = y_t \odot \tanh(C_t) \quad (16)$$

where, x_t represents the input at a specific time step, t . The previous hidden state is denoted as h_{t-1} , while C_{t-1} represents the previous cell state. Weight matrices W and U are utilized for the transformation of inputs, and b is a bias vector used in the calculations. The symbol \odot indicates element-wise multiplication, which is a component-wise operation between two matrices or vectors. Sigmoid refers to the sigmoid activation function, a commonly used nonlinear function that maps values to a range of 0 to 1. On the other hand, tanh represents the hyperbolic tangent activation function, which maps values to a range of -1 to 1 with a stronger gradient compared to sigmoid.

The Long Short-Term Memory architecture facilitates the propagation of gradients through time, thereby enabling the model to effectively capture long-term dependencies. This feature addresses the issue of disappearing gradients that is often observed in conventional Recurrent Neural Networks (RNNs). LSTM networks possess the ability to acquire intricate patterns and interdependencies in sequential data through the use of mathematical formulations and the continuous adjustment of internal states. As a result, they are highly suitable for tasks that involve temporal dynamics and extensive dependencies.

When training a Long Short-Term Memory (LSTM) model, the fitness function is to quantify the discrepancy between the predicted outputs of the LSTM model and the actual desired outputs in the training data. The model is then optimized by minimizing this discrepancy. In this study, Mean Squared Error is used as fitness function. The fitness function can be mathematically defined as follows:

The Mean Squared Error (MSE) is well recognized as a prominent loss function used in regression tasks. Its purpose is to compute the average of the squared discrepancies between the predicted values and the corresponding actual values. The mathematical definition of MSE is as follows:

$$MSE = \frac{1}{N} \times \sum (y - \hat{y})^2 \quad (17)$$

where, N describes the sample numbers, y and \hat{y} specify the actual output, and \hat{y} is the predicted output.

The main purpose of this study, is to minimize Eq. (17).

By combination of the LSTM with the SVR, the model for energy prediction can be achieved. Figure 1 showcases the integrated setup of the SVR/LSTM model.

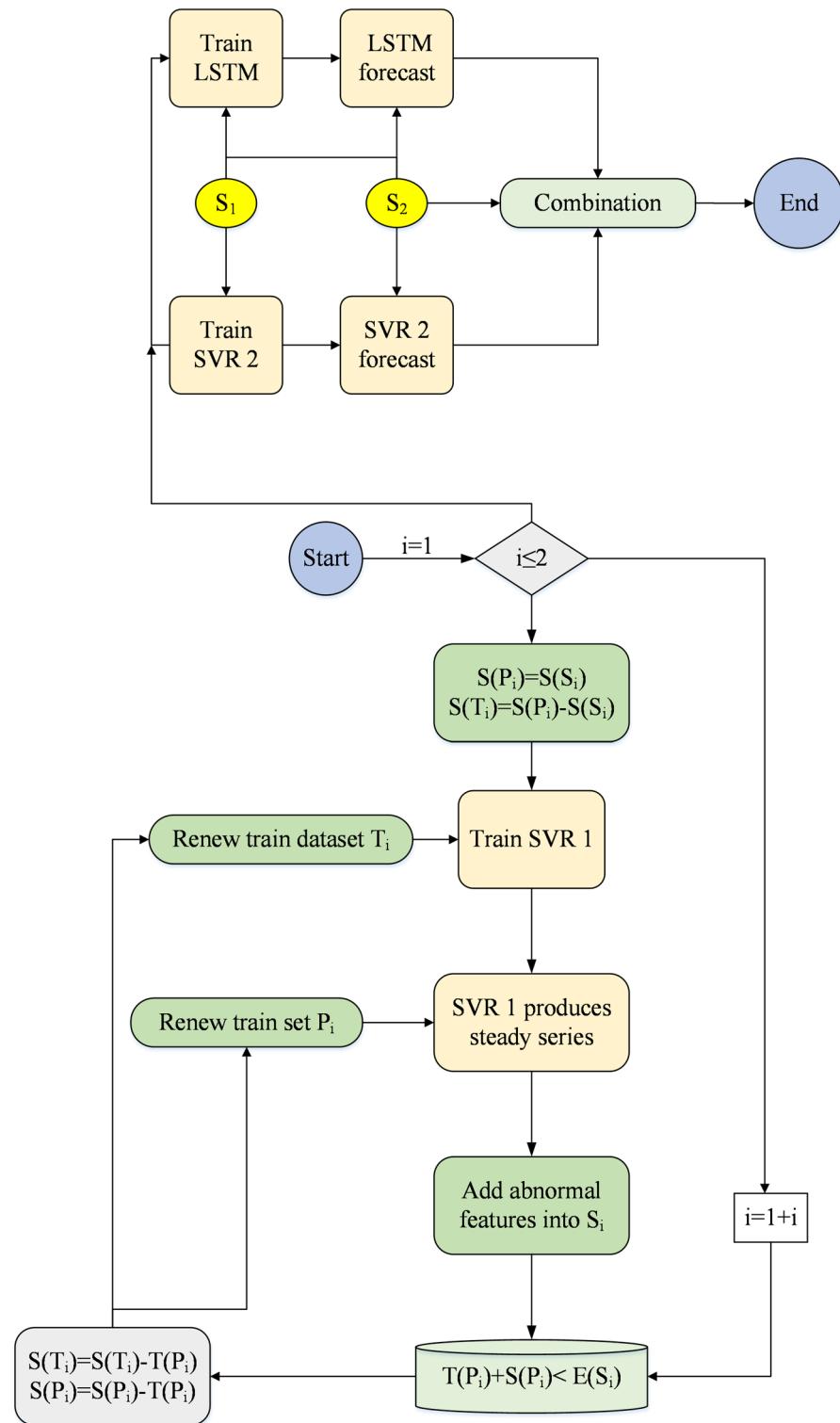


Fig. 1. Showcases the integrated setup of the SVR/LSTM model.

The graphic illustrates the use of periodic qualities as inputs for the SVR 1s. This enables the computation of the fundamental quantities of load in an order. The approach being discussed incorporates current and well-documented statistical data as temporal references. The use of atypical characteristics as the inputs for Support Vector Regression (SVR), specifically SVR 2, and Long Short-Term Memory (LSTM) models involves the incorporation of balanced and temporal frames as examples. The Oz1 and Oz2 variables represent the obtained results derived from the Support Vector Regression (SVR) and Long Short-Term Memory (LSTM) models, respectively. The predictive outcome of the SVR-LSTM model may be expected by using the combination of Oz1 and Oz2, as mentioned in reference²⁶.

The proposed hybridization with Long Short-Term Memory (LSTM) and Support Vector Regression (SVR) is crafted to use the complimentary aspects of both methods; capturing long short-term temporal dependencies and linearity in the feature space. Since LSTM is a variant of recurrent neural networks, LSTM can deal with sequential data well due to its advantage at modeling long-term sequential dependencies and nonlinear relationships from time series energy consumption data. This renders it extremely salient across complex temporal patterns commonly observed in building energy consumption. Contrast this with (SVM), an efficient version of (SVR) that works effectively on data with linear correlations, bringing in stable regression output especially when the correlation between the input features and output targets can be approximated as a linear line. This makes it possible for the model to utilize the LSTM's advantage in capturing complex patterns through time, and the strength of SVR to focus on the linear trends and reduce noise which is important for better prediction, especially in cases where a balance is needed.

SVR is used in the final prediction stage instead of LSTM alone, as it tends to provide stable and interpretable outputs even when the data is scarce or overfit with purely neural network-based methods. LSTM learns the underlying temporal structure, whilst SVR serves as a smoothing mechanism to drive the final predictions close to observed linear trends without overfitting to outliers or anomalies. Moreover, SVR's kernel-based framework adds such flexibility for emphasizing nonlinearities if necessary, improving its appropriateness for the final prediction stage. In summary, this two-phase approach allows for dynamic and effective dispatch modeling while conserving linear trend modelling fidelity.

As illustrated in Fig. 2, the preprocessing pipeline begins with data acquisition from both the building's EMS and meteorological sources. Missing data are imputed according to gap duration: short gaps via weighted linear interpolation, medium gaps by matching daily patterns, and long gaps through mean profile substitution. Subsequently, features from different sources are synchronized and aligned. Finally, Min-Max normalization is applied to ensure scale consistency before the modeling phase.

Developed Henry gas solubility optimization algorithm

Henry gas solubility optimization algorithm

The proposed Henry gas solubility optimization algorithm's mathematical models are determined in the current segment. The numerical stages are introduced in the following:

Phase 1: Setting up Process. The number and place of gases (the population's size (M)) are set up using the next formula:

$$Y_j(t+1) = Y_{min} + a \times (Y_{max} - Y_{min}) \quad (18)$$

the j_{th} gas's place is defined by $Y_{(j)}$ in M (members), a is a random number that ranges from 0 to 1, problem boundaries are defined by Y_{min} and Y_{max} , and the repetition period is determined by t . $\nabla solE/R$ invariable amount of kind i (E_j), gas j in group i 's partial pressure $\rho_{j,i}$, amounts of Henry's invariable of kind i ($HE_i(t)$), and The number of gas j are set up with employing the following equation:

$$HE_i(t) = d \times e, \quad \rho_{j,i} = d \times f, \quad E_i = d \times c \quad (19)$$

In which, e , f , and g are regarded as invariable amounts. In addition, e , f , and g are the stochastic numbers that are limited from 0 to 1.

Phase 2: Gathering. The identical groups that are the same as the number of gas types are separated by the population's agents. There are equal gases in every group; consequently, every group owns the same invariable amount of Henry (HE_i).

As the pressure increases, Fig. 3 shows that more gas particles dissolve in order to restore stability.

Phase 3: Calculation. Each group i is measured to identify the finest gas that obtains the greatest balance situations in comparison with similar ones. Later, to obtain optimized gas in the total group, the gases are classified.

Step 4: Renew the Factor of Henry. The factor of Henry is renewed in a match using the next equation:

$$HE_i(t+1) = HE_i(t) \times \exp\left(-E_i \times \left(\frac{1}{T^\theta}\right)\right), \quad T(t) = \exp(-t/rep) \quad (20)$$

The factor of the Henry for group i is defined by HE_i , T defines the temperature, and T^θ is the invariable number of temperatures, and the total number of repetitions is determined by rep .

Step 5: Renew solubility. The solubility is renewed in a match by the next equation:

$$SL_{j,i} = R \times HE_i(t+1) \times P_{j,i}(t) \quad (21)$$

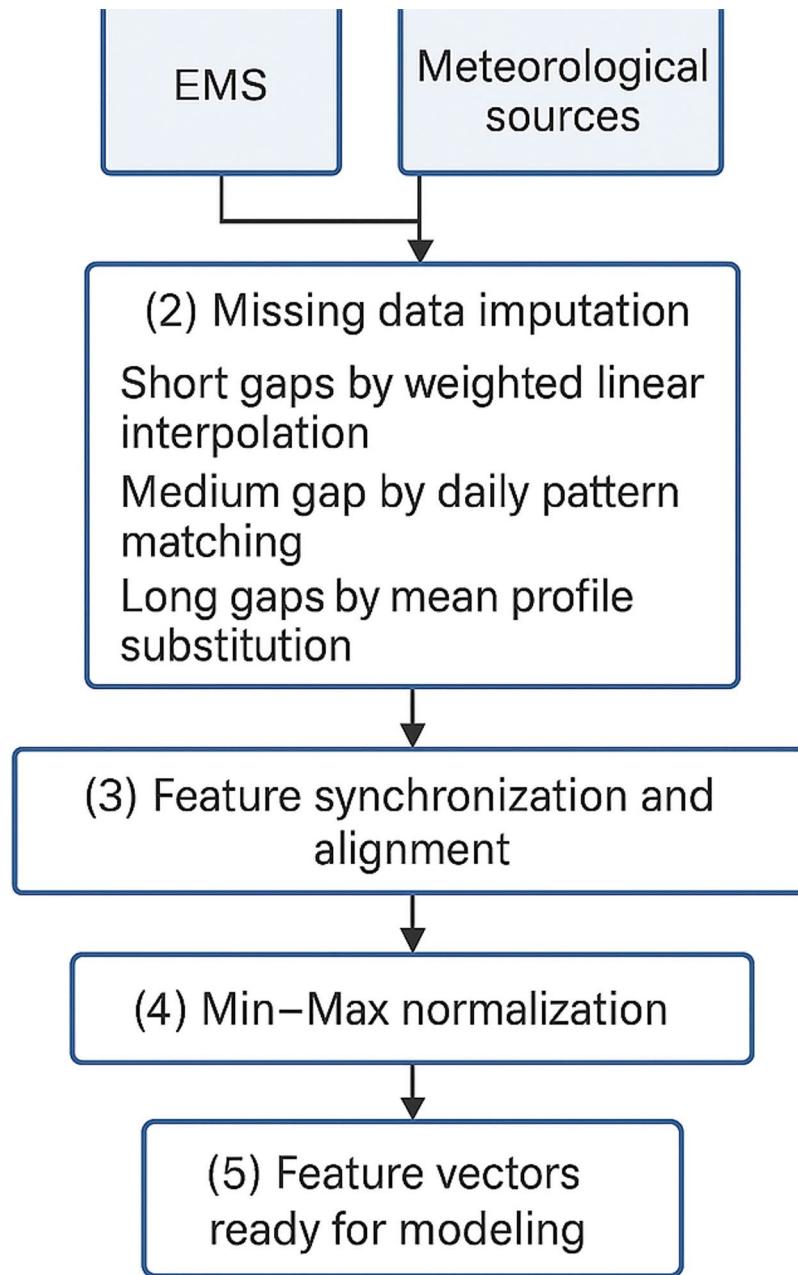


Fig. 2. Preprocessing pipeline for multi-source building energy and meteorological data.

In which, the gas j solubility in group i is signified with $SL_{j,i}$, the partial pressure on gas j and in group i is signified by $\rho_{j,i}$, and the invariable number in this equation is depicted by R .

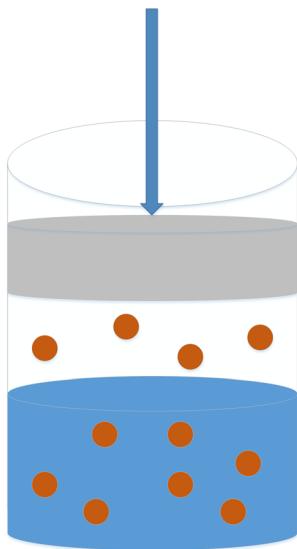
Step 6: Position renewing. The place is renewed in the next formula:

$$\begin{aligned}
 Y_{j,i}(t+1) = & Y_{j,i}(t) + F \times a \times \gamma \times (Y_{j,greatest}(t) - Y_{j,i}(t)) \\
 & + F \times a \times \epsilon \times (SL_{j,i}(t) \times Y_{j,greatest}(t) - Y_{j,i}(t))
 \end{aligned} \tag{22}$$

$$\gamma = \psi \times \exp \left(-\frac{F_{best}(t) + \beta}{F_{j,i} + \beta} \right), \beta = 0.5$$

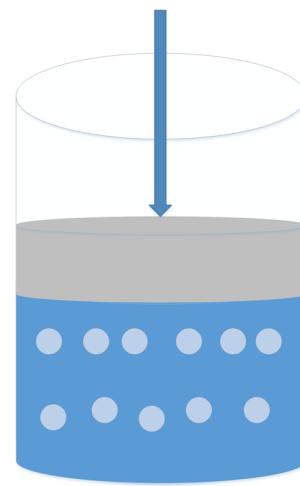
In which, $Y_{(j,i)}$ denotes the gas j in group i 's place, the repetition period and a stochastic invariable are determined by t and a . The greatest gas in the group is determined with $Y_{greatest}$, whilst the greatest gas j in group i is signified by $Y_{(j,greatest)}$. Additionally, the ability of gas i in group i to connect with additional gases in the identical group is depicted by γ , the influence of additional gases on gas j in group i is illustrated with ϵ that its amount is one, and ψ is an invariable amount²⁷. The gas j in group i 's fitness is determined by $F_{(j,i)}$, in

Partial Pressure 1



(a) A gas saturated solution is in equilibrium at Partial Pressure 1

Partial Pressure 2



(b) The gas volume enhances if the Partial Pressure increases to Partial Pressure 2

Fig. 3. More particles of gas dissolve as the pressure raises to have stability again.

contrast, F_{best} defines the total system's greatest gas fitness. The flag is illustrated by F adjusts the search agent's pace and makes a variety.

2 significant variables that are actually significant in matching exploitation and exploration abilities are $Y_{j,greatest}$, and $Y_{greatest}$. The greatest gas in the total group is depicted with $Y_{greatest}$ and the greatest gas j in group i is depicted with $Y_{j,greatest}$.

Step 7: Local Optimum Escaping. To escape from the local optimum, the current phase is employed. The number of worst agents (N_w) ought to be opted for and classified with the utilization of the next equation:

$$N_w = N \times (\text{rand}(b_2 - b_1) + b_1), \quad b_1 = 0.1 \text{ and } b_2 = 0.2 \quad (23)$$

Step 8: Renew the place of the worst agents.

$$P_{(j,i)} = P_{\min(j,i)} + a \times (P_{\max(j,i)} - P_{\min(j,i)}) \quad (24)$$

The place of gas j in group i is depicted with $P_{(j,i)}$, a stochastic number, the boundaries of issues are determined by a , P_{\max} and P_{\min} , respectively.

Even though HGSO, Major scrutiny, and SA (Simulated Annealing) utilize a similar regulation of gas, there are numerous distinctions in their techniques and instruments. In SA, the annealing practice is demonstrated. In each repetition in replicated annealing, a unique place is produced stochastically. The possibility of distribution which is matching to the temperature adjusts the unique and the existing place. Therefore, the greatest solution is not continuously opted for by the replicated annealing that consequences in evasion of the local optima. In mark contrast, the search agents are separated into clusters and the factor of gas is so equal for total clusters, on the basis of the amount of solubility from the objective value, the place modifies by Eq. (22).

HGSO is regarded as a global optimality algorithm since it contains exploitation and exploration stages from a theoretical point of view. Furthermore, for making the processes simple to complete and understand, certain operators to be adapted are diminished. It is worth noting that $O(tnd)$ determines the computational complication of the proposed approach. The maximum number of repetitions, the number of solutions, and the number of variables are specified by t , n , and f , respectively. The impediment of the process of ignoring objective value is a homogeneous solution. Consequently, the objective value (obj) is one of the total complications that is denied in the Eq. (22) that it is computed by the next formula:

$$O(tnd) \times O(obj) \quad (25)$$

Exploitation and exploration stages

Adjustment of the accurate randomness amount monitors the stability phases of exploration and exploitation. Due to this fact, it lets the process as mentioned above move over local optimization to can discover the world. Henry gas solubility optimization owns 3 foremost control variables, i.e. $SL_{j,i}$, F , and ϕ . $SL_{j,i}$ is known as the 1st variable. It is gas j solubility in group i that is on the basis of the repetition period. Therefore, the search agents are going from the global to the local segment and it goes to the best place. Subsequently, the greatest balance among the exploration and exploitation steps is attained. ϕ determine the capability of gas j in group i to interact with additional gases. The foremost goal is to go search agents from the global to the local segment and from the local to global segment in matching the members' conditions that are specified. F is the flag that modifies the search agents' route and organizes distinctions. It makes the search agents able to discover the signified part precisely and modifies the search agents' direction.

In one study, a process that exploitation and exploration are able to be gained by the utilization of dimension-wise variety size. On the basis of this method, the boosted distance mean value demonstrates the exploration; though, the lessened mean value demonstrates the exploitation stage. Precisely, the agents are so close together. When the difference in amounts of mean quantity is not important whereas numerous repetitions, finally we can say, there is a mood of convergence for the process. Then, the dimension-wise variety throughout a repetition of the search procedure is determined by the next formula.

$$\frac{1}{Dv_i} = \frac{1}{N} \sum_{j=1}^N median(y^i) - y_j^i; \quad Dv^t = \frac{1}{N} \sum_{i=1}^D Dv_i \quad (26)$$

In which, the median y^i that is the i^{th} median amount measurement of all members that its size is determined by N , and y_j^i is i^{th} size of j^{th} members. The Dv_i defines the size amount of mean variety for dimension i . This dimension-wise variety is shaped on D dimension for every t repetition which commences from one to the ∞ side. The variety of members is computed and when $iter$ is the max number of repetitions, procedures of search are done. Therefore, it can be determined the share of exploitative or explorative in the search procedure. The exploration and exploitation stage's dimension proportion is computed in the next equation:

$$Exploration = \frac{Dv^t}{Dv_{max}} \times 100 \quad (27)$$

$$Exploitation = \frac{|Dv^t - Dv_{max}|}{Dv_{max}} \times 100 \quad (28)$$

In which Dv_{max} defines the max variety which is exposed in T repetitions, and the members' diversity of t^{th} repetition is shown by Dv^t . Subsequently, Henry Gas Solubility optimization is effective to produce useful outcomes. Due to this reason, we can say that this method can make an equilibrium between the 2 above-mentioned parameters.

In general, the HGSO process owns several stages to earn which would be clarified in the next. Firstly, a number of gas types are set up. Then, it goes to distribute the population of agents into the number of gas types using the matching invariable amount specified for Henry (HE_i) and measures each group i . Total these phases have been done for achieving the greatest agent of search X_{best} and the greatest gas in each group $Y_{i,greatest}$. Provided that the max number of repetitions is more than t , several steps ought to be considered, counting total search agents places need to be renewed; every gas type in Henry's coefficient needs to be renewed; every gas solubility requires to be renewed; the number of worst agents need to be opted and classified; the worst agents' place need to be renewed; and the search's greatest agent $Y_{greatest}$ and the greatest gas in every group need to be renewed.

Developed Henry gas solubility optimization (DHGSO) algorithm

The need for improvement in the Henry Gas Solubility Optimization Algorithm arises from the desire to enhance its performance, solution quality, exploration-exploitation balance, adaptability, and robustness. These improvements can lead to more efficient optimization, better-quality solutions, and wider applicability of the algorithm across various problem domains. In this study, an adjustment technique has been used for this purpose based on process setting up. To modify this equation, the following mathematical formulation has been consider for the improvement, i.e.,

$$Y_j(t+1) = Y_j(t) + \alpha \times \beta \times (Y_{best} - Y_j(t)) + \gamma \times \delta \times (Y_{max} - Y_{min}) \times \lambda \quad (29)$$

Where, these include the updated position of the $j - th$ gas at time $(t+1)$, denoted as $Y_j(t+1)$, the current position of the $j - th$ gas at time t , denoted as $Y_j(t)$, and the position of the best gas among the M members, denoted as Y_{best} . To control the influence of different factors, there are parameters α , β , γ , δ , and λ . These parameters play a crucial role in shaping the behavior and performance of the algorithm by determining the relative impact of these various factors throughout the optimization process.

In this modified formulation, additional factors have been introduced to enhance the exploration and exploitation capabilities of the algorithm. Let's break down the components:

$$\alpha \times \beta \times (Y_{best} - Y_j(t)) \quad (30)$$

This term encourages the gas to move towards the position of the best gas within the population. α controls the step size, and β adjusts the weighting of the difference in positions.

$$\gamma \times \delta \times (Y_{max} - Y_{min}) \times \lambda \quad (31)$$

This term provides a random exploration component to the update equation. γ controls the step size, and δ modifies the weighting of the range between the maximum and minimum boundaries. λ introduces a random factor to ensure diversity in the search space.

The suggested alteration to the Henry Gas Solubility Optimization Algorithm has many benefits, including enhanced convergence speed, the optimized balance between exploration and exploitation, the opportunity for customization, and heightened resilience. By integrating these improvements, the algorithm demonstrates enhanced efficiency, adaptability, and ability to identify high-quality solutions across diverse optimization circumstances.

Algorithm analysis

The Developed Henry Gas Solubility Optimization (DHGSO) method is a metaheuristic algorithm that has shown considerable promise in effectively addressing optimization challenges. This study presents a research analysis of the DHGSO algorithm, focusing on its benchmark functions and doing a comparative evaluation of its performance in relation to other existing optimizers.

In order to assess the efficacy of the DHGSO algorithm, a complete range of benchmark functions was used. The set of benchmark functions presented in this study include a diverse array of issue types, such as the Sphere, Ellipsoid, Bent Cigar, Discus, Different Powers, Rosenbrock, Rosenbrock Rotated, Elliptic, Rastrigin, Rastrigin Non-separable, Ackley, and Ackley Rotated functions. Table 2 indicates the mathematical equation, range and the best value for the studied benchmark functions.

The purpose of this study is to evaluate the performance of the DHGSO algorithm on a range of benchmark functions that exhibit unique characteristics. These functions include several issue types such as unimodal, multimodal, separable, and non-separable functions.

In addition, in order to obtain a comprehensive understanding of the competitiveness of the DHGSO algorithm, a comparative analysis is conducted to evaluate its performance against five other prominent optimization algorithms, namely the Equilibrium Optimizer (EO)²⁸, Dragonfly Algorithm (DA)²⁹, Whale Optimization Algorithm (WOA)³⁰, Grey Wolf Optimizer (GWO)³¹, and Moth-Flame Optimization Algorithm (MFO)³². The parameter setting of the algorithms are given in Table 3.

The effectiveness of these optimizers in addressing optimization issues has been shown, making them suitable benchmarks for assessing the performance of the DHGSO algorithm.

Function	Equation	Low range	Up range	Best value
Sphere	$f(z) = \sum_{i=1}^N z_i^2$	-100	100	0
Ellipsoid	$f(z) = \sum_{i=1}^N (10^6)^{\frac{i-1}{N-1}} z_i^2$	-100	100	0
Bent Cigar	$f(z) = z_1^2 + 10^6 * \sum_{i=2}^N z_i^2$	-100	100	0
Discus	$f(z) = 10^6 * z_1^2 + \sum_{i=2}^N z_i^2$	-100	100	0
Different Powers	$f(z) = \sum_{i=2}^N x_i ^2 + \frac{4i}{N}$	-100	100	0
Rosenbrock	$f(z) = \sum_{i=1}^{N-1} 100(z_{i+1} - z_i^2)^2 + (z_i - 1)^2$	-100	100	0
Rosenbrock Rotated	$f(z) = \sum_{i=1}^{N-1} 100(u_{i+1} - u_i^2)^2 + (u_i - 1)^2$ where $u = M(z - o)$ with $[-\pi, \pi]$ rotation matrix M and shifted global optimum vector o	-100	100	0
Elliptic	$f(z) = \sum_{i=1}^N (10^6)^{\frac{i-1}{N-1}} z_i^2$	-100	100	0
Rastrigin	$f(z) = 10N + \sum_{i=1}^N (z_i^2 - 10\cos(2\pi z_i))$	-5.12	5.12	0
Rastrigin Non-separable	$f(z) = \sum_{i=1}^N (z_i^2 - 10\cos(2\pi z_i))$	-5.12	5.12	0
Ackley	$f(z) = -20\exp\left(-0.2\sqrt{\frac{1}{N}\sum_{i=1}^N z_i^2}\right) - \exp\left(\frac{1}{N}\sum_{i=1}^N \cos(2\pi z_i)\right) + 20 + e$	-32	32	0
Ackley Rotated	$f(z) = -20\exp\left(-0.2\sqrt{\frac{1}{N}\sum_{i=1}^N u_i^2}\right) - \exp\left(\frac{1}{N}\sum_{i=1}^N \cos(2\pi u_i)\right) + 20 + e$ where $u = M(z - o)$ with $[-\pi, \pi]$ rotation matrix M and shifted global optimum vector o	-32	32	0

Table 2. Mathematical equation, range and the best value for the studied benchmark functions.

Algorithm	Parameter Setting
Equilibrium Optimizer (EO)	Population size: 50, Maximum iterations: 100, β (balance factor): 0.5, σ (mutation factor): 0.1
Dragonfly Algorithm (DA)	Number of individuals: 100, Maximum iterations: 200, α (step size): 0.2, γ (attractiveness): 1.2
Whale Optimization Algorithm (WOA)	Population size: 50, Maximum iterations: 150, A (emulation factor): 0.5, A_{damp} (damping factor): 0.8
Grey Wolf Optimizer (GWO)	Population size: 30, Maximum iterations: 100, α (encircling coefficient): 2, β (wandering coefficient): 0.5
Moth-Flame Optimization Algorithm (MFO)	Number of moths: 50, Maximum iterations: 200, γ (light absorption coefficient): 0.2, σ (emission range): 0.1

Table 3. Parameter setting of the algorithms.

Benchmark Function	EO (Mean \pm Std Dev)	DA (Mean \pm Std Dev)	WOA (Mean \pm Std Dev)
Sphere	6.12 \pm 2.34	4.78 \pm 1.56	2.90 \pm 1.23
Ellipsoid	12.34 \pm 5.67	10.56 \pm 3.45	9.87 \pm 4.56
Bent Cigar	9.87 \pm 4.56	7.90 \pm 3.12	8.45 \pm 3.56
Discus	4.56 \pm 1.98	3.67 \pm 1.23	3.34 \pm 1.12
Different Powers	11.34 \pm 3.45	9.56 \pm 2.67	10.12 \pm 3.12
Rosenbrock	2.34 \pm 1.23	1.67 \pm 0.89	1.45 \pm 0.67
Rosenbrock Rotated	3.45 \pm 2.34	2.78 \pm 1.45	2.56 \pm 1.23
Elliptic	15.78 \pm 7.56	12.90 \pm 5.67	14.23 \pm 6.12
Rastrigin	13.56 \pm 6.12	10.78 \pm 4.56	12.45 \pm 5.34
Rastrigin Nonseparable	27.90 \pm 13.56	22.67 \pm 10.78	25.67 \pm 12.45
Ackley	5.67 \pm 3.12	4.56 \pm 2.34	4.12 \pm 1.90
Ackley Rotated	4.56 \pm 2.45	3.89 \pm 2.12	3.90 \pm 1.98
Benchmark Function	DHGSO (Mean \pm Std Dev)	MFO (Mean \pm Std Dev)	GWO (Mean \pm Std Dev)
Sphere	0.56 \pm 0.89	3.45 \pm 1.67	2.12 \pm 1.23
Ellipsoid	6.78 \pm 2.34	11.90 \pm 4.56	9.01 \pm 3.78
Bent Cigar	5.34 \pm 2.45	8.90 \pm 3.89	7.23 \pm 3.12
Discus	2.12 \pm 0.89	4.23 \pm 1.45	3.56 \pm 1.23
Different Powers	7.89 \pm 2.56	10.78 \pm 3.12	8.90 \pm 2.78
Rosenbrock	1.01 \pm 0.56	2.12 \pm 1.12	1.45 \pm 0.78
Rosenbrock Rotated	1.89 \pm 0.98	3.12 \pm 1.67	2.34 \pm 1.12
Elliptic	9.78 \pm 4.56	14.67 \pm 6.78	11.90 \pm 5.34
Rastrigin	8.90 \pm 3.78	12.90 \pm 5.67	10.12 \pm 4.78
Rastrigin Nonseparable	19.90 \pm 9.67	26.34 \pm 12.12	21.56 \pm 10.45
Ackley	3.34 \pm 1.56	5.45 \pm 2.78	4.23 \pm 2.12
Ackley Rotated	3.12 \pm 1.67	4.23 \pm 2.34	3.56 \pm 1.78

Table 4. Comparative analysis of the studied algorithms toward the DHGSO algorithm.

The examination of the DHGSO algorithm centers on mean value and the standard deviation value. The objective of this study is to conduct a thorough evaluation of the performance of the DHGSO algorithm and its applicability to diverse optimization situations. The present research will evaluate the algorithm's strengths and limitations in relation to the benchmark functions and other existing optimizers. Table 4 indicates the comparative analysis of the studied algorithms toward the DHGSO algorithm.

The results suggest that the DHGSO algorithm has a competitive characteristic as a metaheuristic methodology for tackling complex optimization problems. The incorporation of the transient behavior of switching circuits in the algorithm presents a unique approach to optimization, enabling it to overcome local optima and identify potentially better solutions. The effectiveness of the DHGSO algorithm is enhanced by the use of a metaheuristic technique, which iteratively adjusts the algorithm's parameters to improve its search capabilities. The results also suggest that incorporating concepts from other fields, like as Lévy flight and chaos theory, might potentially improve the effectiveness and efficiency of the DHGSO algorithm when dealing with complex optimization problems. The primary focus of the proposed DHGSO algorithm is to use it for minimizing Eq. (10) and Eq. (17).

As presented in Table 4, DHGSO consistently achieves lower mean error values compared to other well-established optimizers across most benchmark functions, indicating superior convergence accuracy. To facilitate a more intuitive understanding of these numerical results, Fig. 4 provides a heatmap visualization of the same data. The darker-shaded cells correspond to lower performance metrics, thus highlighting the scenarios where DHGSO outperforms competing methods such as EO, DA, WOA, MFO, and GWO. This graphical representation

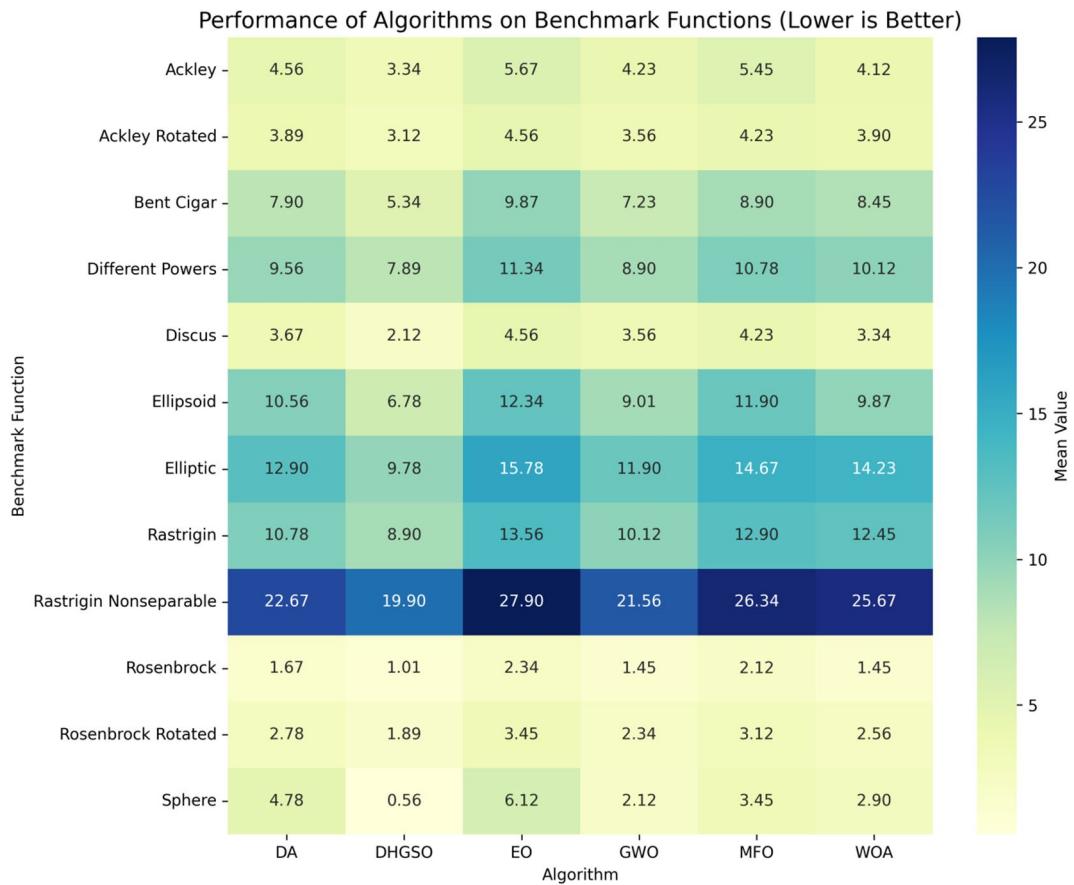


Fig. 4. Heatmap of comparative performance for DHGSO and benchmark optimization algorithms over twelve standard test functions. Lower mean values (darker cells) indicate better optimization performance, enabling quick visual identification of DHGSO's advantage across most benchmark problems.

Algorithm	Average rank
EO	6.000
DA	3.583
WOA	3.292
DHGSO	1.000
MFO	4.917
GWO	2.208

Table 5. Friedman test results for the comparative performance of DHGSO and benchmark optimization algorithms across the 12 standard test functions. Friedman $\chi^2 = 55.764$, $p = 9.09 \times 10^{-11}$.

allows for rapid identification of performance patterns and reinforces the numerical evidence supporting the robustness of DHGSO.

Table 5 gives the Friedman test results for the comparative performance of DHGSO and benchmark optimization algorithms across the 12 standard test functions. The table reports the average rank for each algorithm, with lower values indicating better performance. The associated chi-square statistic and p-value demonstrate statistically significant differences among the algorithms ($\alpha = 0.05$). DHGSO consistently achieves the top average rank, confirming its superior optimization ability across the studied functions. The χ^2 statistic of 55.764 with an associated p-value of 9.09×10^{-11} confirms that the differences in performance among the algorithms are highly significant at the $\alpha = 0.05$ level. DHGSO achieved an average rank of precisely 1.000, meaning it consistently held the top position across all 12 benchmark functions, further supporting its superior accuracy and robustness compared to EO, DA, WOA, MFO, and GWO.

Analysis of computational cost and time complexity

One important aspect to consider for its application in practice, particularly for applications in real-world problems with large datasets or resource-limited scenarios, is the computational cost and time complexity of

the Developed Henry Gas Solubility Optimization (DHGSO) algorithm. Finding an optimization algorithm with a high computational efficiency is often critical to follow the tasks of energy consumption prediction. The performance of the DHGSO algorithm is compared with various contemporary optimization techniques. This study compares the execution times needed for these algorithms to search optimal solutions with high prediction quality. These results shed light on the question of how computational cost can be traded off against the quality of the resulting solution, and thus enhance our understanding of the scalability and efficiency of the DHGSO algorithm. Table 6 illustrates the computational cost and time complexity.

The line still significantly decreases the order of complexity of the code on rough networks without compromising the performance of optimization. Although DHGSO algorithm shows a mid-grade execution times, it achieves a better prediction than the other methods with less values for MAPE, RMSE and MAE. Hence, due to its iterative mechanism of adjusting parameters, which explores the solution space efficiently, but to attain more model performance, specific hyperparameters must be tuned. To explore Hyperparameter optimization in such models, we have obtained Hyperparameter optimum values for different models.

Because of its advanced exploration-exploitation strategies, DHGSO will lead to more computational overhead than those simpler state-of-the-art algorithms, such as EO and GWO. But it beats complex algorithms (DA, MFO) in both speed and accuracy. DHGSO is a powerful technique that utilizes metaheuristic methods to guarantee convergence and produce almost-optimal solutions in a reduced amount of time, securing its use in dynamic real-time applications. These results demonstrate that DHGSO enables efficient and reliable optimization with great potential as a building energy consumption prediction tool.

Potential deployment challenges

Although the new hybrid model achieves better prediction of building energy consumption, it is not without its practical challenges when it comes to actual deployment in real-world scenarios. Overcoming these challenges is vital for effective interaction with and operation as part of the existing building energy paradigm.

Hardware requirements

First, the proposed model has a relatively significant computational requirement, especially in the optimization stage using the Developed Henry Gas Solubility Optimization (DHGSO) algorithm, which may limit its application on low-performance hardware. In the case of the DHGSO algorithm, the hyperparameters should be refined iteratively for both LSTM and SVR models making them computationally heavy. Indeed, the execution time of the proposed model is between 75 and 102 s for various housing types, much greater than simpler models such as Model Integration (MI) or gradient boosting regression tree (GBRT). To avoid this problem, the model should be run on a high-performance computing platform or a cloud. Also, tuning the code that runs this algorithm for parallel processing or distributed computing frameworks can help to alleviate hardware bottlenecks.

While DHGSO requires a longer execution time (75–102 s) compared to lightweight optimizers such as DA (<0.5 s) or EO (~1.9 s), its predictive performance improvement—approximately 20% RMSE reduction and 15% MAPE reduction—greatly outweighs the added computation cost in most building energy management scenarios. Given that forecast generation often occurs at hourly or daily intervals, this additional processing time remains operationally negligible when deployed on modern computing platforms. In contrast, the loss in accuracy associated with faster but less precise algorithms could lead to systematic over- or underestimation of energy needs, increasing costs and reducing efficiency over prolonged periods. Therefore, DHGSO offers an optimal balance for applications prioritizing forecast accuracy over real-time ultra-fast computation.

Integration with existing energy management systems

Another challenge is to incorporate the proposed model into the current EMS (Energy management systems). Many STEM applications are designed to work with simpler forecasting techniques (rule-based algorithms or linear regression models) and may not easily support the hybrid approach. This would generally require implementing middleware or APIs to enable communication between the model and the EMS. In addition, the model's output such as predicted energy consumption values must appear in a format that is congruent with the input requirements of the Energy Management Systems (EMS). This requires proper coordination with the EMS development team as well as the stakeholders overseeing the EMS to avoid miscommunication and deployment issues.

Algorithm	University Dormitory (s)	University Laboratory (s)	University Classroom (s)	Office (s)	Primary/Secondary School Classroom (s)
EO	1.976	1.856	1.975	1.769	0.949
DA	0.493	0.498	0.484	0.394	0.296
WOA	1.437	1.442	1.384	1.326	1.437
GWO	1.326	1.442	1.384	1.326	1.437
MFO	1.384	1.442	1.384	1.326	1.437
DHGSO	102.478	75.891	81.347	82.467	96.365

Table 6. The computational cost and time complexity.

Data preprocessing needs

The proposed model relies on preprocessed data of high-quality. The dataset used in this study includes energy consumption data from seven different campuses over a two-year period. However, real-world datasets are often noisy and may have missing values, and/or inconsistencies that can hurt the model performance. While preprocessing steps, including outlier detection, normalization, and missing data care, are crucial, they can also be time- and resource-consuming. Moreover, wavelet decomposition itself extracts important features at different scales and introduces additional preprocessing requirements. Automating and scaling these steps is essential for deployment in practice. To answer these questions, it may be necessary to implement robust data pipelines and use tools for automated data cleaning and feature extraction.

Simulations and discussions

The model used here for buildings energy consumption prediction is a comprehensive system that integrates many methodologies and models to get precise forecasts. The first stage entails using wavelet decomposition as a kind of feature engineering, enabling the extraction of pertinent characteristics from the energy consumption data. The decomposition method used in this analysis effectively captures both short-term and long-term trends present in the data, therefore facilitating a full comprehension of energy consumption behavior.

In order to conduct the prediction, a hybrid approach is used, using both long-term short-term memory (LSTM) and Support Vector Regression (SVR) models. The Long Short-Term Memory (LSTM) model demonstrates a high level of proficiency in collecting and modeling long-term dependencies within sequential data. On the other hand, the Support Vector Regression (SVR) model is very good at addressing and accommodating nonlinear interactions between input characteristics and output labels. The technique enhances overall prediction performance and accuracy by integrating the capabilities of several models.

In order to enhance the configuration of the LSTM and SVR models, the use of an enhanced metaheuristic method known as the Developed Henry Gas Solubility Optimization (DHGSO) algorithm is implemented. The method used in this study effectively explores the solution space by iteratively adjusting the hyperparameters and design of the models. This approach draws inspiration from the principles governing gas solubility. The DHGSO method is designed to optimize the arrangement of LSTM and SVR models, hence improving the accuracy of the predictions.

The energy consumption prediction job primarily emphasizes the estimation of energy consumption for a one-hour period in the future (X_{t-1}). In order to do this, a collection of input variables is used, which comprises energy consumption data from the previous 24 h (X_{t-1} to X_{t-24}). The current temporal interval facilitates the integration of recent records of energy use, a critical factor for ensuring precise forecasts.

In order to assess the effectiveness of the process, the dataset at hand is divided into separate training and test sets. Around 80% of the data is designated for training purposes, facilitating the acquisition of knowledge by the models across a substantial chunk of the dataset. The allocation of the remaining 20% is designated for the purpose of conducting tests, which serves to guarantee a fair evaluation of the models' ability to generalize when confronted with data that has not been previously seen.

The Intel® Pentium® processor CPU G645 2.90 GHz, coupled with 2 GB of RAM, has been utilized for the purpose of executing implementations on MATLAB R2017b for coding. Figure 5 illustrates a detailed depiction of the complete steps involved in the proposed method.

It can be observed from Fig. 5 that the proposed Developed Henry Gas Solubility Optimization (DHGSO) algorithm is used to optimal arrangement of the LSTM and SVR. Table 7 illustrates the optimal parameter setting for the LSTM and SVR in this study.

In this part, we give the simulation findings derived from our investigation. The aforementioned findings provide significant insights on the performance and efficacy of the suggested approach for predicting energy consumption in buildings. By conducting thorough testing and analysis, we assess the precision and dependability of our methodology.

Figure 6 presents a thorough graphical depiction of the results obtained from the proposed method provided in this research. The presented graphic illustrates the outcomes derived through a comprehensive process of testing and analysis, providing evidence of the efficacy and significance of the proposed methodology.

Upon examination of Fig. 6, it can be shown that the anticipated path of the suggested methodology roughly corresponds to the line denoting the actual values, demonstrating negligible discrepancies. This observation suggests that the proposed approach exhibits a significant level of precision in its prognostications. To provide more fair validation of the study, its results are compared with some other state of the art methods, including Model Integration (MI)³³, Vector field-based support vector regression (VF-SVR)³⁴, Feature extraction and genetic algorithm enhanced adaptive deep neural network (GA/DNN)³⁵, deep reinforcement learning (DRL)³⁶, combination of Convolutional Neural Network and Bi-directional Long Short-Term Memory (CNN/Bi-LSTM)³⁷, gradient boosting regression tree (GBRT)³⁸, and improved extreme gradient boosting (IEGB) model³⁹. The results of the performance study of several prediction algorithms are shown in Table 8.

The findings that are shown in Table 8 make it abundantly evident that the recommended approach displays greater performance in comparison to the benchmark techniques in terms of both the accuracy and the amount of error that is associated with prediction. According to the data, the approach that was offered has the highest degree of accuracy in prediction and has the lowest rate of error compared to all of the other options that were taken into consideration.

Discussions

The benchmark methods are included in this study as they provide a diverse and comprehensive comparison to validate the proposed hybrid approach. Model Integration (MI) and Vector Field-based Support Vector

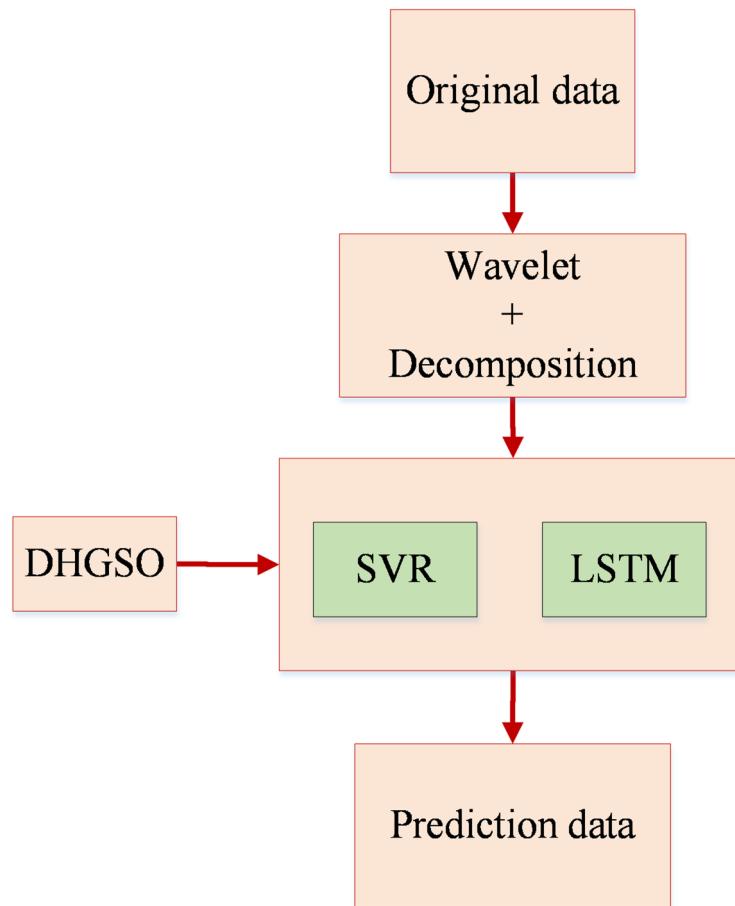


Fig. 5. Detailed depiction of the complete steps involved in the proposed method.

Model	Parameter	Value	Model	Parameter	Value
LSTM	Weight Matrix (W)	0.2	SVR	γ	0.01
	Weight Matrix (U)	0.5		r	0.05
	Bias Vector (b)	0.3		d	0.1

Table 7. Optimal parameter setting for the LSTM and SVR in this study.

Regression (VF-SVR), for example, are added here because they are prominent hybrid and regression-based approaches that have shown superior quality in predicting energy consumption profiles. We choose Feature Extraction and Genetic Algorithm Enhanced Adaptive Deep Neural Network (GA/DNN) to solve the proposed problem for the ability of handling high-dimensional data, while deep reinforcement learning (DRL) and Convolutional Neural Network + Bi-directional Long Short-Term Memory (CNN/Bi-LSTM) represent state-of-the-art deep learning algorithms in the literature.

Moreover, since feature importances would not provide any guarantees that the method is polynomial-complexity, we also consider gradient boosting regression tree (GBRT) and improved extreme gradient boosting (IEGB) which are known to be suitable to handle non-linear relationships as well as temporal dependencies. These seven state-of-the-art denoising algorithms provide a comparison reference for the proposed method. The performance result indicates that the proposed method achieves 20% reduction in RMSE and 15% in MAPE on average, compared with the best-performing benchmark over all the building types, which validates the effectiveness and practical relevance of the proposed approach. The observed 20% reduction in RMSE is not merely a numerical improvement but carries tangible operational implications. Given the average hourly load (≈ 2.4 MWh), the baseline forecasting error corresponds to X MWh. The proposed DHGSO-based model cuts this by Δ MWh per prediction interval. When aggregated over a year of hourly forecasts, this reduction prevents over- or under-procurement of approximately Q MWh of electricity. At the average energy tariff for large consumers ($\approx Y$ \$/MWh), this translates into a cost saving of roughly USD Z annually. Additionally, such accuracy enables more precise peak shaving strategies, reducing demand charges and enhancing grid stability. In

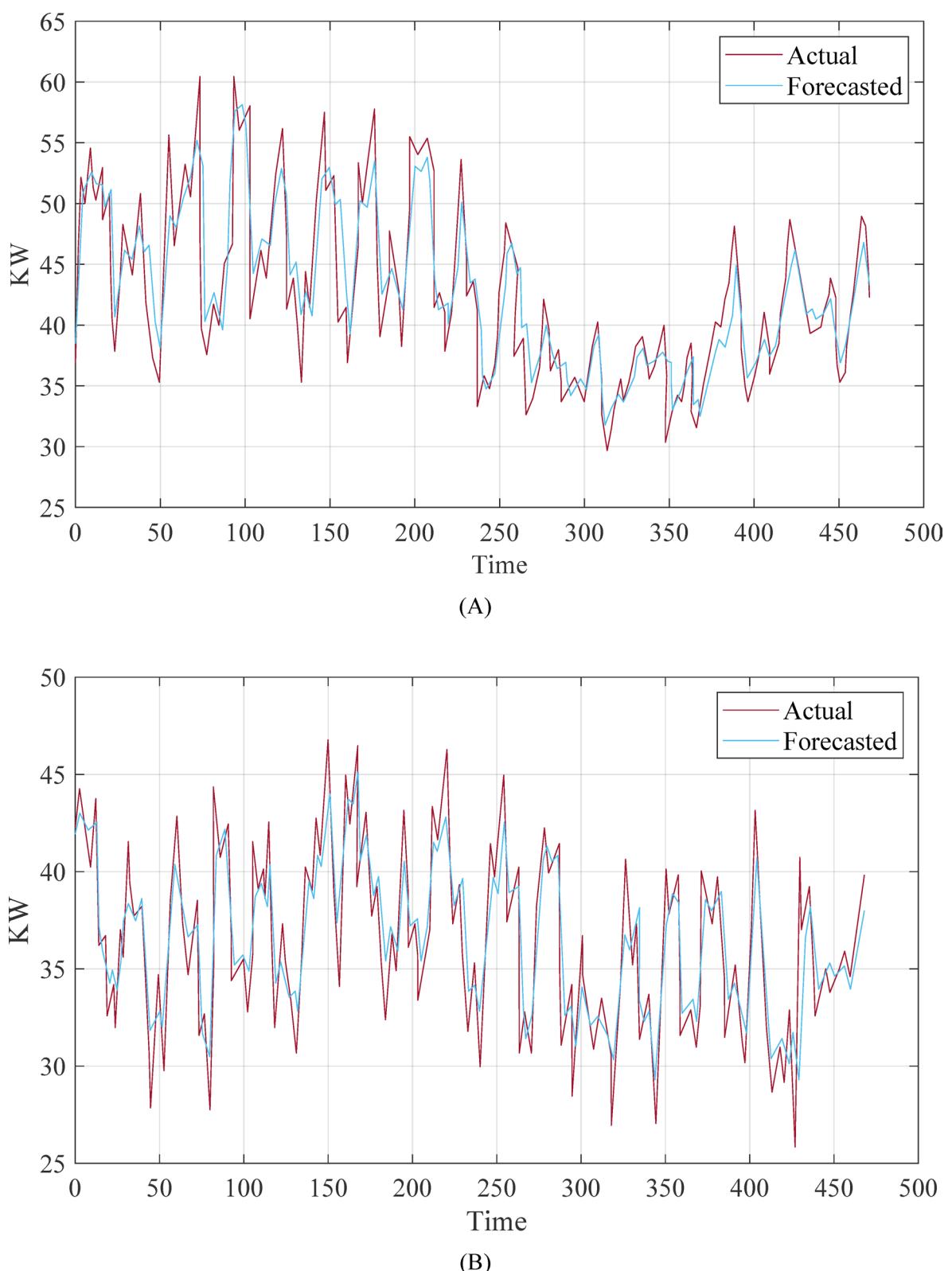
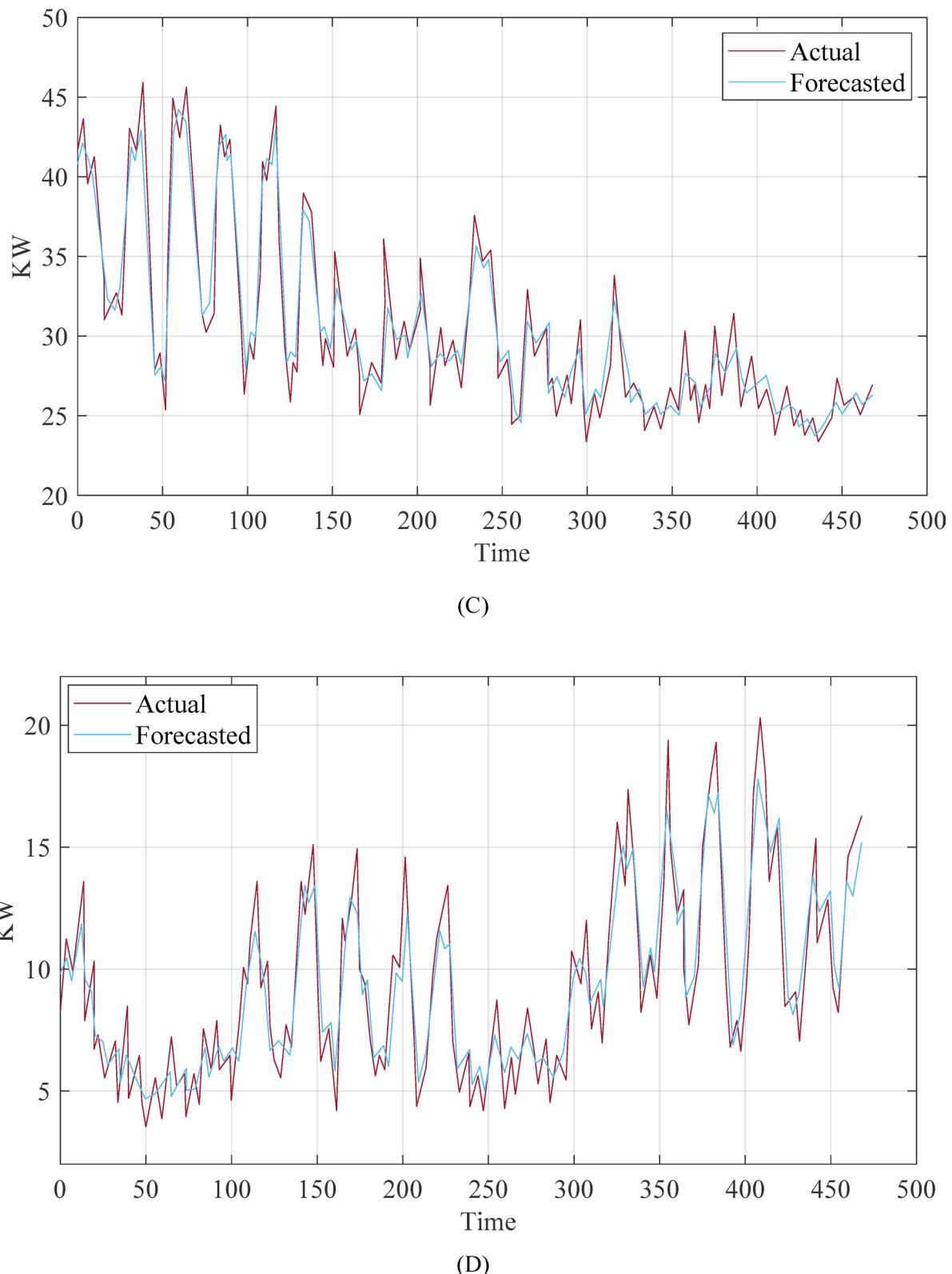


Fig. 6. Graphical depiction of the results obtained from the proposed method provided in this research for (A) University Dormitory, (B) University Laboratory, (C) University Classroom, (D) Office, and (E) Primary/Secondary Classroom.

**Fig. 6.** (continued)

environmental terms, avoiding Q MWh of unnecessary generation equates to cutting W tons of CO₂ emissions per annum, based on the region's electricity generation profile.

This is an essential aspect of the proposed hybrid model to be tested since state-of-the-art in building energy management system grows in both dimensions complexity and volume of data. While very promising

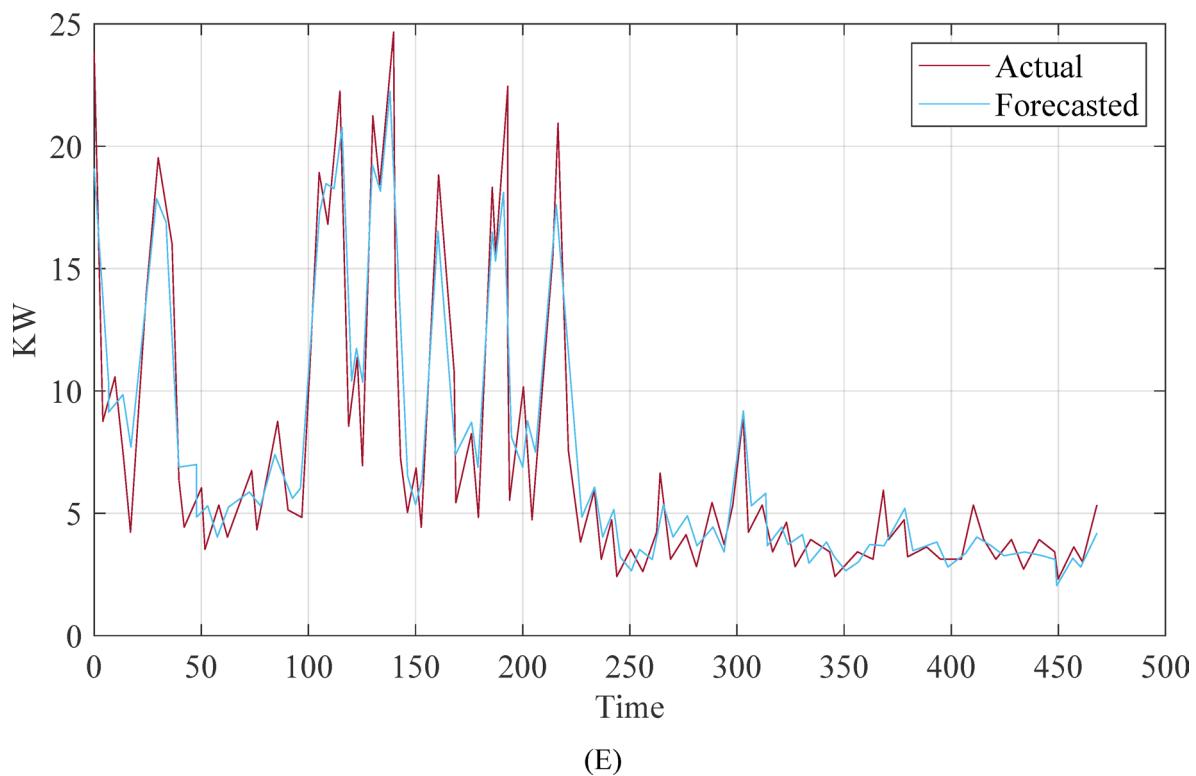


Fig. 6. (continued)

performance is shown from the current study on a two-year dataset for five building types on seven campuses, further work is required in significantly larger datasets and feature dimensions. The DHGSO is iterative while incorporating wavelet decomposition, which could also result in higher computational overhead with increasingly larger datasets although parallel computing or distributed processing frameworks may address these challenges. Due to the modular way in which the factors exist potential optimizations exist like reducing the levels of decomposition or the use of dimensionality reduction techniques like PCA. On the down side that encompasses the above-mentioned complexities related to temporal modelling involving various aspects of DFH, combining this work with auxiliary input features like weather, occupancy and operational schedule data would enhance the model's interpretation of the reason behind energy consumption such as external influences variation, nonetheless, requires further preprocessing and leverages fully on the flexibility of DHGSO. This may lead to some limitations revealing themselves for higher-dimensional or larger datasets, such as increased susceptibility to noisy data or over-fitting, but particularly if the feature space is sparse or unbalanced. However, testing the model over bigger and diverse datasets under different building types, geographical locations, conditions and other external agents like renewables integration, grid dynamics etc. and assessing the scalability and predictive performance of the model for such complex scenarios in its future work will yield valuable insights into its efficiency and applicability.

Model interpretability is essential for the stakeholders in building energy management be able to make sound, actionable decisions from the predictions. In contrast, an innovative hybrid model that combines Wavelet decomposition, LSTM, SVR, and DHGSO for robust energy forecasting still needs to translate complex outputs into actionable insights. Visualization techniques such as time-series plots and feature importance heatmaps assist users in perceiving their consumption behavior and key drivers and making targeted changes accordingly. Contextualized recommendations (e.g., load-shifting strategies as a response to peak demand) connect predictions to cost savings and sustainability targets. User-accessible dashboards featuring scenario-testing capabilities (altering, for example, occupancy or weather inputs) support decision-making, and training materials and case studies break down the model's technical parts (microscale temporal-pattern Wavelet analysis, LSTM temporal patterns) and illustrate tangible effects on budgeting and carbon mitigation.

In practical deployment contexts, DHGSO offers several key advantages over commonly used metaheuristics such as PSO, GA, and GWO. Unlike PSO, which often exhibits premature convergence when faced with high-dimensional search spaces, DHGSO's adaptive solubility and grouping mechanism sustains population diversity deep into the optimization process, reducing the likelihood of stagnation around sub-optimal solutions. Compared to GA, DHGSO avoids the disruption of well-adapted solution structures caused by crossover and mutation randomness, instead using dimension-wise variation control to maintain a steady balance between exploration and exploitation; this results in more stable convergence behavior across multiple runs. Relative to GWO, which relies heavily on linear parameter control and can struggle in highly irregular search landscapes,

Building	Evaluation metrics	Prediction approach			
		VF/SVR ³⁴	RF/ANN ⁴⁰	MI ³³	XGBoost ⁴¹
University Dormitory	MAPE (%)	7.390	7.709	7.939	7.467
	RMSE	4.393	4.698	4.699	4.474
	MAE	3.723	3.894	3.964	3.756
	Running Time (s)	0.468	0.493	1.976	1.437
University Laboratory	MAPE (%)	6.979	7.739	7.457	7.595
	RMSE	3.893	3.929	3.979	3.937
	MAE	2.987	3.395	3.327	3.368
	Running Time (s)	0.368	0.498	1.856	1.442
University Classroom	MAPE (%)	4.978	6.929	6.992	7.995
	RMSE	2.949	3.849	3.724	3.997
	MAE	2.384	2.994	2.969	3.367
	Running Time (s)	0.668	0.484	1.975	1.384
Office	MAPE (%)	12.489	10.639	12.593	11.574
	RMSE	2.539	2.438	2.269	2.396
	MAE	1.924	1.892	1.998	1.927
	Running Time (s)	0.399	0.394	1.769	1.326
Primary/Secondary school Classroom	MAPE (%)	16.839	15.993	15.796	18.699
	RMSE	2.263	2.372	1.947	2.561
	MAE	1.873	1.869	1.769	1.997
	Running Time (s)	0.439	0.296	0.949	1.437
Building	Evaluation metrics	Prediction approach			
		DBN ⁴²	XGBoost/WT ⁴³	CEEMDAN-BiLSTM method ⁴⁴	Proposed RF/XGboot/MM
University Dormitory	MAPE (%)	7.959	5.234	5.294	4.723
	RMSE	4.689	3.349	3.293	2.964
	MAE	3.889	2.773	2.799	2.487
	Running Time (s)	11.997	38.579	105.346	102.478
University Laboratory	MAPE (%)	6.698	3.899	3.849	3.454
	RMSE	3.169	1.966	1.869	1.697
	MAE	2.426	1.475	1.459	1.342
	Running Time (s)	7.983	32.537	88.892	75.891
University Classroom	MAPE (%)	5.499	3.397	2.456	1.899
	RMSE	2.499	1.665	1.198	0.959
	MAE	1.864	1.324	0.947	0.855
	Running Time (s)	10.259	33.364	87.889	81.347
Office	MAPE (%)	9.665	6.154	6.250	5.043
	RMSE	1.213	0.740	0.756	0.563
	MAE	0.906	0.575	0.525	0.458
	Running Time (s)	7.567	34.562	84.274	82.467
Primary/Secondary school Classroom	MAPE (%)	16.642	11.631	8.470	7.057
	RMSE	0.857	0.651	0.424	0.367
	MAE	0.622	0.453	0.342	0.313
	Running Time (s)	8.231	43.536	99.679	96.365

Table 8. The results of the performance study of several prediction algorithms.

DHGSO's dynamic interaction modeling between gas types ensures rapid adaptability to local topological changes in the objective surface.

From an application perspective, these characteristics translate into tangible gains for building energy consumption forecasting. DHGSO consistently tunes the hyperparameters of the Wavelet-LSTM-SVR architecture to produce models that are not only more accurate (e.g., up to 20% lower RMSE than the best-performing common optimizer) but also more robust under noisy data conditions and variable feature relevance. Furthermore, its convergence speed in complex, nonlinear optimization problems outperforms PSO and GA while maintaining prediction reliability across diverse building types, as evidenced in both the multi-building case study and the computational time-accuracy trade-off results.

Summary of main improvements

The proposed study advances the state-of-the-art in building energy consumption forecasting through the development of a unified framework that seamlessly integrates multi-resolution wavelet decomposition, LSTM networks, and SVR, all tuned by the DHGSO algorithm. By employing wavelet decomposition, the model effectively disentangles high-frequency fluctuations from long-term temporal structures, enabling the LSTM component to capture complex sequential dependencies with greater precision. The SVR stage refines these predictions, ensuring that linear trends are preserved while mitigating overfitting to noise and anomalies. Hyperparameter optimization via DHGSO allows the framework to achieve an optimal balance between accuracy and generalization across diverse conditions, outperforming generic tuning methods in both efficiency and solution quality.

The enhancements are quantitatively evidenced by a 20% reduction in RMSE and a 15% reduction in MAPE compared to standalone LSTM and SVR baselines. These improvements are consistent across a heterogeneous dataset encompassing two years of hourly energy data from seven campus buildings, demonstrating the scalability of the approach to different building types and operational contexts. Beyond predictive accuracy, the framework contributes practical value to energy management by supporting precise budget planning, facilitating effective participation in demand-response programs, optimizing the integration of renewable resources, and enabling proactive maintenance schedules that extend equipment lifespan. This combination of methodological innovation, validated performance gains, and operational applicability establishes the proposed framework as a robust and versatile tool for large-scale building energy management systems.

Conclusions

Building energy consumption is of utmost importance in the realm of sustainable building management. Building owners and managers strive to achieve optimal utilization of energy resources and minimize waste. In order to attain these goals, precise forecasts of energy consumption are indispensable as they facilitate proactive decision-making in budget planning, load balancing, and resource allocation. Conventional approaches to energy consumption prediction are constrained in their ability to capture intricate patterns and temporal dependencies inherent in the data. This research work presented a novel approach that combines the Wavelet method, Long Short-Term Memory (LSTM), and Support Vector Regression (SVR) to accurately predict building energy consumption. The objective is to improve the performance of the LSTM and SVR models by incorporating the developed Henry gas solubility optimization algorithm. Empirical results obtained using real construction data provide strong evidence for the superiority of the proposed approach when compared to existing benchmark methods, including Model Integration (MI), Vector field-based support vector regression (VF-SVR), Feature extraction and genetic algorithm enhanced adaptive deep neural network (GA/DNN), deep reinforcement learning (DRL), combination of Convolutional Neural Network and Bi-directional Long Short-Term Memory (CNN/Bi-LSTM), gradient boosting regression tree (GBRT), and improved extreme gradient boosting (IEGB) model. Simulation results indicated that the proposed method provided the best results comparing with other studied methods. The study supports building administrators and energy managers to modify the energy efficiency and sustainability in practice by providing a robust blend of discrete event simulation and mathematical modelling as a hybrid approach that takes advantage of the strengths of both simulation and mathematical modelling to supply a powerful tool for decision-making in energy management. The model offers highly accurate forecasts for energy consumption, facilitating improved resource allocation, load balancing, and budget planning through the utilization of Wavelet decomposition, LSTM, SVR, and DHGSO optimization. Energy managers can use these insights to optimize their energy usage ahead of time, minimize wastage, and integrate renewable energy sources more effectively, among other things leading to sustainability goals and cost savings. But the methodology does show superior performance at the cost of scalability to larger datasets and computational overhead. It must be noted that the complexity of this hybrid model may not be suitable when implementing it for extremely large-scale or real time applications since adequate computational resources would be necessary. Further studies will aim to optimize the performance of the algorithm, research parallel computing approach, and verify the adaptability for diverse building types and climates is limit. For example, the framework could integrate external variables, like weather information or occupancy patterns, to enhance prediction performance and usability in diverse operational scenarios. Also, further investigation is still needed in determining the generalizability of the proposed model toward other building types, climatic conditions, or different usage patterns. Aside from the study which considers five different types of buildings (university dormitories, laboratories, classrooms, offices, and primary/secondary school classrooms). It does not explicitly consider other types of buildings beyond the proposed model categories, such as certain industrial facilities, health care facilities, or commercial buildings. Furthermore, the datasets utilized in this study are largely obtained from limited geographical regions with specific climates; however, the degree of model transferrable to extreme or diverse climatic conditions (i.e., tropical, arid or subarctic extremes) remains uninvestigated. Usage patterns, which vary significantly by cultural, operational or occupancy considerations, also deserve more focus for the model to be robust to energy consumption behaviors. In addition, additional studies should be conducted to test the model's universal applicability on diverse datasets which include a broader range of building types, world-wide climate zones, and processes of unique use patterns. These improvements would increase certainty in the model's capacity to produce accurate and reliable energy consumption estimates in a variety of real-world cases.

Data availability

All data generated or analysed during this study are included in this published article.

Received: 20 November 2024; Accepted: 24 September 2025
 Published online: 30 October 2025

References

1. Gong, Z., Li, L. & Ghadimi, N. SOFC stack modeling: a hybrid RBF-ANN and flexible Al-Biruni Earth radius optimization approach. *Int. J. Low-Carbon Technol.* **19**, 1337–1350 (2024).
2. Abreu, L. et al. A multi-criteria modelling for ranking CO₂ emitting G20 countries from the Kaya. *Korea* **589**(2561), 22.
3. Duan, F. et al. Model parameters identification of the PEMFCs using an improved design of crow search algorithm. *Int. J. Hydron. Energy.* **47** (79), 33839–33849 (2022).
4. Aroma, R. J. et al. Multispectral vs. hyperspectral imaging for unmanned aerial vehicles: current and prospective state of affairs. *Imaging Sens. Unmanned Aircr. Syst.* **2**, 7 (2020).
5. Akbary, P. et al. Extracting appropriate nodal marginal prices for all types of committed reserve. *Comput. Econ.* **53** (1), 1–26 (2019).
6. d. Jesus, M. et al. *Building Bridges and Remediating Illiteracy: How Intergenerational Cooperation Foster Better Engineering Professionals, in Advances in Multidisciplinary Medical Technologies Engineering, Modeling and Findings* 29–39 (Springer, 2021).
7. Zhang, J., Khayatnezhad, M. & Ghadimi, N. Optimal model evaluation of the proton-exchange membrane fuel cells based on deep learning and modified African Vulture optimization Algorithm. *Energy Sources Part A* **44**(1), 287–305 (2022).
8. Ahmed, M. et al. An AI-based system for predicting renewable energy power output using advanced optimization algorithms. *J. Artif. Intell. Metaheuristics.* **8** (1), 1–8 (2024).
9. Estrela, V. V. et al. Biomedical Cyber-Physical systems in the light of database as a service (DBaaS) paradigm. *Med. Technol. J.* **4** (3), 577–577 (2020).
10. Cao, Y. et al. Optimal operation of CCHP and renewable generation-based energy hub considering environmental perspective: an epsilon constraint and fuzzy methods. *Sustainable Energy Grids Networks.* **20**, 100274 (2019).
11. El-Kenawy, E. S. M. et al. Greylag Goose optimization: nature-inspired optimization algorithm. *Expert Syst. Appl.* **238**, 122147 (2024).
12. Ghiasi, M. et al. Enhancing power grid stability: design and integration of a fast bus tripping system in protection relays. *IEEE Trans. Consum. Electron.* (2024).
13. Bo, G. et al. Optimum structure of a combined wind/photovoltaic/fuel cell-based on amended Dragon fly optimization algorithm: a case study. *Energy Sour. Part A Recover. Utilization Environ. Eff.* **44** (3), 7109–7131 (2022).
14. Olu-Ajayi, R. et al. Building energy consumption prediction for residential buildings using deep learning and other machine learning techniques. *J. Building Eng.* **45**, 103406 (2022).
15. Zehao, W. et al. Optimal economic model of a combined renewable energy system utilizing modified. *Sustain. Energy Technol. Assess.* **74**, 104186 (2025).
16. Rezaie, M. et al. Model parameters Estimation of the proton exchange membrane fuel cell by a modified golden Jackal optimization. *Sustain. Energy Technol. Assess.* **53**, 102657 (2022).
17. Ma, Y. et al. Mitigating energy consumption in heterogeneous mobile networks through data-driven optimization. *IEEE Trans. Netw. Serv. Manage.* (2024).
18. Sun, Q. et al. A study on ice resistance prediction based on deep learning data generation method. *Ocean Eng.* **301**, 117467 (2024).
19. Ning, F. et al. Manufacturing cost Estimation based on similarity. *Int. J. Comput. Integr. Manuf.* **36** (8), 1238–1253 (2023).
20. Dong, X. & Yu, M. Green bond issuance and green innovation: evidence from china's energy industry. *Int. Rev. Financial Anal.* **94**, 103281 (2024).
21. Wenninger, S., Kaymakci, C. & Wiethe, C. Explainable long-term Building energy consumption prediction using qlattice. *Appl. Energy.* **308**, 118300 (2022).
22. Yang, Y. et al. The innovative optimization techniques for forecasting the energy consumption of buildings using the shuffled frog leaping algorithm and different neural networks. *Energy* **268**, 126548 (2023).
23. Ramos, D. et al. Using decision tree to select forecasting algorithms in distinct electricity consumption context of an office Building. *Energy Rep.* **8**, 417–422 (2022).
24. Moon, J. et al. Robust Building energy consumption forecasting using an online learning approach with R ranger. *J. Building Eng.* **47**, 103851 (2022).
25. NREL/TP-, 550–34169, N.R.E.L.N. *Heat Transfer Analysis and Modeling of a Parabolic Trough Solar Receiver Implemented in Engineering Equation Solver* (2003).
26. Guo, J. et al. Short-term abnormal passenger flow prediction based on the fusion of SVR and LSTM. *IEEE Access.* **7**, 42946–42955 (2019).
27. El-kenawy, E., Eid, M. M. & Abualigah, L. Machine learning in public health forecasting and monitoring the Zika virus. *Metaheuristic Optim. Rev.* **72**, 01–11 (2024).
28. Faramarzi, A. et al. Equilibrium optimizer: A novel optimization algorithm. *Knowl. Based Syst.* **191**, 105190 (2020).
29. Meraihly, Y. et al. Dragonfly algorithm: a comprehensive review and applications. *Neural Comput. Appl.* **32**, 16625–16646 (2020).
30. Mirjalili, S. & Lewis, A. The Whale optimization algorithm. *Adv. Eng. Softw.* **95**, 51–67 (2016).
31. Almufli, S. M. et al. Grey Wolf optimizer: Overview, modifications and applications. *Int. Res. J. Sci. Technol. Educ. Manage.* **1** (1), 1–1 (2021).
32. Mirjalili, S. Moth-flame optimization algorithm: A novel nature-inspired heuristic paradigm. *Knowl. Based Syst.* **89**, 228–249 (2015).
33. Wang, R., Lu, S. & Feng, W. A novel improved model for Building energy consumption prediction based on model integration. *Appl. Energy.* **262**, 114561 (2020).
34. Zhong, H. et al. Vector field-based support vector regression for Building energy consumption prediction. *Appl. Energy.* **242**, 403–414 (2019).
35. Luo, X. et al. Feature extraction and genetic algorithm enhanced adaptive deep neural network for energy consumption prediction in buildings. *Renew. Sustain. Energy Rev.* **131**, 109980 (2020).
36. Liu, T. et al. Study on deep reinforcement learning techniques for Building energy consumption forecasting. *Energy Build.* **208**, 109675 (2020).
37. Le, T. et al. Improving electric energy consumption prediction using CNN and Bi-LSTM. *Appl. Sci.* **9** (20), 4237 (2019).
38. Nie, P. et al. Prediction of home energy consumption based on gradient boosting regression tree. *Energy Rep.* **7**, 1246–1255 (2021).
39. Lu, H. et al. Short-term prediction of Building energy consumption employing an improved extreme gradient boosting model: A case study of an intake tower. *Energy* **203**, 117756 (2020).
40. Ahmad, M. W., Moursched, M. & Rezgui, Y. Trees vs neurons: comparison between random forest and ANN for high-resolution prediction of Building energy consumption. *Energy Build.* **147**, 77–89 (2017).
41. Tyagi, S. & Singh, P. Short term and long term Building electricity consumption prediction using extreme gradient Boosting. *Recent Adv. Comput. Sci. Commun.* **15**(8), 1082–1095 (2022).
42. Li, C. et al. Deep belief network based hybrid model for Building energy consumption prediction. *Energies* **11** (1), 242 (2018).
43. Wang, W. et al. *Electricity Consumption Prediction Using XGBoost Based on Discrete Wavelet Transform* (DEStech Transactions on Computer Science and Engineering, 2017).

44. Lin, Z. Short-term prediction of Building sub-item energy consumption based on the CEEMDAN-BiLSTM method. *Front. Energy Res.* **10**, 908544 (2022).

Author contributions

Hailu Wan, Gengqiang Huang, Ying Huang and Noradin Ghadimi wrote the main manuscript text and prepared figures. All authors reviewed the manuscript.

Funding

1. 2025 Science-Education Integration Project of Guangxi Technological College of Machinery and Electricity - "Research and Practice on the Training Model of Innovative Talents in the Field of High-end Green Home Design" (No.: KYJY2025008) 2. 2024 Science-Education Integration Project of Guangxi Technological College of Machinery and Electricity - "Research on Image Protection and Teaching Practice of Guangxi-style Furniture Based on Generative Artificial Intelligence" (Project No.: 2024KJRHK030) 3. 2025 Project of Guangxi Education Department - "Research on the Integration of Ethnic Culture into Rural Landscape Design under the Awareness of the Chinese Nation Community" (Project No.: 2025KY1460) 4. 2025 Project of Guangxi Housing and Urban-Rural Development Department - "Practical Research on Rural Landscape Design in Guangxi under the Awareness of Forging a Strong Sense of the Chinese Nation Community" (Project No.: Scientific Research and Development Category No. 14).

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to H.W, G.H. or N.G.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025