Environment and Ecology Research 13(1): 25-37, 2025 http://www.hrpub.org
DOI: 10.13189/eer.2025.130103

An Empirical Study of Meteorological Data-driven Electric Energy Consumption Prediction for Smart Campuses

Qiangjun Liu

Engineering Technology Management Program, International College, Krirk University, Thailand

Received August 16, 2024; Revised December 21, 2024; Accepted January 13, 2025

Cite This Paper in the Following Citation Styles

(a): *[1] Qiangjun Liu , "An Empirical Study of Meteorological Data-driven Electric Energy Consumption Prediction for Smart Campuses," Environment and Ecology Research, Vol. 13, No. 1, pp. 25 - 37, 2025. DOI: 10.13189/eer.2025.130103.*

(b): *Qiangjun Liu (2025). An Empirical Study of Meteorological Data-driven Electric Energy Consumption Prediction for Smart Campuses. Environment and Ecology Research, 13(1), 25 - 37. DOI: 10.13189/eer.2025.130103.*

Abstract As the global energy crisis and environmental pollution intensify, energy conservation and emission reduction have become essential objectives for the development of smart campuses at universities and colleges worldwide. Effective management of energy consumption in these environments relies on accurate power consumption prediction models. This paper investigates the prediction of power consumption for smart campuses through a meteorological data-driven model. By analyzing key meteorological factors, such as temperature, humidity, wind speed, and solar radiation, which influence electricity consumption on campus, we propose a prediction model using a genetic algorithm-backpropagation (GA-BP) neural network combined with sample entropy value and assignment methods. This model utilizes actual data on campus electricity consumption, augmented with multi-source meteorological data for training and validation purposes. Experimental results indicate that the model achieves high prediction accuracy across various meteorological data-driven conditions, thus enhancing the optimization of energy consumption management on campuses. The findings of this research offer theoretical support for energy-saving strategies in smart campuses and serve as a valuable reference for predicting energy consumption in similar public buildings.

Keywords Smart Campus, BP Neural Network, Power Consumption Prediction, Meteorological Data, Energy Saving and Emission Reduction

1. Introduction

1.1. Increasing Global Climate Change and Environmental Problems

The challenges posed by global climate change and environmental degradation are escalating. In response, governments and organizations are implementing measures to mitigate these issues. A key global objective has emerged: reducing energy consumption and carbon emissions while advancing a low-carbon economy. In this context, the education sector—especially higher education institutions—plays a crucial role due to its significant impact on carbon emissions and energy usage. Therefore, initiating energy and carbon reduction initiatives on university campuses addresses global environmental policies and fosters social responsibility and sustainable development.

1.2. Environmental Protection and Sustainable Development

The world currently faces numerous environmental challenges, including climate change, energy shortages, and ecological degradation. As key institutions in developing future talent, educational establishments have a responsibility to promote environmental stewardship. Creating smart campuses not only reduces energy use and carbon emissions but also raises students' awareness of environmental issues. Given the growing global emphasis on sustainability, campuses—serving as centers for education and research—must actively demonstrate green and sustainable practices.

Educational institutions can effectively assess and manage resources related to energy consumption, photovoltaic power generation, and other factors. This approach helps minimize their ecological footprint and encourages a sustainable lifestyle. While the advancement of smart educational frameworks is essential, the development of smart campuses should also prioritize environmental protection and responsible resource utilization. By leveraging data to monitor and optimize energy use, schools can work towards their goals of resource conservation and environmental protection, thereby laying a strong foundation for a sustainable future.

Data analysis allows institutions to evaluate the specific impacts of various activities on the environment, supporting the creation of strategies that mitigate negative effects. This method not only reduces energy consumption but also significantly decreases carbon footprints.

1.3. Enhance the Efficiency of Campus Management and Thereby Promote Energy Conservation

Predicting electricity energy consumption is essential for reducing energy use and carbon emissions, thereby enhancing the management efficiency of a smart campus. This approach aligns with the environmental protection policies set forth by national and local governments and simultaneously boosts the university's social image and overall competitiveness. Moreover, it offers valuable insights for other educational institutions. By promoting energy savings, optimizing resources, and implementing green technologies, a smart campus not only mitigates environmental impact but also fosters environmental awareness and practical experience among students.

In summary, promoting the development of green smart campuses is essential in the context of global climate change and environmental pollution. Global warming has led countries to focus on controlling carbon emissions. Green smart campuses, as a significant direction for educational development and energy management transformation, can reduce carbon emissions, promote the concept of green development, and foster students' environmental awareness. These campuses utilize information technology to create an intelligent environment that optimizes resource management and enhances educational quality. Additionally, encouraging the low-carbon transformation of smart campuses not only aligns with global environmental protection policies but also strengthens the campus's social responsibility and capacity for sustainable development. This initiative holds considerable significance for the future.

2. Literature Review

The concept of a smart green campus aims to minimize environmental impacts and operational costs through the integration of green energy solutions and smart technologies [1]. Menezes et al. [2] demonstrated that photovoltaic solar power systems are economically viable for campus environments and can promote the advancement of sustainable technologies. Fonseca et al. [3] explored the retrofitting of university campus buildings to achieve near-zero energy targets while ensuring high levels of on-site renewable energy generation. Meteorological factors significantly influence electrical energy consumption patterns, making them essential variables in energy forecasting models. Understanding how these conditions affect energy use allows campus facilities to adjust their energy management strategies in real time, thereby improving operational efficiency and reducing costs [4]. Additionally, neural network-based machine learning approaches have proven effective in predicting short-term energy consumption trends [5]. By harnessing this technology, campuses can further enhance their energy management initiatives. Through the use of machine learning algorithms, researchers can enhance energy consumption prediction models, thereby increasing their accuracy. This advancement enables smart campuses to make informed decisions regarding energy usage and optimization [6]. It provides valuable insights into power management practices and informs the decision-making process [7].

As a result, this approach facilitates faster data analysis and decision-making, leading to optimized electrical energy consumption within campus facilities [8]. Machine learning models have demonstrated significant potential in accurately predicting electrical energy consumption. They can achieve high levels of accuracy in identifying energy usage patterns and optimizing management strategies within smart campuses [9]. This integrated approach not only predicts electrical energy consumption with greater precision but also outperforms traditional prediction methods [10]. Consequently, this results in more reliable predictions and the development of optimized electrical energy management strategies [11]. These strategies can adapt to dynamic energy demands and enhance resource allocation, ultimately improving overall electrical energy efficiency.

Khoshrou and Pauwels [12] proposed using multivariate linear regression and machine learning algorithms for short-term load forecasting. Their approach incorporates meteorological data—such as temperature and wind speed—as well as historical electricity consumption data. This method is particularly effective in campus environments, where precise meteorological data can significantly enhance the accuracy of energy forecasting and aid in developing more efficient energy management strategies. Zhao and Magoulès [13] noted that machine learning energy prediction models, including regression models and neural networks, can process extensive meteorological and real-time energy consumption data. This capability leads to more accurate prediction outcomes, thereby optimizing energy scheduling and management. Mohandes et al. [14] argued that, in smart campus scenarios, artificial neural networks can efficiently handle complex meteorological data and deliver precise predictions of electrical energy consumption. This accuracy is crucial for managing electrical energy in real time and for reducing carbon emissions. Foucquier et al. [15] emphasized that a data-driven modeling approach, when combined with meteorological data, can substantially enhance the accuracy of building energy efficiency predictions. For smart campuses, this method effectively forecasts electricity consumption and supports efforts to optimize energy usage. Amin and Mourshed [16] identified real-time local weather data as essential for optimizing energy management and accurately predicting energy and environmental behaviors. Additionally, Wu et al. [17] developed a tracking control strategy that fuses multiple sources of meteorological data under varying weather conditions, significantly improving the power generation efficiency of photovoltaic systems.

The construction of green smart campuses aligns with national environmental policies while enhancing social responsibility and promoting sustainable development. These campuses aim to improve energy efficiency, reduce carbon emissions, promote the use of green technologies, and enhance environmental quality. Intelligent management systems, such as smart lighting, energy efficiency monitoring, and green building technologies, can significantly lower energy consumption and carbon emissions. Additionally, environmental monitoring systems track real-time parameters, including air quality and temperature, to ensure a safe and comfortable campus environment. Despite the notable advantages of green smart campuses, the high initial investment required for implementation can impede the transition to low-carbon operations in educational institutions. This paper aims to provide theoretical support for energy and carbon reduction initiatives through meteorological data-driven research on predicting electrical energy consumption in smart campuses.

3. Research Methodology

3.1. Problem Description

This study aims to investigate how meteorological data affects energy consumption in smart campuses. By examining existing cases and practical experiences, we will use data analysis to forecast trends in campus energy consumption. Additionally, we will propose optimization strategies for energy management on campuses to enhance environmental protection and resource utilization. The study will focus on developing and optimizing the energy management strategy for smart campuses, covering aspects such as technology selection, implementation steps, and investment budgeting. We will conduct a detailed investigation into the current state of energy consumption at GXVT College, identify the main challenges in its transition to a smart campus, and analyze the effectiveness of existing transformation efforts. Furthermore, we will evaluate successful experiences and provide recommendations for optimization. This research will also offer practical insights and guidance for similar smart campus projects.

3.2. Model Assumptions

Hypothesis 1: Multiple linear regression is employed to simultaneously evaluate the impact of various factors on energy consumption. This method allows for the prediction of energy usage based on variables such as temperature, time of day, weather conditions, and other relevant elements.

Hypothesis 2: To address complex nonlinear relationships, regression predictions can be generated using a backpropagation (BP) neural network model. This approach effectively predicts nonlinear changes in energy consumption over time.

Hypothesis 3: The combination of Genetic Algorithm-Backpropagation (GA-BP) models with other neural network approaches can leverage their collective advantages to enhance prediction performance. These integrated models can be effectively utilized for energy consumption prediction and analysis.

The model described above can assist in researching energy consumption prediction on smart campuses. It analyzes the intricate patterns in energy consumption data, facilitates accurate predictions, and optimizes strategies for reducing energy use and carbon emissions.

3.2.1. Multiple Linear Regression Models

The general form of the multiple linear regression model is:

$$\hat{y} = b_0 + b_1x_1 + b_2x_2 + \cdots + b_nx_n \quad (1)$$

In the equation, $\hat{y}$ represents the estimated value of the dependent variable y ; $b_0, b_1, b_2, \dots, b_n$ denote the parameters, i.e., the regression coefficient.

3.2.2. BP Neural Networks

The backpropagation (BP) algorithm is a widely used method for training artificial neural networks. Initially proposed in 1986 by a group of American researchers led by Rumelhart and Williams, this technique combines the concept of backward error propagation with a multilayer feedforward neural network, resulting in what is known as a BP neural network. Often referred to as gradient descent, this optimization method adjusts the weights during the training phase to minimize the loss function. The BP algorithm consists of two main processes: forward propagation and backpropagation. In the forward propagation phase, the input signal moves through the neural network until it produces an output value, allowing for the calculation of the loss function. In the subsequent backpropagation stage, the gradient of the loss function is computed for each weight in the network. These gradients are then used to update the weights, aiming to minimize the overall loss. To fully understand the BP algorithm, it is crucial to first comprehend its underlying network structure. The representation of the BP neural network is shown in Figure 1.

The system consists of an input layer, a hidden layer, and an output layer, alongside the desired outputs, connection weights for each layer, activation functions, thresholds for each neuron, and a global error metric. The input vector $X_1, X_2, \dots, X_n$ represents the original energy consumption data for each building. The output vector $Y_1, Y_2, \dots, Y_m$, denotes the pblueicted energy consumption values for those buildings. The hidden layer input data is represented by $\mathrm{h}f_1, \mathrm{h}f_2, \dots, \mathrm{h}f_l$, while the hidden layer output data is denoted by $ho_1, ho_2, \dots, ho_l$. The desiblue output vector is indicated by $d_1, d_2, \dots, d_m$. The connection weights from the input layer to each neuron in the hidden layer are labeled as w_{ih} denotes the connection weights from the input layer to each neuron in the hidden layer; how denotes the connection weights from the hidden layer to each neuron in the output layer. The backpropagation (BP) neural network primarily consists of two phases. In the first phase, the input signal is forwarded through the network. The number of nodes in the input layer n, the number of nodes in the hidden layer l, and the number of nodes in the output layer m are determined based on the input-output sequences of the sample data (X,Y), with the activation function being a sigmoid function. Finally, k sets of data are selected in the dataset as inputs and corresponding desired outputs:

$$x(k) = \left(\mathrm{x}_1(k), \mathrm{x}_2(k), \dots, \mathrm{x}_n(k)\right) \tag{2}$$

$$d_0(k) = \left(d_1(k), d_2(k), \dots, d_n(k)\right) \tag{3}$$

The input to the i-th node of each neuron in the hidden layer of the neural network:

$$hf_i(k) = \sum_{i=1}^{n} w_{ih} x_i(k) - a_i \quad {\scriptstyle i=1,2,\dots,l} \tag{4}$$

The output from the i-th node in the hidden layer of a neural network:

$$ho_i(k) = f\left(hf_i(k)\right) \quad i = 1,2,\dots,l \tag{5}$$

The input to the j-th node of each neuron in the output layer of the neural network is:

$$yf_j(k) = \sum_{j=1}^{l} w_{ho} ho_i(k) - b_j \quad {\scriptstyle i=1,2,\dots,m} \tag{6}$$

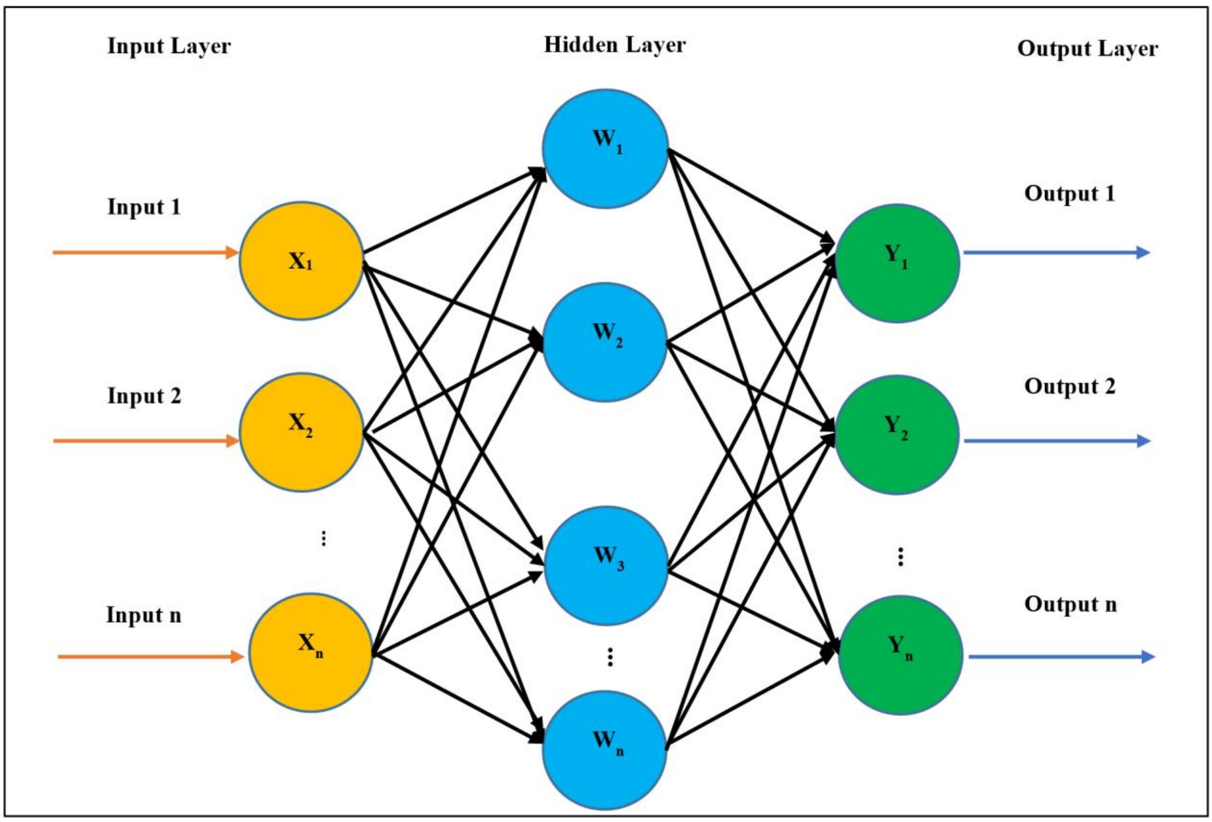

Figure 1. BP Neural Network Diagram

The output of each neuron in the output layer of the neural network can be expressed for each node of the j-th neuron as follows:

$$yo_j(k) = f\left(yf_j(k)\right) \quad (7)$$

Error function:

$$E = \frac{1}{2}\sum_{o=1}^{m}\left(d_o(k) - y_o(k)\right)^2 \quad (8)$$

The second stage is the back propagation of the error: the partial derivatives of the error function are calculated for each neuron in the output layer by using the desired output value and the actual output value $\delta_o(k)$:

$$\frac{\partial \mathrm{E}}{\partial w_{ho}} = \frac{\partial E}{\partial yf_j}\frac{\partial yf_j}{\partial w_{ho}} \quad (9)$$

$$\frac{\partial yf_j(k)}{\partial w_{ho}} = \frac{\partial(\sum_h^1 w_{ho}ho_i(k) - b_j)}{\partial w_{ho}} \quad (10)$$

$$\frac{\partial \mathrm{E}}{\partial yf_j} = \frac{\partial(\frac{1}{2}\sum_{o=1}^{m}(d_o(k) - yo_j(k)))^2}{\partial w_{ho}} = -\left(d_o(k) - yo_j(k)\right)\hat{f}\left(yf_j(k) - \delta_o(k)\right) \quad (11)$$

$$\frac{\partial \mathrm{E}}{\partial hf_i(k)} = -(\textstyle\sum_{o=1}^{m}\delta_o(k)w_{ho})\hat{f}(hf_i(k)) - \delta_i \quad (k) \quad (12)$$

The partial derivatives of each neuron in the output layer, along with the output values of each neuron in the hidden layer, are utilized to adjust the connection weights. where η represents the learning step size:

$$\triangle\, w_{ho}(k) = -\eta\frac{\partial E}{\partial W_{HO}} = \eta\delta_0(k)ho_i(k) \quad (13)$$

$$w_{ho}(k+1) = w_{ho}(k) + \eta\delta_0(k)ho_i(k) \quad (14)$$

The partial derivatives of each neuron in the hidden layer and the inputs of each neuron in the input layer are used to correct the connection weights:

$$\triangle\, w_{in}(k) = -\eta\frac{\partial e}{\partial hf_i(k)}\ \frac{\partial hf_i(k)}{\partial w_{ih}} = \delta_i(k)x_i(k) \quad (15)$$

$$w_{ih}(k+1) = w_{ih}(k) + \eta\delta_i(k)x_i(k) \quad (16)$$

The resulting network error is assessed to see if it meets the specified criteria. The training algorithm will terminate under two conditions: first, if the error meets the anticipated requirements, and second, if the number of iterations exceeds the predefined maximum limit. If the criteria are not met and the maximum iteration threshold has not been reached, the algorithm will proceed to the next learning cycle.

3.2.3. Combined GA-BP Modeling

The BP neural network offers several capabilities, including approximation, parallelism, fault tolerance, and self-learning, which makes it widely applicable in model classification and prediction. However, it also has significant limitations, such as issues with multiple feedback loops, slow convergence rates, and a propensity to become trapped in local minima, which can reduce the accuracy of prediction models. In contrast, the Genetic Algorithm (GA) encodes operational parameters without requiring prior knowledge of the system. By starting the search with a diverse set of strings, GA explores a vast search space, minimizing the risk of converging to a local optimum and enhancing global optimization.

In this study, we combine the strengths of both the BP neural network and the GA algorithm. Specifically, we use the GA to optimize the weights and thresholds of the BP neural network model, resulting in the development of the GA-BP optimization network model, as illustrated in Figure 2.

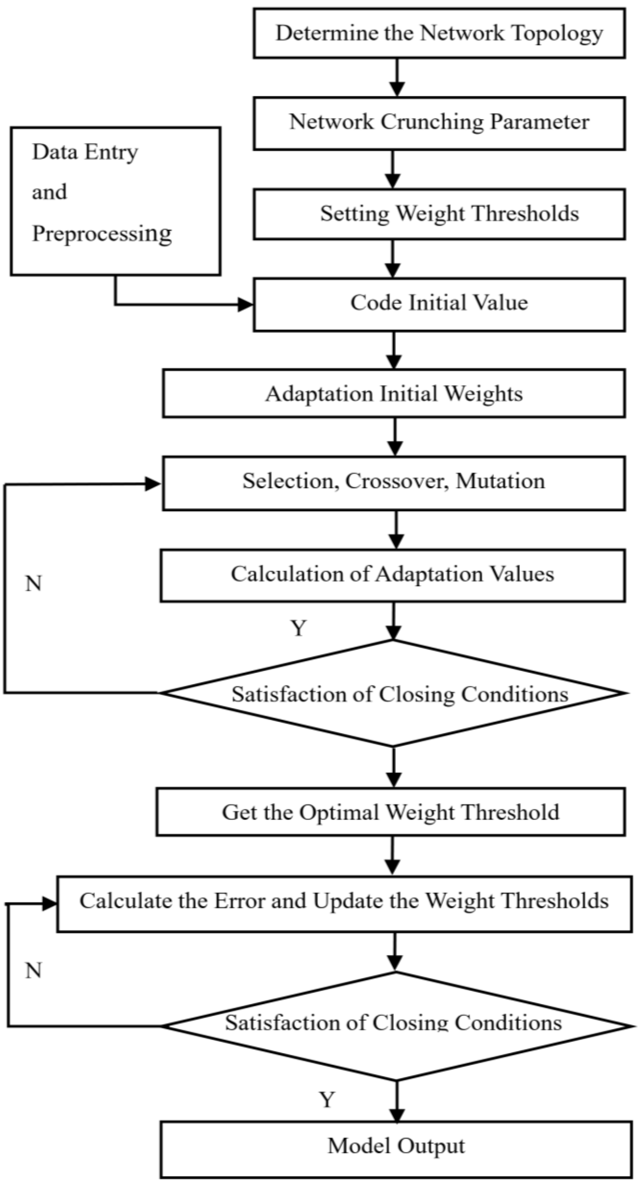

Figure 2. Block diagram of GA-BP modeling process

This study employs real number encoding to concatenate the weights and thresholds of the backpropagation (BP) neural network into a sequential vector, denoted as $\zeta = [V_1, W_1, B_1, B_2]$. In this context, V_1 signifies the connection weights between the hidden and input layers, W_1 indicates the connection weights between the output and hidden layers, B_1 represents the threshold for the hidden layer neurons, and B_2 denotes the threshold for the output layer neurons. The transfer function of the hidden layer is defined as:

$$f(x) = \frac{1}{(1+e^{-x})} \tag{17}$$

Let the input sample data M be denoted as $X_P = (X_{P,1}, X_{P,2}, X_{P,3}, \dots, X_{P,n})$, $p = 1,2,3,\dots,M$, the corresponding output for the samples is $Y_P = (Y_{P,1}, Y_{P,2}, Y_{P,3}, \dots, Y_{P,n})$, $p = 1,2,3,\dots,M$, V_P is the implicit layer to input layer neuron connection weights for the Pth sample; W_P is the output layer to implied layer neuron connection weights for the Pth sample; B_p^1 represents the threshold of the hidden layer for the P-th sample, while B_p^2 denotes the threshold of the output layer for the same sample. The output value M_p obtained after BP learning is:

$$M_p = W_P \frac{1}{1+e^{-[\sum_{i=1}^{R} V_P \times i + B_p^1]}} + B_p^2 \tag{18}$$

The adaptation function is chosen as:

$$F = \frac{1}{\sum_{p=i}^{M}(Y_p - M_P)^2} = \frac{1}{\sum_{p=i}^{M}(Y_p - M_P \frac{1}{1+e^{e^{-[\sum_{i=1}^{R} V_P \times i + B_p^1]}}} + B_P^2)^2} \tag{19}$$

The selection operator employs the roulette selection method. Let the probability of selecting an individual be denoted as P_i. Since a higher fitness value indicates a greater error and, therefore, a lower likelihood of selection, we use the inverse of the fitness value. Consequently, we can express this relationship as follows:

$$f_i = \frac{k}{F_i} \tag{20}$$

$$p_i = \frac{f_i}{\sum_{i=1}^{n} f_i} \tag{21}$$

where F_i is the individual i fitness value; k is a constant; n is the population size.

The crossover operation at position k between the *i*-th chromosome X_i and the *m*-th chromosome X_m is described as follows:

$$X_{ij} = X_{ij}(1-b) - X_{ij}b \tag{22}$$

$$X_{mj} = X_{mj}(1-b) - X_{mj}b \tag{23}$$

where b is a random number. $b \in [0,1]$.

Subsequently, apply mutation operations to select the, the *j*-th individual with the first *i*-th gene y_{ij} individual for modification:

$$y_{ij} = \begin{cases} y_{ij} + (y_{ij} - y_{max})\left(1 - \frac{g}{G_{max}}\right) r_2, r \geq 0.5 \\ y_{ij} + (y_{min} - y_{ij})\left(1 - \frac{g}{G_{max}}\right) r_2, r < 0.5 \end{cases} \tag{24}$$

y_{max} represents the upper limit of gene y_{ij}, while y_{min} denotes the lower limit of the same gene. Additionally, r_2 is defined as a random number; g refers to the number of iterations, G_{max} indicates the maximum number of evolutionary cycles, and r is constrained within the interval [0, 1].

$$minE_p < \epsilon \tag{25}$$

Where ε is the total allowable error. Evaluate the results, if the convergence criterion is satisfied, the set of solutions corresponding to the applicability degreeF_p, max. If the convergence criterion is satisfied, the set of solutions corresponding to the applicability degree is the solution of the problem, and the computation is finished; otherwise, we go to step 2 for selection, crossover, and mutation to generate the next generation of populations. The quantity of hidden layer nodes l in the architecture of a Backpropagation BP neural network can be calculated using the following formula:

$$l = \sqrt{n+m} + a \tag{26}$$

In Equation (26), n denotes the number of nodes in the input layer, m signifies the number of nodes in the output layer, and a is a constant defined within the range [1,10]. The genetic algorithm GA significantly enhances the performance of the BP neural network by facilitating rapid convergence and preventing local minima, as illustrated in Figure 2.

3.3. Parameter Setting

3.3.1. Parameter Variables

The study of energy consumption prediction in smart campuses, driven by meteorological data, encompasses essential principles and definitions for selecting research variables. (1) Energy Consumption-Related Variables: This includes data on the usage of various energy resources, particularly electricity, within the campus. Analyzing these variables helps to explore ways to use energy and resources more efficiently. (2) Variables related to campus environmental monitoring: Environmental indicators, including temperature and humidity, are utilized as significant variables. Employ predictive models to project future trends in electricity consumption, facilitating resource planning and optimization based on these projections (refer to Parameter Variable Definition Table 1).

Table 1. Definition of variables

variant	notation	Definitions and explanatory notes
Campus electrical energy consumption	Et	Energy consumption of electricity used by external power supply offices
Photovoltaic power	Eg	Generated photovoltaic energy power
Weather category	TW	Weather category: sunny, cloudy, cloudy, rainy.
Temp	T	Maximum, minimum and average temperatures, etc.
Air velocity	WS	Wind speed rating
Atmospheric precipitation	AP	Liquid water, etc., that falls to the ground from clouds in the sky is collectively called precipitation.
Solar radiation index (SDI)	DHI	The solar radiation index is a quantitative measure of the intensity of solar radiation.
Direct solar radiation index (DSRI)	DNI	The direct irradiance of the sun is defined as the amount of solar radiation energy received per unit area per second.
Total horizontal solar radiation	GHI	Total solar irradiation on a horizontal surface is the total amount of solar radiation received by a horizontally placed surface per unit area.
Fixed inclination radiation	GTIF	PV fixed tilt radiation refers to the angle between the surface of the PV module and the surface water plane.
Track tilt radiation	GTIT	Photovoltaic tracking of tilted radiation refers to the process of maximizing the reception of solar radiation by adjusting the tilt angle of the PV module.
Azimuth	A	The azimuth of a PV plant is the angle between the projection of the normal to the plane of the PV module on the horizontal plane and the direction due south.
Relative humidity	RH	Relative humidity is defined as the ratio of the partial pressure of water vapor in the air to the saturated water vapor pressure at a given temperature, expressed as a percentage.
Ground pressure	P	Ground pressure generally refers to the atmospheric pressure at a certain location on the ground.
Zenith	ZA	The solar zenith angle is the angle between the direction of incidence of light and the direction of the zenith.

3.3.2. Model Evaluation

Error measure selection, this paper adopts three evaluation indexes, namely, mean absolute percentage error (MAPE), root mean square error (RMSE) and mean absolute error (MAE), to evaluate the proximity between the predicted value of energy consumption and the real value. The smaller values of MAPE, RMSE and MAE in campus energy consumption prediction represent the more accurate energy consumption prediction. The evaluation indexes are shown in Eq. (27)-Eq. (29) below:

$$E_{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y(i)-y(\acute{\iota})}{y(i)}\right| \tag{27}$$

$$E_{RMSE} = \frac{1}{n}\sqrt{\sum_{i=1}^{n}(y(i)-y(\acute{\iota}))^2} \tag{28}$$

$$E_{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y(i)-y(\acute{\iota})| \tag{29}$$

In the formula:$y(\acute{\iota})$, and $y(i)$ are the predicted energy consumption value and the real energy consumption value at the i the predicted energy consumption value and the real energy consumption value at the moment of time; n represents the total number of test samples.

3.4. Data Processing

3.4.1. Sample Data

The GXVT campus is situated in the Nanning area of Guangxi, characterized by hot summers and warm winters. The campus primarily relies on electric energy for its energy needs. This paper will analyze electric energy consumption data from the GXVT campus over the past three years (2021-2023), along with relevant meteorological data (Figure 3: Weather Type Statistics).

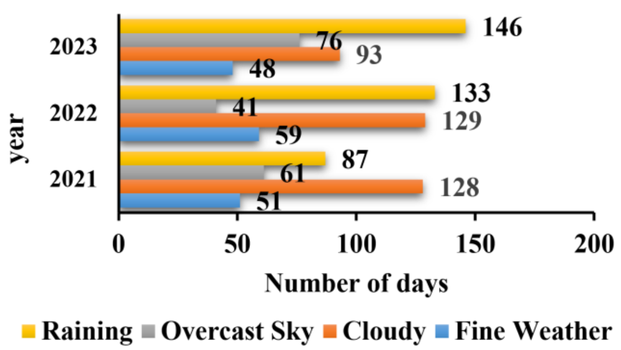

Figure 3. Statistical map of weather types

3.4.2. Data Processing

3.4.2.1. Sample Entropy Method

To quantitatively analyze the relationship between weather type and campus energy consumption, we employ the sample entropy method to quantify the weather type. This method helps describe the complexity and regularity of the system. The steps of the algorithm are as follows:

(1) Transform the original time series of length N into am dimensional vector$x_m(i)$ with the formula.

$$x_m(i) = [x(i), x(i+1), \dots, x(i+m-1)] \quad i = 1,2,3,\dots, N-m+1 \tag{30}$$

(2) Calculate the distance between each pair of vectors and identify the maximum value.

$$d_m[x_m(i), x_m(j)] = \max[x_m(i+k) - x_m(j+k)], 0 \le k \le m-1 \tag{31}$$

(3) Determine the tolerance threshold r, and define the number of distances less than r to the $N-m$. The ratio of the number of distances less than r is defined as $B_i^m(r)$:

$$B_i^m(r) = \frac{1}{N-m+1} num\{d[x_m(i), x_m(j)] < r\},$$
$$i = 1 = 1,2,3,\dots, N-m+1; i \neq j \tag{32}$$

(4) Calculate the average value of $B_i^m(r)$ between i and $N-m+1$.

$$B^m(r) = \frac{1}{N-m+1} \sum_{i=1}^{N-m+1} B_i^m(i) \tag{33}$$

(5) Increment m by 1 (m = m + 1), then repeat steps 1 through 4 to perform the calculations.

$$B^{m+1}(r) = \frac{1}{N-m} \sum_{i-1}^{N-m} B_i^{m+1}(r) \tag{34}$$

(6) Such that the sample is first defined as.

$$S_n(m,r) = \lim_{N\to\infty} \{-\ln[B^{m+1}(r) / B^m(r)]\} \tag{35}$$

Since N will not be infinitely large in the real situation, the sample full-fledged of Eq. (43) can again be written as.

$$\mathrm{SampEn}(m, r, N) = -\ln[B^{m+1} / B^m(r)] \tag{36}$$

Typically, r is defined as $r = (0.1 \text{ to } 0.25)E_{std}$ (E_{std} is the standard deviation of the original sequence). A value that is too large may lead to a loss of information, while a value that is too small increases the likelihood of errors. The reconstruction dimension, m, should be set to either 1 or 2. Setting m greater than 2 necessitates a substantial amount of data, whereas m equal to 1 yields insufficient information. When selecting the data length, N, empirical findings suggest that a value between 100 and 10,000 is optimal.

3.4.2.2. Quantification of Weather Types

(1) Sample entropy quantification of weather type method T1

Unlike meteorological factors, weather type serves as a descriptive parameter. Therefore, this study focuses on energy consumption data from four specific weather types: sunny, cloudy, overcast, and rainy days. To comprehensively examine the factors influencing energy consumption, we employ the sample entropy method for quantitative analysis of these weather types. The sample entropy parameters used in this study are set as follows: r=0.2Estd, m=2, N=30. For instance, the quantification of weather types in May 2021 is presented in Table 2(Sunny S, cloudy C, overcast Wo, rainy R).

Table 2. Entropy value of weather type samples in May 2021

dates	Type of weather	Sample entropy value	dates	Type of weather	Sample entropy value	dates	Type of weather	Sample entropy value
1	S	0.423	11	C	0.354	21	S	0.447
2	S	0.385	12	WO	0.369	22	S	0.412
3	C	0.367	13	S	0.382	23	R	0.385
4	R	0.344	14	S	0.437	24	R	0.385
5	C	0.359	15	C	0.393	25	S	0.433
6	WO	0.346	16	C	0.357	26	C	0.412
7	S	0.394	17	WO	0.391	27	C	0.406
8	S	0.379	18	S	0.368	28	WO	0.376
9	C	0.361	19	S	0.378	29	C	0.384
10	R	0.346	20	WO	0.396	30	S	0.397

(2) Sample entropy quantification of weather type method T2

In this paper, the authors draw on the experience of power system dispatchers to utilize values between [0, 1] to qualitatively represent the impact of climate sensitivities on forecasted loads. As shown in Table 3 (Sunny S, Cloudy C, Overcast Wo, Light Rain LR, Moderate Rain MR, Heavy Rain HR):

Table 3. Table of weather eigenvalue values

weather conditions	S	C	CO	LR	MR	HR
S	1	0.9	0.85	0.75	0.7	0.65
C	0.9	0.8	0.75	0.65	0.6	0.55
CO	0.85	0.75	0.7	0.6	0.55	0.5
LR	0.75	0.65	0.6	0.5	0.45	0.4
MR	0.7	0.6	0.55	0.45	0.4	0.35
HR	0.65	0.55	0.5	0.4	0.35	0.3

Sample entropy quantifies the weather type as a descriptive parameter, distinct from traditional meteorological factors. In this study, four weather types—sunny, cloudy, overcast, and rainy days—are selected for analysis. The sample entropy method is employed to quantify and analyze these weather types, facilitating a more comprehensive examination of the factors influencing energy consumption (see Table 3). The authors draw on insights from power system dispatchers to use values within the range of [0, 1] to qualitatively represent the impact of weather sensitivities on forecasted loads (refer to Table 3). These two methods for quantifying weather data serve as input for the BP model, enabling a more accurate comparative analysis of the results.

4. Results and Analyses

4.1. Regression Model Analysis

Through regression analysis of meteorological data, campus energy consumption, and photovoltaic (PV) power generation, it is demonstrated that meteorological factors significantly influence energy consumption. Specifically, PV power generation exhibits a positive correlation with weather temperature, relative humidity, ground total irradiance (GTI) for both fixed and tracking tilt configurations, atmospheric precipitable water, surface air pressure, wind speed, zenith angle, and direct normal irradiance (DNI) index. However, it shows a negative correlation with the diffuse horizontal irradiance (DHI) index (As shown in tables 4 and 5).

This paper hypothesizes that multiple linear regression models are inadequate for accurately analyzing the nonlinear characteristics of meteorological data (e.g., temperature, humidity, solar radiation), which in turn impacts the prediction accuracy of meteorological data, campus energy consumption, and photovoltaic power generation. This analysis lays the groundwork for future research employing the BP combination model.

Table 4. Analysis of Campus Energy Consumption and the Impact of Weather Factors

variant	B	standard error	standardized factor	t	Sig.
(Constant)	4730.244	1040.460		9.023	.000
TW	414.663	152.174	.184	2.295	.023
T	79.493	14.573	0.237	5.455	0.000
AP	-31.894	2.071	-0.22	-15.402	.019
RH	-.105	.011	-.047	-9.270	.000
P	-5.239	2.471	-.172	-2.120	.036
WS	-21.089	75.064	-.020	-.281	.000

Dependent variable: campus energy consumption

Table 5. Analysis of photovoltaic power generation and the impact of weather factors

variant	B	standard error	standardized factor	t	Sig.
(Constant)	2.905	.545		5.327	.000
TW	14.472	10.174	.804	0.539	.000
T	.127	.002	.047	51.449	0.000
AP	-.084	.004	-.016	-24.036	.000
RH	.054	.001	.024	38.613	0.000
P	-.002	.001	-.001	-3.219	.001
WS	-.066	.006	-.004	-10.786	.000
ZA	-.025	.001	-.025	-33.543	.000
GTIF	.100	.000	.934	560.007	0.000
GTIT	.091	.000	.854	429.977	0.000
DNI	-.082	.000	-.702	-295.981	0.000
DHI	-.065	.000	-.227	-175.546	0.000

Dependent Variable: Photovoltaic Power

4.2. Comparative Analysis of BP Models

4.2.1. Analysis of BP Neural Network Structure and Parameter Setting

Backpropagation (BP) neural networks are often susceptible to overfitting when addressing complex problems. Overfitting occurs when the model demonstrates high performance on the training set but fails to generalize effectively to the test set or new data, leading to poor model generalization. To mitigate the risk of overfitting, it is essential to determine the optimal number of neurons in the hidden layer based on prior sample data, reduce the number of network layers or neurons per layer, and opt for a simpler network architecture. In the context of modeling campus electric energy consumption (Et) and campus photovoltaic power (Eg), the number of hidden layer neurons in the BP neural network is established using a trial calculation method. This trial calculation method involves two steps.

(1) First, the empirical formula (26) is employed to ascertain the optimal range for the number of hidden layer neurons, denoted as l. In this context, n represents the number of neurons in the input layer, while m indicates the number of neurons in the output layer. Additionally, the variable a is a constant, with a value range between 1 and 10.

(2) Replace the number of neurons in the hidden layer within the specified range in the established network for further calculations, and identify the optimal number of hidden layer neurons based on the fitting test results.

In the modeling process of campus electric energy consumption (Et), with parameters n=6 and m=1, the range for the number of neurons in the hidden layer is determined to be between 3 and 13, as indicated in Step 1. When the number of neurons in the hidden layer varies from 3 to 13, the training data and the root mean square error (RMSE) are utilized; a smaller RMSE signifies a better fit for the network model. Table 6 illustrates the variation in RMSE between the network output and the desired values as the number of neuron nodes in the hidden layer changes. It is evident from the table that as the number of neurons increases, the RMSE reaches its minimum at 9 neurons. Therefore, selecting 9 neurons in the hidden layer allows the network to achieve an optimal fit. Consequently, the number of neurons in the hidden layer of the BP neural network for the campus electrical energy consumption (Et) measurement model is chosen to be 9. In a similar manner, for the campus photovoltaic (PV) model, the number of neurons is also set to 9. Furthermore, in the context of campus photovoltaic power (Eg), where n=11 and m=1, Step 1 indicates that the range for the number of neurons in the hidden layer is between 4 and 14. Thus, selecting 9 neurons in the hidden layer again results in a better fitting effect, leading to the conclusion that the number of neurons in the hidden layer of the BP neural network for the campus photovoltaic power measurement model should also be set to 9.

4.2.2. Combining Genetic Algorithms to Optimize BP Neural Networks to Reduce the Risk of Overfitting

This paper presents a genetic algorithm (GA) to optimize the backpropagation (BP) neural network. Initially, the weights of the BP network are optimized using GA, followed by the application of the BP algorithm to accurately search for the optimal solution of these weights. The GA-BP neural network model enhances the structure and parameter selection of the BP neural network through the genetic algorithm, thereby improving the model's performance and robustness. The number of neurons in the hidden layer is determined using a trial-and-error method, resulting in a GA-BP neural network that exhibits greater stability compared to traditional BP neural networks. By leveraging the strengths of both genetic algorithms and BP neural networks, the GA-BP neural network effectively addresses nonlinear data and demonstrates strong performance across various fields. Furthermore, it significantly reduces the risk of overfitting, thus enhancing the model's performance on new data.

4.2.3. Comparative Analysis of Training Results

Multiple models for regression (BP) analysis were analyzed via Matlab and compared.

The dataset used for this study comprises historical monitoring data from January 2021 to December 2023, which includes meteorological variables such as weather type and temperature. In this research, the first 80% of the dataset is designated as the training set, while the remaining 20% serves as the test set. The T1-GA-BP model proposed in this paper is compatible and is analyzed against both the GA-BP model and the BP model that utilizes unprocessed data. The result presented in Tables 7 and 8, indicate that the T1-GA-BP prediction model is effective for forecasting energy consumption across different weather conditions. Furthermore, its prediction accuracy surpasses that of the T2-GA-BP model, which does not utilize sample entropy processing. Previous analyses have shown that meteorological factors significantly influence campus energy consumption. This study demonstrates that an electric energy consumption prediction model that integrates meteorological data can significantly contribute to the development of a smart campus. The model not only aids in optimizing energy usage but also enables the real-time detection and response to anomalies in energy consumption. Consequently, it enhances the overall efficiency of electric energy management on smart campuses, facilitating a balanced approach between total campus energy consumption, external power supply usage (Et), and energy generated from photovoltaic sources (Eg).

In summary, by analyzing the meteorological data and electrical energy consumption of the GXVT campus, the campus can initiate several key strategies. These include generating electricity through rooftop photovoltaic systems, accelerating the installation of energy-efficient smart lighting, transitioning from traditional PCs to cloud desktop PCs for hands-on training, and increasing the number of electric vehicles alongside the corresponding charging infrastructure. Although these measures require significant capital investment, they are essential for developing a comprehensive grsmart campus. A systematic and planned approach is necessary for optimizing the construction process (Figure 4). The analysis of meteorological and energy data serves to enhance the feasibility of establishing a holistic energy consumption management platform for the campus. This platform will support distributed photovoltaic power generation, promote campus sustainability, and create an efficient pathway to optimize the development of the Smart Campus, ultimately fostering an improved educational environment.

Table 6. RMSE for varying numbers of neurons in the hidden layer (N)

N	3	4	5	6	7	8
RMSE	0.1855	0.1812	0.1828	0.1804	0.1806	0.1796
N	9	10	11	12	13	
RMSE	0.1776	0.1782	0.1785	0.1785	0.1787	

Table 7. Campus Energy Consumption - Prediction Errors of BP Model vs. BP Combined Model

Model	**MAE**	**MAPE**	**EMSE**
BP	5.1712	0.053748	6.4899
GA-BP	4.7836	0.048652	5.8346
T2-GA-BP	4.2673	0.042631	4.2638
T1-GA-BP	3.8914	0.038673	4.9563

Table 8. Photovoltaic Power Generation - Prediction Errors of BP Model Versus BP Combined Model

Model	**MAE**	**MAPE**	**EMSE**
BP	4.5195	0.052198	6.2546
GA-BP	3.9247	0.048235	5.4546
T2-GA-BP	3.0872	0.031324	4.0821
T1-GA-BP	3.6361	0.039372	4.8438

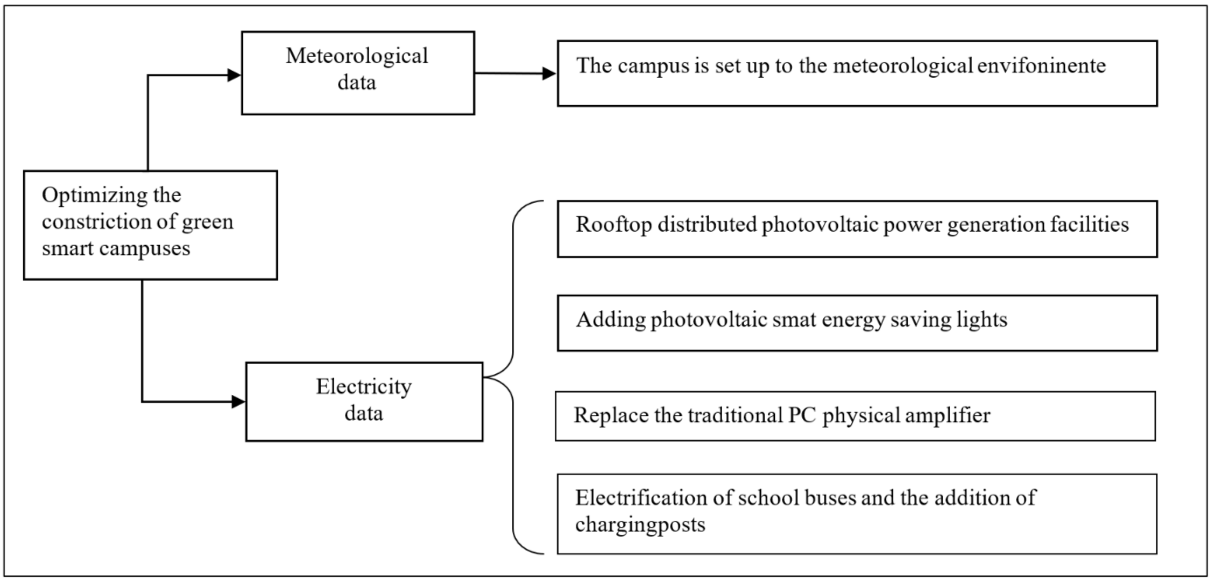

Figure 4. Optimising smart campus construction based on meteorological and electrical energy consumption data

5. Conclusions

Weather data can be utilized to determine the optimal time for automatically activating the lighting in the campus's public infrastructure. Similarly, it can help address the issue of managing equipment, air conditioning, and lighting in teaching facilities by enabling them to automatically shut off when unoccupied. This approach effectively reduces energy waste.

By utilizing regression analysis on gas and energy consumption data, researchers can develop effective retrofit strategies and accurately predict energy usage. This method enables a thorough assessment of energy demand on campus, supporting the implementation of targeted energy-saving measures. Predicting energy consumption allows institutions to optimize resource allocation, reduce waste, and improve overall energy efficiency. Additionally, accurate forecasts of energy use can help lower energy costs, which is particularly important for budget preparation and energy procurement.

Post-decarbonization, forecasting energy consumption becomes essential for evaluating the effectiveness of implemented measures and assessing their impact on carbon emissions reduction. Furthermore, proactive energy use forecasting can increase environmental awareness among students and faculty, fostering the establishment of a green campus. This approach also lays the groundwork for long-term planning and decision-making, ensuring that the campus energy system can adapt to future demands and changes.

In summary, predicting energy consumption is essential for the low-carbon transformation of smart campuses. It serves as a key factor in guiding and optimizing transformation strategies, while also enhancing the overall standards of energy management within the institution.

I recommend implementing several effective carbon reduction measures in addition to increasing photovoltaic power generation, expanding green planting areas, and upgrading energy-efficient office and teaching equipment. These measures include: (1) Improving energy efficiency through building renovations, such as enhancing insulation and installing high-efficiency windows and doors. (2) Implementing smart systems for controlling lighting and temperature. (3) Promoting sustainable transportation options and providing charging stations for electric vehicles. (4) Adopting water conservation practices, including the installation of efficient fixtures and the use of rainwater harvesting. (5) Utilizing low-carbon products and services. Additionally, it is crucial to encourage a paperless office environment, offer environmental education, implement a low-carbon campus policy, and promote waste separation and recycling. Collectively, these initiatives will reduce the carbon footprint of campuses and cultivate a culture of sustainability.

The development of a comprehensive management platform for a green intelligent campus facilitates real-time dynamic data management related to campus weather, electricity, and energy consumption. By leveraging big data and artificial intelligence, the platform enables effective data mining, as well as forecasting of weather and energy consumption. This initiative aims to continuously enhance the construction of the green intelligent campus, ultimately establishing a fully integrated, iconic model of sustainability and intelligence.

REFERENCES

[1] Abdulmouti, H., Skaf, Z., & Alblooshi, S. “Smart Green Campus: The Campus of Tomorrow.” 2022 Advances in Science and Engineering Technology International

Conferences (ASET), pp. 21-24, Feb. 2022.

[2] Menezes, J., Cury, J., & Souza, L. “Sustainable Practices Improving the University Campus: Feasibility of A Photovoltaic System.” Journal of Chemistry, Environmental Sciences and its Applications, vol. 7, no. 2, pp. 43-53, 2021. https://doi.org/10.15415/jce.2021.72006

[3] Fonseca, P., Moura, P., Jorge, H. and de Almeida, A., “Sustainability in university campus: options for achieving nearly zero energy goals.” International Journal of Sustainability in Higher Education, vol. 19, no. 4, pp. 790-816, 2018. https://doi.org/10.1108/IJSHE-09-2017-0145

[4] Tokareva, G., Sanzhapov, R., Savenkov, M., & Ilyin, D. “Prediction models for electricity consumption under influence of meteofactors.” Vestnik of Astrakhan State Technical University Series Management Computer Science and Informatics, pp. 99-106, 2018. https://doi.org/10.24143/2072-9502-2018-4-99-106

[5] Shaikh, A., Nazir, A., Khan, I., & Shah, A. “Short term energy consumption forecasting using neural basis expansion analysis for interpretable time series.” Scientific Reports, vol. 12, no. 1, 2022. https://doi.org/10.1038/s41598-022-26499-y

[6] Babanazarov, N. “Advancing energy efficiency: harnessing machine learning for smart grid management.” E3s Web of Conferences, vol. 524, Article Number 01003, 2024. https://doi.org/10.1051/e3sconf/202452401003

[7] Kifor, C., Olteanu, A., & Zerbes, M. “Key performance indicators for smart energy systems in sustainable universities.” Energies, vol. 16, no. 3, pp. 12-46, 2023. https://doi.org/10.3390/en16031246

[8] Fernández-Caramés, T. and Fraga-Lamas, P. “Towards next generation teaching, learning, and context-aware applications for higher education: a review on blockchain, iot, fog and edge computing enabled smart campuses and universities.” Applied Sciences, vol. 9, no. 21, pp. 44-79, 2019. https://doi.org/10.3390/app9214479

[9] Balaji, S. and Karthik, S. “Energy prediction in iot systems using machine learning models.” Computers Materials & Continua, vol. 75, no. 1, pp. 443-459, 2023. https://doi.org/10.32604/cmc.2023.035275

[10] Ghazal, T., Noreen, S., Said, R., Khan, M., Siddiqui, S., Abbas, S., … & Ahmad, M. “Energy demand forecasting using fused machine learning approaches.” Intelligent Automation & Soft Computing, vol. 31, no. 1, pp. 539-553, 2022. https://doi.org/10.32604/iasc.2022.019658

[11] Wang, S. “Predication of smart building energy consumption based on deep learning algorithm.” Journal of Autonomous Intelligence, vol. 6, no. 2, pp. 691, 2023. https://doi.org/10.32629/jai.v6i2.691

[12] Khoshrou, A., & Pauwels, E. J. “Short-term scenario-based probabilistic load forecasting: A data-driven approach.” Applied Energy, vol. 238, pp. 1258-1268, 2019. https://doi.org/https://doi.org/10.1016/j.apenergy.2019.01.155

[13] Zhao, H.-x., & Magoulès, F. “A review on the prediction of building energy consumption.” Renewable and Sustainable Energy Reviews, vol. 16, no. 6, pp. 3586-3592, 2012. https://doi.org/https://doi.org/10.1016/j.rser.2012.02.049

[14] Mohandes, S. R., Zhang, X., & Mahdiyar, A. “A comprehensive review on the application of artificial neural networks in building energy analysis.” Neurocomputing, vol. 340, pp. 55-75, 2019. https://doi.org/https://doi.org/10.1016/j.neucom.2019.02.040

[15] Foucquier, A., Robert, S., Suard, F., Stéphan, L., & Jay, A. “State of the art in building modelling and energy performances prediction: A review.”Renewable and Sustainable Energy Reviews, vol. 23, pp. 272-288, 2013. https://doi.org/https://doi.org/10.1016/j.rser.2013.03.004

[16] Amin, A., & Mourshed, M. “Weather and climate data for energy applications.” Renewable and Sustainable Energy Reviews, vol. 192, 2024. https://doi.org/10.1016/j.rser.2023.114247

[17] Wu, H., Chou, X.W., Yang, G.T., & Liu, C.A. “Optimisation of photovoltaic tracking system by integrating meteorological data.” Journal of Huazhong University of Science and Technology (Natural Science Edition), vol. 41, no. S1, pp. 204-207, 2013. https://doi.org/10.13245/j.hust.2013.s1.003