Building Energy Management: Integrating K-means with Sparrow Search Algorithm Optimized BiLSTM for Anomaly Detection and Prediction

Xuwei Liu[1], Lisha Si[2], Wenting Tang[3], Yongde Li[3*]
[1]Campus Construction Office, China Jiliang University, Hangzhou, 310018, China
[2]Audit Department, Zhejiang Sci-Tech University, Hangzhou, 310018, China
[3]Audit Department, Zhejiang Gongshang University, Hangzhou, 310018, China
*Corresponding author's email: YongsongLiu0102@outlook.com

Abstract: Currently, building energy conservation and emission reduction are crucial for environmental protection and sustainable development. In response to the difficulty in accurately quantifying building energy consumption and the low efficiency of existing management, this study develops a building energy management system that integrates hybrid anomaly detection and intelligent prediction algorithms. The outcomes indicate that the highest determination coefficient of the prediction model reaches 0.96, and the mean absolute error of energy consumption prediction in spring is 3.33%, with a mean absolute percentage error of 1.01%. Compared with benchmark models such as sparrow search algorithm and particle swarm optimization algorithm, this method reduces the number of convergence iterations by 40.8%, further reduces the average absolute error by 14.2%, saves training time by 44.7%, and reduces memory usage by 18.5%. This method has significant advantages in building energy consumption prediction and anomaly detection, which can effectively improve the efficiency of building energy management and provide data support for energy efficiency enhancement and pollutant reduction. This provides new solutions for energy efficiency optimization of university buildings, promoting the realization of green buildings and energy efficiency enhancement and pollutant reduction.
Keywords: Building energy, Anomaly detection, BiLSTM, K-means, DBSCAN

1. INTRODUCTION

According to data from the International Energy Agency (IEA), the building sector accounts for over 30% of global final energy consumption and nearly 40% of direct and indirect carbon dioxide emissions, exerting a profound environmental impact. Energy conservation and emissions reduction in this sector are crucial for mitigating climate change [1]. Public buildings such as office towers and university facilities possess significant energy-saving potential due to their multifunctional nature and high energy consumption (EC). Consequently, research into campus building EC is essential. However, outdated energy efficiency management—relying on manual experience rather than precise data analysis and intelligent tools—has become a barrier to improving efficiency and reducing pollutants [2]. Data mining technologies offer novel solutions for energy management and fault diagnosis. They enable real-time monitoring and analysis of equipment operation, detect abnormal fluctuations in EC for precise adjustments, and predict failures and enable preventive maintenance through historical data, thereby reducing energy waste and economic losses [3]. Among these, Building Energy Management Systems (BEMS) can monitor building energy consumption (BEC) in real time, playing a significant role in enhancing energy efficiency, reducing

pollutants, and lowering operational costs [4]. BEMS is widely used in various fields, and existing research can be roughly divided into energy management optimization, EC prediction models, or anomaly detection. Researchers have made significant progress in optimizing energy management. For example, the research in references [5] - [8] focuses on optimizing cost, equipment, resources, and management models through BEMS. Regarding the EC prediction model and anomaly detection field, scholars in references [9] - [12] have conducted in-depth analysis by combining artificial intelligence (AI) and deep learning techniques.

In summary, although BEMS has made significant progress in the design and application of existing research, it cannot effectively solve the problems of inaccurate data processing, insufficient model optimization, and poor adaptability to multiple scenarios in existing research. For example, some studies use single traditional K-means clustering or standard LSTM prediction models that are sensitive to parameters and susceptible to outlier interference, resulting in poor generalization ability under different operating conditions or building types.

Therefore, this study innovatively proposes a framework combining ensemble learning and metaheuristic optimization, whose core contribution lies in:

(1) A multi-level robust clustering mechanism combining entropy weight (EW) and K-means algorithm (EWKM) and density-based spatial clustering of applications with noise (DBSCAN) is proposed.

(2) An improved sparrow search algorithm (ISSA) is designed, which introduces optimal set initialization, reverse learning, and dynamic weighted search mechanism to achieve automatic optimization of bidirectional long short-term memory (BiLSTM) hyperparameters. ISSA-BiLSTM method effectively improves the convergence speed and prediction accuracy (PA) of the model, overcoming the problem of traditional optimization algorithms easily falling into local optima.

(3) By combining real-time monitoring with dynamic prediction, real-time diagnosis of energy consumption anomalies and dynamic adjustment of management strategies have been achieved. This provides a more adaptable solution for energy management across multiple building types and operating conditions.

The comparison of different studies is shown in Table 1.

Table 1. Comparison of Different Studie

Research method	Research results	Author
Dynamic programming and Bald Eagle search optimization algorithm	Cost reduction: 15.3%; Convergence: 280 iterations	H. Youssef et al.[5]
Interactive Intelligent BEMS	Control accuracy: 94.2%; Response time: 1.8s	M. U. Shoukat et al.[6]
Life Cycle Assessment	Energy efficiency: 23.5% improvement	A. Polyanska et al.[7]
Standardization of BEMS development using open-source modeling libraries	Model development time: reduced by 35%	L. Maier et al.[8]
AI driven BEMS	Real-time capability: 95.8%; Operational efficiency: 28.7%	E. Tabaku et al.[9]
Entangled LSTM architecture	Prediction Mean Absolute Error (MAE): 4.85%; Convergence: 310 iterations	K. Sharma et al.[10]

CNN-LSTM hybrid model	Prediction R^2: 0.89; Training time: 4.2h	A. Kristian et al.[11]
Neural network+logistic regression+decision tree	Accuracy: 98.96%; Generalization: 82.3%	B. Mehdizadeh et al.[12]

2. METHODS AND MATERIALS

2.1. DESIGN OF BEMS

University buildings often have the characteristics of large land area, multiple building types, and complex functional divisions, among which the EC environment is complex and the types of EC are diverse. The manual management of energy has low efficiency and high cost, making it impossible to achieve effective scientific management of EC [13]. Therefore, universities have designed an intelligent BEMS to achieve refined energy-saving and consumption reduction management, and improve energy efficiency. The functional framework of BEMS is shown in Figure 1.

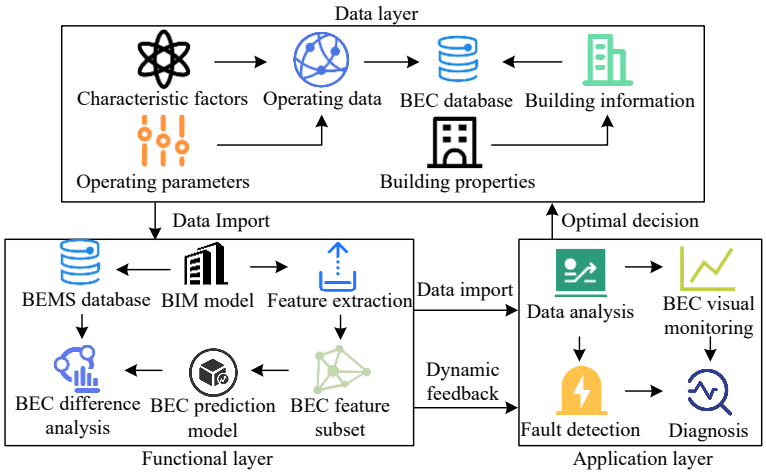


Figure 1. Functional framework of BEMS

In Figure 1, the design of BEMS includes three main layers: data layer, functional layer, and application layer. The data layer bears accountability for collecting data from various devices and sensors inside the building. The collected high-dimensional data are imported into the Building Information Modeling (BIM) in the functional layer to provide data support for the implementation of lower level algorithms. Key EC characteristics are extracted from the BEC database and the data are used to establish a BEC prediction model, which predicts future EC grounded on historical data. A subset of BEC features is obtained through feature extraction, and after training, the BEC prediction model is used to obtain forecasted values. Data mining and analysis techniques are utilized in the application layer to conduct in-depth analysis of the collected BEC data and generate visual reports. The core technologies of this layer include AI, machine learning, and big data analysis, which help managers understand the current situation and trends of BEC, thereby achieving anomaly diagnosis.

2.2. DATASET DESCRIPTION AND DATA PREPROCESSING

The experimental dataset used in this study was sourced from the ASHRAE global database (publicly available, doi.org/10.6078/D1F671). This dataset covers the global building stock of 109033 real buildings over the past 20 years. This study extracted power data from one of the university buildings from January to October 2024, with a collection interval of once per hour and

a total of 6480 data records. The dataset is divided into 70% training set, 15% validation set, and 15% testing set. Given the presence of data anomalies and noise interference, the BEC data requires cleansing to uphold data quality. The data preprocessing stage first performs cleaning and quality enhancement on the raw data. The handling of missing values shall be carried out in accordance with ASHRAE Guideline 14-2014 and ISO 50001:2018 standards. Linear interpolation is used to fill in missing data collected by sensors with a duration not exceeding 1 hour. Data that are missing for more than 24 hours will be excluded and considered a system failure requiring equipment inspection. The outlier processing uses is Inter-Quartile Range (IQR) for identification and removal, with upper and lower limits of S_u and S_d, respectively. The expression is shown in equation (1) [14].

$$S_u = q1 - 1.5 \times \mathrm{IQR}, S_d = q3 + 1.5 \times \mathrm{IQR} \quad (1)$$

In equation (1), q_1 and q_3 are the first and third quartiles. On this basis, standardize all continuous variables to improve data consistency and model training stability.

2.3. BEC DATA ANOMALY DIAGNOSIS BASED ON EWKM AND DBSCAN

The anomalies in BEC data have multiple sources, including global discrete point anomalies caused by equipment or sensor failures, as well as local collective anomalies caused by special events such as holidays and large-scale events. A single algorithm is difficult to balance two types of anomaly detection: traditional clustering methods such as K-means are sensitive to noise, while DBSCAN, although good at identifying local anomalies, lacks the ability to weight high-dimensional features. Aiming at the multimodal, context sensitive, and high-dimensional noise characteristics of EC data, EWKM and DBSCAN methods were designed specifically. The implementation steps for abnormal diagnosis of BEC data in the research design are shown in Figure 2.

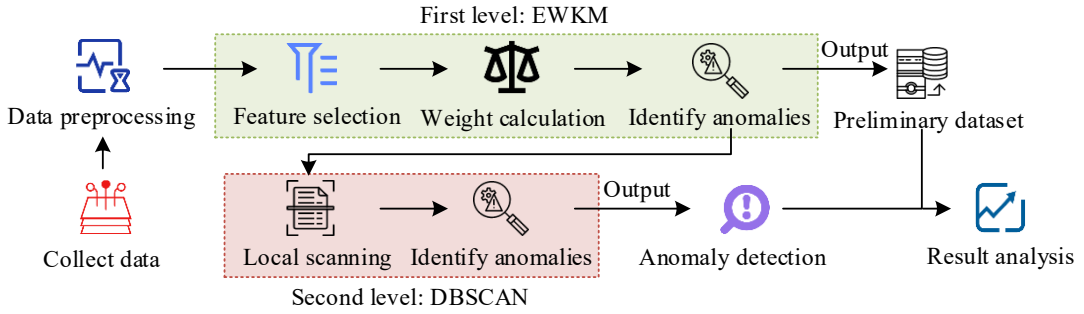


Figure 2. Implementation steps for BEC data anomaly diagnosis

In Figure 2, the study gathers BEC data encompassing various metrics like electricity usage, water consumption, air conditioning EC, temperature, and equipment status. The first level adopts EW method weighted K-means clustering, which achieves soft noise reduction through feature weight allocation, suppressing noise while retaining the main energy consumption patterns, and preliminarily screening global discrete anomalies. The second level further identifies contextual collective anomalies that EWKM failed to capture through DBSCAN. This hybrid strategy significantly improves the anomaly detection accuracy of high-dimensional energy consumption data in complex scenarios through functional complementarity [15]. The combination of EW and K-means utilizes EW to weight features, thereby transforming the distance measurement standard of K-means. The specific steps are as follows: Firstly, the EW method is employed to calculate the weight of each feature. Features with lower entropy values represent greater information content, and therefore should be given higher weights. EW is used to measure the importance of features and

ensure that the K-means clustering algorithm focuses on important features. The research collects the EC influencing factors of i buildings, and establishes a BEC factor dataset $F=\{f_1,f_2,...,f_j,...,f_i\}, f_j=\{f_{j1},...,f_{ju},...,f_{jn}\}$, where f_{jn} is a dataset of n influencing factors collected for Building j. Secondly, during the iteration process of the K-means algorithm, the weights calculated using EW are used to redefine the distance between data points and cluster centers. The traditional Euclidean distance has been improved to a weighted Euclidean distance, and its calculation formula is shown in equation (2) [16].

$$J=(\mathbf{W},\mathbf{Z},\mathbf{\Lambda})=\sum_{l=1}^{k}\left[\sum_{j=1}^{i}\sum_{u=1}^{n}w_{lj}\lambda_{lu}\left(z_{lu}-x_{ju}\right)^2+\gamma\sum_{u=1}^{n}\lambda_{lu}\,lg\,\lambda_{lu}\right]\lambda \tag{2}$$

In equation (2), $\mathbf{W}$ is the allocation matrix, with a size of $k\times i$. $\mathbf{Z}$ is the current cluster center matrix, $\mathbf{\Lambda}$ is the matrix corresponding to λ. λ is the weight corresponding to each cluster, with dimension $k\times m$. γ is the parameter, $\gamma>0$. Finally, perform weighted clustering, where the algorithm allocates data points and updates cluster centers based on this weighted distance. This indicates that features with high information content have greater weight in determining the attribution of data points. To avoid entropy calculation failure caused by zero values, add a small perturbation term 1e-8 to the normalized data. The number of clusters K in the EWKM algorithm is automatically determined by the Davies Bouldin Index (DBI) and silhouette coefficient [17]. To further identify anomalies, the study chose the DBSCAN algorithm. The DBSCAN can effectively identify noise and outliers, and is suitable for detecting outliers in BEC data. Based on EWKM clustering results, DBSCAN algorithm is applied to detect noise points and outliers, and abnormal data points are marked. The process design of DBSCAN algorithm is in Figure 3.

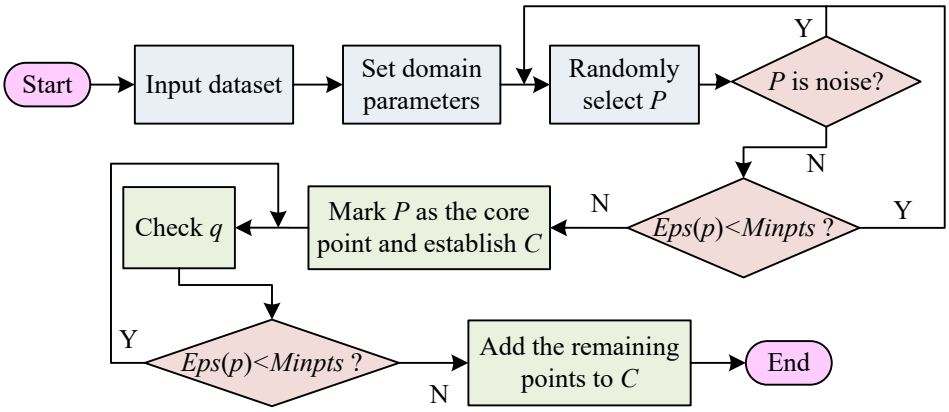


Figure 3. DBSCAN algorithm process

In Figure 3, the dataset is input, and the algorithm sets the domain parameter neighborhood Eps and neighborhood density threshold $Minpts$. The algorithm randomly selects object P from the dataset that has not been checked yet. If P is not processed, that is, it is not classified as a cluster or marked as noise, then Eps is checked. If the number of objects included is not less than $Minpts$, the algorithm marks P as the core point and establishes a new cluster C. All points of P are added to C. All unprocessed objects q in C have their Eps checked. If there are at least $Minpts$ objects, the algorithm adds these objects to C. If q does not belong to any cluster, the algorithm will add q to C. The steps are repeated, the unprocessed objects in C are further checked, and the current candidate set C is empty. The steps are repeated until all objects are

assigned to a cluster or marked as noise.

2.4. BEC PREDICTION MODEL BASED ON ISSA AND BILSTM

In addition to anomaly diagnosis to identify faults in energy management, EC prediction is one of the important steps in BEMS. The feature subset obtained from clustering results is input into the EC prediction model, and BiLSTM can process time series prediction data in BEC prediction. The BiLSTM model combines forward and backward LSTM units by introducing a bidirectional structure, enabling it to learn from both past and future data simultaneously. The forward LSTM unit processes data in the forward order of the sequence, while the reverse LSTM unit processes data in the reverse order [18]. BiLSTM can more comprehensively capture the correlation information in time series, improving the understanding and representation of sequence features The forward and reverse Hidden Layer (HL) states are $\vec{h}_t$ and $\overleftarrow{h}_t$. The expressions involved in this process are shown in equation (3).

$$\begin{cases} \vec{h}_t = \sigma\left(W_1 x_t + W_2 \vec{h}_{t-1} + b_1\right) \\ \overleftarrow{h}_t = \sigma\left(W_1^{'} x_t + W_2^{'} \overleftarrow{h}_{t-1} + b_2\right) \\ f_t = \sigma\left(W_{\mathrm{f}} \cdot \left[h_{t-1}, x_t\right] + b_{\mathrm{f}}\right) \\ I_t = \sigma\left(W_I \cdot \left[h_{t-1}, x_t\right] + b_I\right) \\ \tilde{C}_t = \mathrm{Tanh}\left(W_C \cdot \left[h_{t-1}, x_t\right] + b_C\right) \\ O_t = \sigma\left(W_O \cdot \left[h_{t-1}, x_t\right] + b_O\right) \\ h_t = O_t * \mathrm{Tanh}\left(C_t\right) \end{cases} \tag{3}$$

In equation (3), σ is the sigmoid activation function, and x_t is the dataset of EC impact factors input at time t. W_1 and $W_1^{'}$ are the weight matrices from the input data to the HL state in the forward and backward LSTM units, respectively. b_1 and b_2 are the forward and reverse output layer bias controls, respectively. W_2 and $W_2^{'}$ are the weights of the HL state from time t to time $t-1$ for the forward and backward LSTM units, respectively. f_t is the forget gate, which determines which information to discard from the cellular state. W_f, b_f represents the weight and bias of f_t, respectively. I_t is the input gate. W_I, b_I represents the weight and bias of I_t respectively. O_t is the output gate. W_O, b_O represents the weight and bias of O_t, respectively. C_t and $\tilde{C}_t$ represent cell state and candidate cell state, respectively. W_C, b_C represents the weight and bias of $\tilde{C}_t$, respectively. Tanh is the activation function. The final BEC prediction value Y_t after combining forward and reverse directions is expressed in equation (4) [19].

$$Y_t = \alpha_1 \vec{h}_t + \alpha_2 \overleftarrow{h}_t \tag{4}$$

In equation (4), α_1 and α_2 are the positive and negative factors. The prediction model is trained and the prediction results are optimized to ensure accuracy and stability. To address the shortcomings of time-consuming and unstable manual parameter tuning, ISSA is studied for automatic parameter tuning of BiLSTM to improve the accuracy and efficiency of the model. The

process of BEC prediction model is shown in Figure 4.

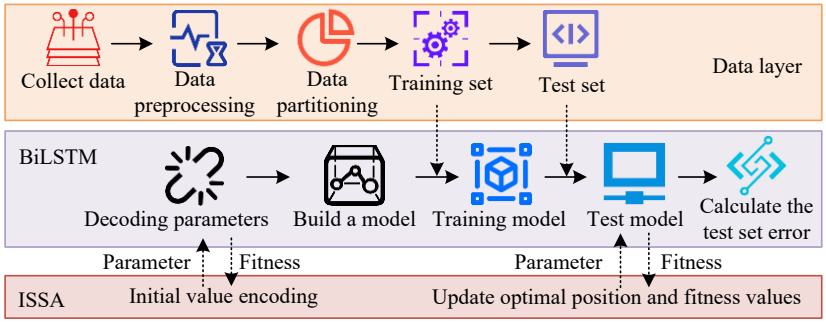


Figure 4. BEC prediction model process flow

In Figure 4, the ISSA-BiLSTM model applies the ISSA to hyper-parameter optimization of the BiLSTM model. It randomly generates a set of hyper-parameters for the BilSTM model to form the initial population for the ISSA. The BiLSTM model is trained using the training dataset and the model's PA is evaluated using the validation dataset. The PA is used as the fitness value of the ISSA. The ISSA updates hyper-parameters based on fitness values, guiding the hyper-parameter search of the BiLSTM model and continuously iterating and optimizing it. When the preset number of iterations or fitness value is reached and there is no further improvement, the algorithm terminates. The ISSA in this model is responsible for finding the optimal hyper-parameter combination for the BiLSTM model, thereby improving the PA and generalization ability of the model [20].

2.5. ISSA DESIGN

To overcome the shortcomings of traditional SSA in optimizing BiLSTM hyperparameters, such as slow convergence speed, weak global search ability, and susceptibility to local optima, and significantly improve the accuracy of BEC prediction, this study proposes ISSA. The core improvement of ISSA lies in the introduction of three key mechanisms: the good point set mechanism, reverse learning mechanism, and optimization search mechanism. These improvements directly improve the prediction accuracy of BiLSTM from three aspects. One is to initialize the hyperparameter candidate solutions using the good point set, making their distribution more uniform and diverse, avoiding slow convergence caused by random initialization, and laying a solid foundation for optimization. The second is to explore the reverse solution of the symmetric region of the current optimal solution through the reverse learning mechanism, which helps the algorithm to escape from local optima and find the global optimal or better hyperparameter combination more easily. The third is through the optimization search mechanism, the step size is adaptively adjusted during iteration, with a focus on global exploration in the early stage and fine local development in the later stage, achieving stable and efficient approximation of the optimal solution after balancing. The flowchart of ISSA is shown in Figure 5.

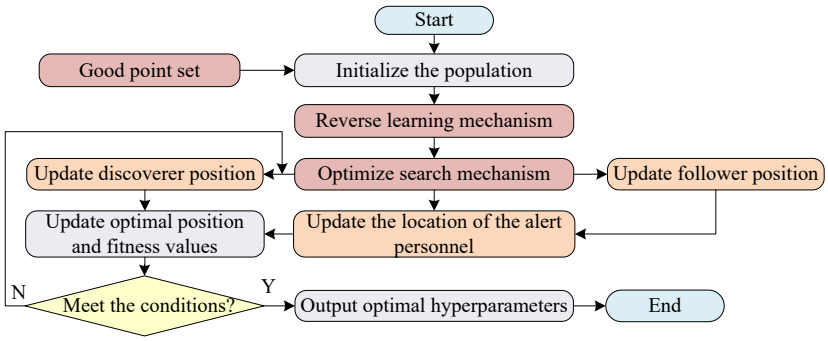


Figure 5. Process diagram of ISSA

In Figure 5, the ISSA first initializes the population and uses the good point set mechanism to initialize the good point set. When an individual updates its location, in addition to updating rules based on the current location, it also incorporates reverse learning mechanisms and optimization search mechanisms. The principle of the good point set is to set G_s as a unit cube in s-dimensional Euclidean space. If the good point $e \in G_s$, the expression for the good point set $J_n^{'}(k)$ is shown in equation (5) [21].

$$\begin{aligned} &J_n^{'}(k)=\left\{\left(\left\{e_1^{(n)}\cdot k\right\},\left\{e_2^{(n)}\cdot k\right\},...,\left\{e_s^{(n)}\cdot k\right\}\right),1\le k\le n\right\}, \\ &e=\left\{2cos\left(2\pi k/q\right),1\le k\le s\right\} \end{aligned} \tag{5}$$

In equation (5), $\left\{e_s^{(n)}\cdot k\right\}$ denotes the decimal part and n' is the number of points. q is the min prime number that satisfies $(q-3)/2\ge s$. $J_n^{'}(k)$ is mapped to the search space as shown in equation (6).

$$x_i(j)=\left(s_j^{'}-c_j^{'}\right)\cdot\left\{e_j^{i}\cdot k\right\}+c_j^{'} \tag{6}$$

In equation (6), $s_j^{'}$ and $c_j^{'}$ are the upper and lower bounds of the j th dimension. ISSA initializes the population through a set of optimal points and incorporates a reverse learning mechanism when updating the positions of followers and alerts. The reverse learning mechanism enables individuals to learn in reverse during the search process, avoiding falling into local optimal regions. This mechanism can guide individuals to explore different solution space regions, thereby enhancing the ability of global search. The reverse solution $R_k^{'}$ of individual X_k is shown in equation (7).

$$R_k^{'}=n_r\cdot\left(s_j^{'}+c_j^{'}\right)+x_k \tag{7}$$

In equation (7), n_r is a random number, $n_r\in(0,1)$. Additionally, to balance the global search and local development of the algorithm, an optimization search mechanism is introduced to boost demographic diversity and utilization of the current solution. This mechanism improves the search efficiency of the algorithm by dynamical weighting factor w, enhancing individuals' search ability in the solution space, avoiding falling into local optima, and raising the algorithm's global

exploration ability. w adopts a linear decreasing strategy for adjustment. The formula for updating the d-dimensional position $X_{i,d}^{t+1}$ of the i th individual in the $t+1$ th generation of the population is shown in equation (8).

$$X_{i,d}^{t+1}=\begin{cases} w\cdot X_{i,d}^{t}+c_1\cdot rand\left(X_{best}^{t}-X_{i,d}^{t}\right)+ & R_2<s_2 \\ c_1\cdot rand\left(X_{i,h}^{t}-X_{i,d}^{t}\right), & \\ X_{i,d}^{t}+\delta\cdot n_s, & R_2\geq s_2 \end{cases} \tag{8}$$

In equation (8), c_1 and c_2 are both learning factors. X_{best}^{t} is the global optimal value for the t th generation. $X_{i,h}^{t}$ is the historical optimal value for the i th individual. n_s is a standard normal distribution random number. R_2 is a uniform random number, $R_2\in[0,1]$. s_2 is the alert threshold, $s_2\in[0.5,1.0]$. The formula for updating the position of the alert personnel is shown in equation (9) [22].

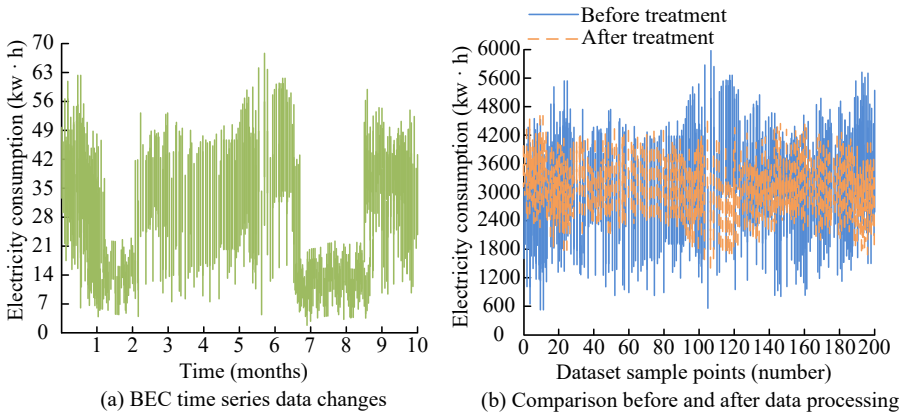
$$X_{i,d}^{t+1}=\begin{cases} X_{best}^{t}+\theta\left(X_{i,d}^{t}-X_{best}^{t}\right), & f_i\neq f_g \\ X_{best}^{t}+K_r\cdot\left(\dfrac{X_{i,d}^{t}-X_{i,h}^{t}}{\left|f_i-f_w\right|+\varepsilon}\right), & f_i=f_g \end{cases} \tag{9}$$

In equation (9), θ is a random number that conforms to the standard normal distribution. K_r is a uniform random number, $K_r\in[-1,1]$. ε is a smaller number to prevent the denominator from being one. f_w is the fitness value of the sparrow in the worst position.

3. RESULTS

3.1. PERFORMANCE ANALYSIS OF BEC FEATURE SELECTION

The experimental software used included Python 3.9 and PyTorch 1.12, physical simulation software EnergyPlus v9.6, visualization tools Grafana v8.3 and Tableau 2022.1. Hardware Dual Intel Xeon Gold 6248R (48 cores), GPU NVIDIA Tesla V100S-32GB, Memory 384GB DDR4 ECC. The trend chart of time series data and the data analysis before and after BEC data processing are shown in Figure 6.



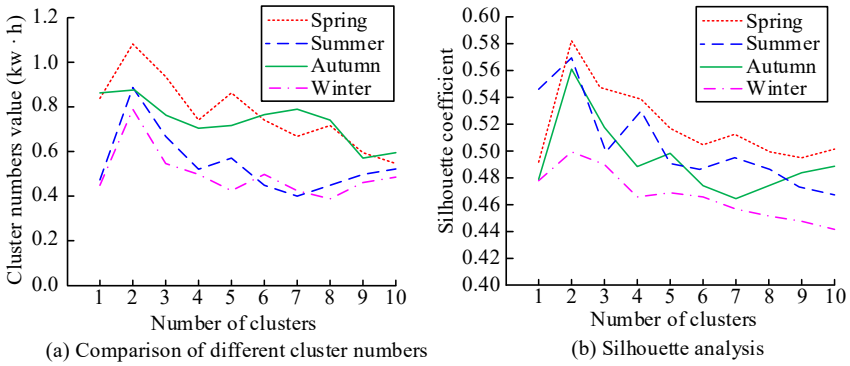
(a) BEC time series data changes

(b) Comparison before and after data processing

Figure 6. Changes in different time series data and changes before and after data processing

Figure 6 (a) shows the time series trend of electricity consumption for different months. The data shows a strong seasonal correlation, with significant differences (p<0.01) in the distribution of electricity consumption data between January, July, and August compared to other months. In January, July, and August, due to a sharp decline in building utilization, there were a large number of extremely low EC days, accounting for over 60%. Therefore, some EC data from January, July, and August were excluded from the model training data. Figure 6 (b) shows the data results after anomaly detection and cleaning. The outliers and noise in the dataset were removed, and the overall trend was more stable. The electricity consumption was more concentrated between 1800kw·h and 4200kw·h. Compared to the original data, the cleaned data exhibited higher reliability.

To determine the optimal number of clusters, the range of cluster values was set to [1, 10], which means calculating the values of different cluster numbers for each season's electricity consumption data to find the most suitable cluster number. Simultaneously perform silhouette analysis to validate optimal cluster numbers per season. The results of these experiments are shown in Figure 7.



(a) Comparison of different cluster numbers (b) Silhouette analysis

Figure 7. EWKM clustering results of data

In Figure 7 (a), the highest values for spring, summer, autumn, and winter were 1.10, 0.88, 0.85, and 0.79, respectively. Therefore, the optimal number of clusters for all four seasons was 2. This was because the values of spring, summer, and winter reached their highest when there were two clusters, indicating that EWKM had the best clustering effect on the current season's electricity consumption data. The result of silhouette analysis is shown in Figure 7 (b). The average silhouette coefficient reaches its peak when the number of clusters is 2. The average silhouette coefficient is 0.537 in spring, 0.529 in summer, 0.518 in autumn, and 0.472 in winter. This indicates that the two clusters have achieved the best balance between cohesion within the clusters and separation between clusters in each season.

To further analyze the abnormal situation of BEC, the electricity consumption data for spring and autumn 2024 were tested separately. The electricity consumption results within 60 days are shown in Figure 8.

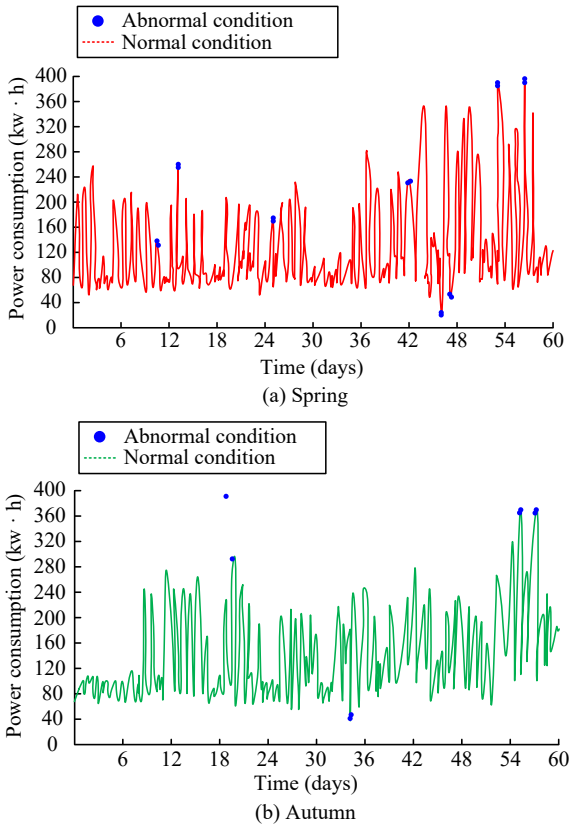


Figure 8. Abnormal electricity consumption detection in spring and autumn

In Figure 8 (a), the DBSCAN algorithm successfully identified 16 abnormal points in spring and autumn electricity consumption data. These outliers mainly occurred during some atypical electricity usage periods, due to electricity fluctuations caused by weekends or special events. In Figure 8 (b), a total of 8 anomalies were detected in autumn electricity consumption, which occurred between 10am and 3pm. This was related to large-scale course arrangements and activities. For further validation, Tukey's Fence method was employed in statistical testing to corroborate the identified anomalies, specifically 16 in spring and 8 in autumn. The results showed that these abnormalities deviated significantly from the normal distribution (p<0.01).

3.2. ISSA PERFORMANCE ANALYSIS

To evaluate the contribution of different mechanisms in ISSA, ablation experiments were conducted. The results of the ablation experiment are shown in Table 2.

Table 2. Ablation study of ISSA components

Component	Convergence iterations	Changes in convergence iterations	MAE (%)	MAE changes	Local optima escapes	Changes in local optima escapes
ISSA	142	/	3.33	/	89%	/
No reverse learning	178	+25%	3.74	+12.3%	62%	-30.3
No good point set	165	+16%	3.65	+9.6%	71%	-20.2%
Basic SSA	218	+54%	5.12	+53.8%	34%	-61.8%

In Table 2, ISSA only requires 142 iterations to converge, while SSA requires 218 iterations. The good point set initialization and optimization search mechanism introduced by ISSA effectively improves the search efficiency of the algorithm, enabling it to find the optimal solution at a faster speed. The local optimal escape rate of ISSA is as high as 89%, while SSA is only 34%. ISSA has improved by 55%, mainly due to the reverse learning mechanism. This mechanism allows the algorithm to explore the reverse region of the current optimal solution, effectively escaping the local optimal trap and having a higher probability of finding the global optimal solution. The MAE of ISSA-BiLSTM is 3.33%, which is its final predictive performance on an independent test set.

To confirm the optimization effect of ISSA, the study selected Particle Swarm Optimization (PSO), Grey Wolf Optimization (GWO), and SSA for comparison. The study chose to conduct experiments on the Congress on Evolutionary Computation (CEC) test function set, which serves as a standard test function set covering various complex problems [23]. The changes in fitness values on high-dimensional and low dimensional test functions are shown in Figure 9.

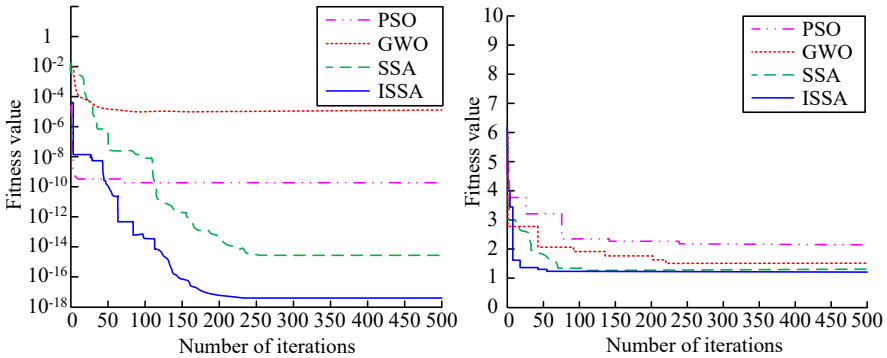


Figure 9. Changes in fitness values on high-dimensional and low dimensional test functions

In Figure 9 (a), for high-dimensional unimodal functions, the ISSA had better optimization performance than other algorithms. The standard deviation, mean, and optimal value of ISSA were 9.15×10^{-17}, 0.96×10^{-17}, and 2.18×10^{-18}, respectively. The optimal values for PSO, GWO, and SSA were 6.74×10^{-9}, 5.21×10^{-7}, and 5.65×10^{-15}, respectively. In Figure 9 (b), the standard deviation, mean, and optimal value of ISSA were 1.011, 1.005, and 1.050, respectively. The optimal values for PSO, GWO, and SSA were 2.045, 1.420, and 1.091, respectively. The optimization effect of SSA low dimensional test function was closer to ISSA, PSO had the worst optimization effect, and GWO had average optimization effect.

The study compares the performance and PA of ISSA-BiLSTM with baseline models. The relevant results are shown in Table 3.

Table 3. Performance comparison of different models

Model	Training time (h)	Memory usage (GB)	MAE (%)	Convergence iteration times	NSE	WI	PBIAS (%)
Vanilla BiLSTM	3.8	9.2	4.51	320	0.82	0.88	+5.2
SSA-BiLSTM	2.9	8.7	3.87	240	0.89	0.92	+2.8
ISSA-BiLSTM	2.1	7.5	3.33	142	0.93	0.94	-1.3

In Table 3, The difference in MAE between ISSA-BiLSTM and Vanilla BiLSTM is significant

(t=8.15, p<0.001, Cohen's d=1.52, Belonging to large effect quantity). There was a significant difference in MAE between ISSA-BiLSTM and SSA-BiLSTM (t=5.72, p<0.001, Cohen's d=1.21, belonging to large effect quantity). When Nash Sutcliffe Efficiency (NSE)>0.9, it indicates a high degree of fit between the predicted and actual values of the model, and the closer Willmott Index (WI) is to 1, the more accurate the model's prediction. The NSE of ISSA-BiLSTM is 0.93, WI is 0.94, and its Percent Bias (PBIAS) is -1.3%. Compared to vanilla BiLSTM, it only increased by 2.3% while still maintaining its optimal performance. ISSA-BiLSTM saves 44.7% of time and reduces memory usage by 18.5% compared to Vanilla BiLSTM. The good point set mechanism ensures uniform distribution of the initial population, reduces ineffective searches, and reduces the number of iterations by 40.8% compared to SSA-BiLSTM. Despite a 13.8% reduction in memory, the ISSA dynamic weighting strategy still resulted in a 14.2% decrease in MAE compared to SSA-BiLSTM.

To verify the effectiveness of the proposed method, it was compared with Transformer-based model, Informer, and Temporal Convolutional Network (TCN). The results are shown in Table 4.

Table 4. Performance results of different methods

Method	Training time (h)	Identification accuracy (%)	F1-Score	Parameter quantity (M)
Proposed method	2.1	92.6	0.93	1.2
Transformer	6.8	85.9	0.89	3.5
Informer	5.2	88.4	0.91	2.8
TCN	3.5	83.5	0.85	1.8
K-means+LSTM	4.2	78.9	0.72	1.5

In Table 4, the training time of the proposed method is 2.1 hours. From the perspective of multi scenario adaptation, the proposed method has the highest recognition accuracy of 92.6%. This relies on the dynamic parameter tuning advantage of ISSA BiLSTM in the proposed method, which automatically adjusts the time window and hidden layer nodes for campus buildings. In contrast, the fixed structure in Transformer based models requires retraining, resulting in the longest training time of 6.8 hours and a recognition accuracy of only 85.9%. The F1 Score of the proposed method is 0.93, which is 0.21 higher than K-means+LSTM.

3.3. PERFORMANCE ANALYSIS OF BEC PREDICTION MODEL

To determine the optimal values of hyper-parameters for the prediction model, ISSA was used for tuning the BiLSTM model. The iterative process of obtaining the optimal values of hyper-parameters is shown in Figure 10.

In Figure 10 (a), the number of cycles set at the end of the final iteration was 9. The learning rate was 8.9×10^{-3}. By optimizing the learning rate through ISSA, the BiLSTM model could converge quickly while maintaining stability. The final optimized learning rate could prevent oscillation and instability while ensuring convergence speed. In Figure 10 (b), the first HL contained 65 nodes, and the second one had 84 nodes. The size of the HL directly affected the expressive power of BiLSTM. The optimized node numbers were 65 and 84, which not only enhanced the model's capability but also ensured reasonable computational efficiency and avoided the risk of overfitting.

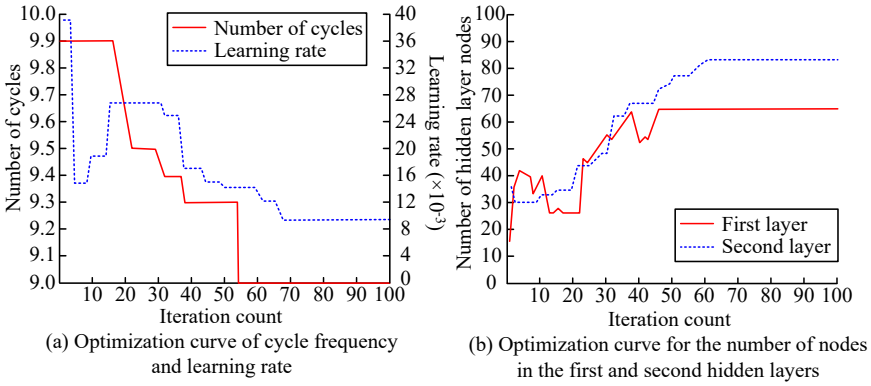


Figure 10. Optimization iteration curve variation

To confirm the PA of the proposed BEC prediction model, BiLSTM, SSA-BiLSTM, PSO-BiLSTM, and GWO-BiLSTM models were selected for comparison. The study trained and tested these 5 models in each season, calculated R^2, Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE) for each item. The PA results of different models in the four seasons are in Figure 11.

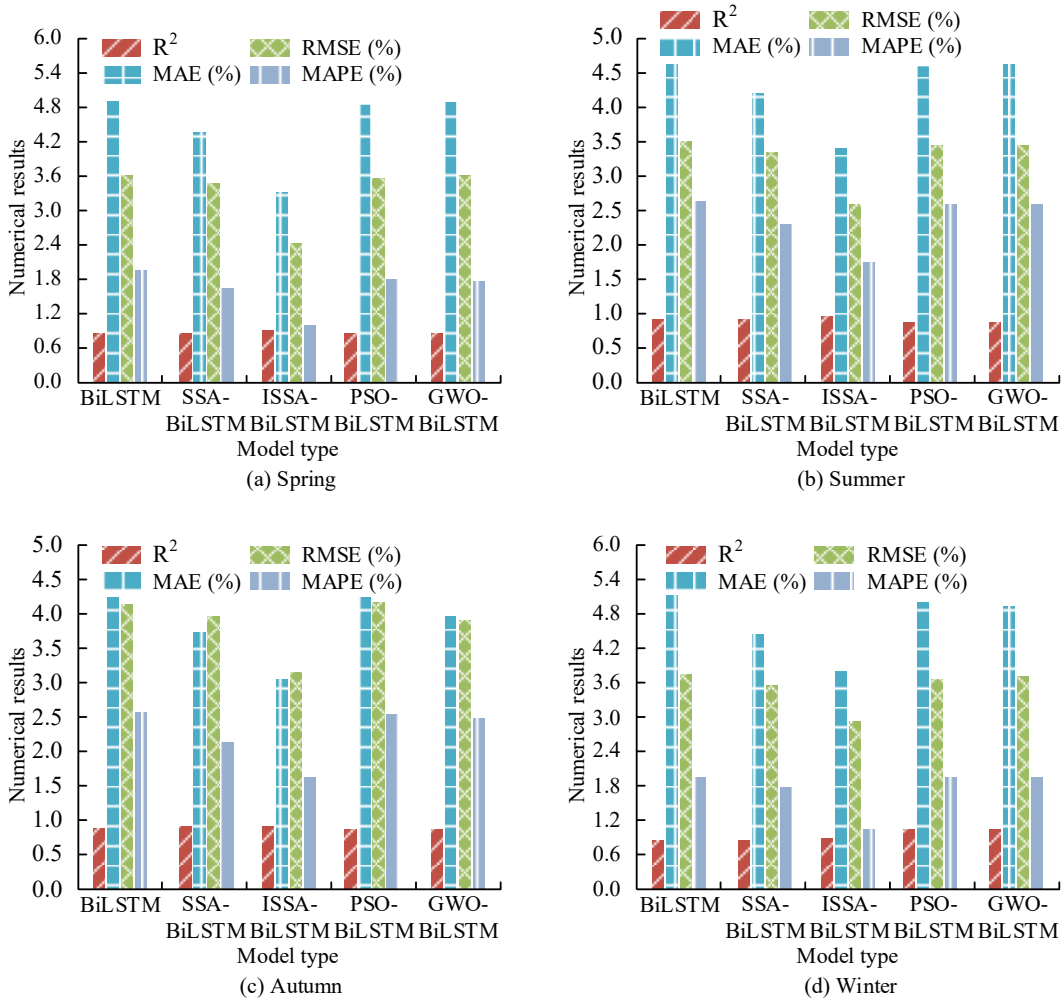


Figure 11. BEC PA outcomes of various models

Figures 11 (a) to (d) show the PA results of different models in the four seasons of spring, summer, autumn, and winter, respectively. The R^2 values of ISSA-BiLSTM were generally high in all seasons, showing stronger predictive ability than other models in spring, summer, and winter, with the highest reaching 0.96. The MAE of ISSA-BiLSTM in spring is 3.33%, which is 31.5% lower than that of PSO-BiLSTM (4.86%). The MAPE of ISSA-BiLSTM is 1.01%. ISSA-BiLSTM could more accurately predict the EC of buildings. Compared to single BiLSTM, PSO, and GWO, it had stronger global search ability when optimizing model parameters, avoiding getting stuck in local optima. ISSA-BiLSTM showed higher accuracy in all seasons, with better R2 and MAPE values than other comparison models, and had good performance in BEC prediction.

4. DISCUSSION AND CONCLUSION

To achieve building energy efficiency enhancement and pollutant reduction, BEMS was developed, and data anomaly detection was achieved through optimized clustering algorithms. Meanwhile, a BiLSTM model was built in combination with ISSA to predict BEC. The results showed that the clustering algorithm proposed in the study resulted in the MAE and RMSE of the model being 18.14 and 20.33, respectively. The MAE and RMSE were significantly reduced, and the PA of the model was improved. The research of M. Ordouei et al. did not use data mining techniques. Instead, they chose a reinforcement learning framework to analyze the EC of various uses in cities, and found that the annual EC of cities was reduced by at least 35% [24]. In contrast, the data mining technique EMKM used in the study could effectively handle outliers and noise in the data, further improving the PA of the model. This is because EWKM, which combines EW and K-means, assigns weights to different features based on data-driven methods, which can solve the problems of inconsistent feature dimensions and uneven contribution in high-dimensional data.

In addition to K-means clustering and DBSCAN clustering algorithms, M.E. Ramachandran et al. proposed a decision tree, naive Bayes, and SVM method to process clustering patterns and classified and processed raw data for power load forecasting. The results showed that the accuracy of the method reached 96.4% [25]. Compared with decision tree clustering, the combination of EWKM and DBSCAN in this study improved the F1 score of anomaly detection from 0.72 to 0.93. M. A. Alghamdi et al. used stacked LSTM combined with meta learners and fast Fourier transform to analyze power consumption for EC prediction, and the results showed that the RMSE of this method was 0.020 [26]. Although these methods could handle classification problems, the BiLSTM model used in this study was better suited for handling time series data and nonlinear relationships, resulting in better performance in BEC prediction.

This study proposed ISSA, which incorporated three mechanisms based on SSA: the good point set mechanism, reverse learning mechanism, and optimization search mechanism. The number of cycles for ISSA was set to 9, the learning rate was 8.9×10-3, the number of nodes in the first hidden layer was 65, and the number of nodes in the second hidden layer was 84. By optimizing the hyper-parameters of the BiLSTM model through ISSA, the PA and training efficiency of the model on BEC data were improved. Therefore, this study was able to achieve better prediction results. However, the proposed model did not take into account the impact of CO_2 emission reduction data on the BEMS system. Future research can couple relevant CO_2 emission reduction data with carbon quota trading systems to deeply analyze the economic optimization path of BEMS.

FUNDING

No funding received.

REFERENCE

[1] R. K. T., "Energy Efficiency and Integration of Renewable Energy Sources in Architectural Design: Techno-Economic Possibilities of Azerbaijan," The American Journal of Applied Sciences, vol. 7, no. 3, pp. 27-44, Mar. 2025, 10.37547/tajas/Volume07Issue03-05A.

[2] P. Bouquet, I. Jackson, M. Nick, and A. Kaboli, "AI-based forecasting for optimised solar energy management and smart grid efficiency," Int. J. Prod. Res., vol. 62, no. 13, pp. 4623-4644, Sep. 2024, 10.1080/00207543.2023.2269565.

[3] A. Heidari, Y. Peyvastehgar, and M. Amanzadegan, "A systematic review of the BIM in construction: From smart building management to interoperability of BIM & AI," Archit. Sci. Rev., vol. 67, no. 3, pp. 237-254, Aug. 2024, 10.1080/00038628.2023.2243247.

[4] L. Y. Rock, F. P. Tajudeen, and Y. W. Chung, "Usage and impact of the internet-of-things-based smart home technology: a quality-of-life perspective," Univ. Access Inf. Soc., vol. 23, no. 1, pp. 345-364, Nov. 2024, 10.1007/s10209-022-00937-0.

[5] H. Youssef, S. Kamel, M. H. Hassan, L. Nasrat, and F. Jurado, "An improved bald eagle search optimization algorithm for optimal home energy management systems," Soft Comput., vol. 28, no. 2, pp. 1367-1390, May. 2024, 10.1007/s00500-023-09471-4.

[6] M. U. Shoukat, L. Yan, J. Zhang, Y. Cheng, M. U. Raza, and A. Niaz, "Smart home for enhanced healthcare: exploring human machine interface oriented digital twin model," Multimed. Tools Appl., vol. 83, no. 11, pp. 31297-31315, Sep. 2024, 10.1007/s11042-023-16875-9.

[7] A. Polyanska, Y. Pazynich, K. Mykhailyshyn, D. Babets, and P. Toś, "Aspects of energy efficiency management for rational energy resource utilization," Rud.-Geolo.-Naft. Zb., vol. 39, no. 3, pp. 13-26, Apr. 2024, 10.17794/rgn.2024.3.2.

[8] L. Maier, D. Jansen, F. Wüllhorst, M. Kremer, A. Kümpel, T. Blacha, and D. Mü ller, "AixLib: an open-source Modelica library for compound building energy systems from component to district level with automated quality management," J. Build. Perform. Simul., vol. 17, no. 2, pp. 196-219, Aug. 2024, 10.1080/19401493.2023.2250521.

[9] E. Tabaku, E. Vyshka, R. Kapç iu, A. Shehi, and E. Smajli, "Utilizing artificial intelligence in energy management systems to improve carbon emission reduction and sustainability," J. Ilm. Ilmu Terap. Univ. Jambi, vol. 9, no. 1, pp. 393-405, Mar. 2025, 10.22437/jiituj.v9i1.38665.

[10] K. Sharma, Y. K. Dwivedi, and B. Metri, "Incorporating causality in energy consumption forecasting using deep neural networks," Ann. Oper. Res., vol. 339, no. 1, pp. 537-572, Jul. 2024, 10.1007/s10479-022-04857-3.

[11] A. Kristian, T. S. Goh, A. Ramadan, A. Erica, and S. V. Sihotang, "Application of AI in optimizing energy and resource management: Effectiveness of deep learning models," Int. Trans. Artif. Intell., vol. 2, no. 2, pp. 99-105, Jun. 2024, 10.33050/italic.v2i2.530.

[12] B. Mehdizadeh Khorrami, A. Soleimani, A. Pinnarelli, G. Brusco, and P. Vizza, "Forecasting heating and cooling loads in residential buildings using machine learning: a comparative study of techniques and influential indicators," Asian J. Civ. Eng., vol. 25, no. 2, pp. 1163-1177, Jul. 2024, 10.1007/s42107-023-00834-8.

[13] A. M. Usman and M. K. Abdullah, "An assessment of building energy consumption characteristics using analytical energy and carbon footprint assessment model," Green Low-Carbon Econ., vol. 1, no. 1, pp. 28-40, Mar. 2023, 10.47852/bonviewGLCE3202545.

[14] W. N. Digitemie and I. O. Ekemezie, "A comprehensive review of Building Energy Management Systems BEMS) for improved efficiency," World J. Adv. Res. Rev., vol. 21, no.

3, pp. 829-841, Mar. 2024, 10.30574/wjarr.2024.21.3.0746.

[15] H. Zhang, C. Yu, M. Zeng, T. Ye, D. Yue and C. Dou, "Homomorphic encryption-based resilient distributed energy management under cyber-attack of micro-grid with event-triggered mechanism," IEEE Trans. Smart Grid, vol. 15, no. 5, pp. 5115–5126, April, 2024, 10.1109/TSG.2024.3390108.

[16] X. Wang, H. Wang, B. Bhandari, and L. Cheng, "AI-empowered methods for smart energy consumption: A review of load forecasting, anomaly detection and demand response," Int. J. Precis. Eng. Manuf.-Green Technol., vol. 11, no. 3, pp. 963-993, Sep. 2024, 10.1007/s40684-023-00537-0.

[17] C. N. Egwim, H. Alaka, O. O. Egunjobi, A. Gomes, and I. Mporas, "Comparison of machine learning algorithms for evaluating building energy efficiency using big data analytics," J. Eng. Des. Technol., vol. 22, no. 4, pp. 1325-1350, Sep. 2024, 10.1108/JEDT-05-2022-0238.

[18] Y. K. Ahranjani, M. Beiraghi, and R. Ghanizadeh, "Short time load forecasting for Urmia city using the novel CNN-LTSM deep learning structure," Electr. Eng., vol. 107, no. 1, pp. 1253-1264, May. 2025, 10.1007/s00202-024-02361-4.

[19] A. A. Nasser and A. S. A. Alghawli, "Evaluation and clustering of health security performance in Africa: A comparative analysis through the entropy-TOPSIS-K-means approach," Afr. Secur. Rev., vol. 33, no. 3, pp. 330-348, Jun. 2024, 10.1080/10246029.2024.2367967.

[20] S. Jia, H. Yuan, L. Zhang, Z. Liang, and Z. Yinfei, "Detecting outliers in crowdsourced bathymetric data based on intelligently optimized 3D-DBSCAN algorithm," Mar. Geod., vol. 47, no. 1, pp. 41-58, Oct. 2024, 10.1080/01490419.2023.2279086.

[21] A. Bhattacharya and P. Majumdar, "Artificial intelligence-machine learning algorithms for the simulation of combustion thermal analysis," Heat Transf. Eng., vol. 45, no. 2, pp. 176-193, Feb. 2024, 10.1080/01457632.2023.2178282.

[22] D. Hu, H. Sun, P. Mehrabi, Y. A. Ali, and M. Al-Razgan, "Application of artificial intelligence technique in optimization and prediction of the stability of the walls against wind loads in building design," Mech. Adv. Mater. Struct., vol. 31, no. 19, pp. 4755-4772, May. 2024, 10.1080/15376494.2023.2206208.

[23] P. Kumar and P. Samui, "Reliability-based load and resistance factor design of an energy pile with CPT data using machine learning techniques," Arab. J. Sci. Eng., vol. 49, no. 4, pp. 4831-4860, Sep. 2024, 10.1007/s13369-023-08253-2.

[24] M. Ordouei, A. Broumandnia, T. Banirostam, and A. Gilani, "Optimization of energy consumption in smart city using reinforcement learning algorithm," Int. J. Nonlin. Anal. Appl., vol. 15, no. 1, pp. 277-290, Jul. 2024, 10.22075/IJNAA.2022.29258.4102.

[25] M. E. Ramachandran, R. R. Singh, G. B. Balachandran, D. P. Mohan, P. W. David, M. Anantharaman, and N. Ganesan, "Energy monitoring for renewable energy system using machine learning algorithms," Recent Adv. Electr. Electron. Eng., vol. 17, no. 10, pp. 966-975, Oct. 2024, 10.2174/0123520965258879231011182850.

[26] M. A. Alghamdi, S. Abdullah, and M. Ragab, "Predicting energy consumption using stacked LSTM snapshot ensemble," Big Data Min. Anal., vol. 7, no. 2, pp. 247-270, Jun. 2024, 10.26599/BDMA.2023.9020030.

Submitted: 06.08.2025.
Revised: 10.09.2025.
Accept: 21.10.2025.