U.P.B. Sci. Bull., Series C, Vol. 88, Iss. 3, 2026 ISSN 2286 – 3540

RESEARCH ON 5G BASE STATION ENERGY EFFICIENCY BASELINE CONSTRUCTION TECHNIQUE BASED ON PRINCIPAL COMPONENT ANALYSIS AND ARTIFICIAL NEURAL NETWORKS IN THE PERSPECTIVE OF HIGH-DIMENSIONAL DATA ANALYSIS

Yao RAO[1], Yuan CHENG[2] *, Xinlei BAI[3], Sheng DING[4], Shen XIAO[5], Ziyi LIU[6]

Aiming at the high-dimensional characteristics of the data presented in the process of energy efficiency assessment of 5G base stations, this paper proposes a novel baseline construction method for energy efficiency of 5G base stations by organically combining the principal component analysis (PCA) with the self-encoder technology. A systematic data preprocessing process is designed in the research process, followed by linear dimensionality reduction by PCA technology, which projects the high-dimensional data into the low-dimensional space with the help of orthogonal transformations, thus effectively retaining the main variation information of the data. In order to enhance the effect of dimensionality reduction, the study introduces a self-encoder to realize nonlinear dimensionality reduction, and designs a bootstrap constrained self-encoder based on the PCA results. On this basis, this study utilizes the feature data after dimensionality reduction to construct an energy efficiency baseline model using BP neural network. The experimental results confirm that the method has significant advantages in terms of accuracy, efficiency and stability, effectively solves the key problem of high-dimensional data analysis in the energy efficiency assessment of 5G base stations, provides a solid theoretical basis and technical support for the optimization of the energy efficiency of 5G base stations, and has far-reaching significance in promoting the green development of the entire 5G network.

Keywords: Principal component analysis; self-encoder; BP neural network; 5G base station; energy efficiency baseline; high-dimensional data

[1] State Grid Electric Power Research Institute Wuhan Energy Efficiency Evaluation Co., Ltd. 430070, Wuhan, Hubei
[2]* State Grid Electric Power Research Institute Wuhan Energy Efficiency Evaluation Co., Ltd. 430070, Wuhan, Hubei, Corresponding author, E-mail: YuanChengYCG@outlook.com
[3] State Grid Hebei Electric Power Co., Ltd. Marketing Service Center, 050021 Shijiazhuang, Hebei
[4] State Grid Electric Power Research Institute Wuhan Energy Efficiency Evaluation Co., Ltd. 430070, Wuhan, Hubei
[5] State Grid Electric Power Research Institute Wuhan Energy Efficiency Evaluation Co., Ltd. 430070, Wuhan, Hubei
[6] State Grid Electric Power Research Institute Wuhan Energy Efficiency Evaluation Co., Ltd. 430070, Wuhan, Hubei

1. Introduction

As the latest communication network that has been put into use, the fifth-generation mobile communication technology (5G) has a brand-new network architecture with peak rates in Gbps, millisecond transmission latency and hundreds of billions of connectivity, which opens up a new era of three-way interconnection and interaction between humans, machines and things [1-3]. However, compared to 4G base stations, 5G base stations consume about three times as much power, and to achieve full network coverage, 5G requires about three times as many base stations as 4G base stations, resulting in huge energy consumption [4]. At the 2021 World 5G Conference, Xiao Yaqing, Minister of Industry and Information Technology, said that China has deployed and constructed 993,000 5G base stations [5-6]. This also means that the country will continue to promote the comprehensive construction of 5G networks, and the number of 5G base stations will increase.

Currently, the equipment that generates energy efficiency in 5G networks is mainly distributed in four major parts: transmission, power supply, room air conditioning, and base station, of which the energy consumption accounted for by the base station alone is more than 80% of the energy consumption of all 5G networks [7]. Data provided by the China Communications Standards Association shows that the no-load power consumption of the main equipment of 5G base stations of various operators is about 2.2-2.3 kW, and the full-load power consumption is about 3.7-3.9 kW [8-10]. Moreover, since 5G base stations are deployed in one-stop multi-frequency mode, the maximum power consumption generated by the base station in this mode is likely to reach 10-20 kW, and the power consumption of the base station will be increased if the multi-operator sharing scenario is considered [11-13]. At the same time, the number of base stations that need to be deployed will also increase by a multiple of 4G base stations because 5G base station RF signals will decay with propagation [14]. Therefore, the high-power consumption problem brought by 5G base stations has become an important factor limiting the construction of 5G networks.

From the above literature, the high-dimensional data generated by the 5G base station is the time series collected by each monitoring unit inside the base station, which is very suitable for the artificial neural network, but the artificial neural network loses its learning ability for the data with higher dimensionality. In order to improve the ability of artificial neural networks to handle high-dimensional data, high-dimensional data can be downscaled by principal component analysis (PCA) during the training process, thus enhancing their applicability in 5G base station energy efficiency management and optimization. Therefore, this study aims to explore the preprocessing methods applicable to high-dimensional energy efficiency data, investigate the application of linear dimensionality reduction techniques based on PCA, design a nonlinear dimensionality reduction method combined with a self-encoder, construct an energy efficiency baseline model based on BP neural networks, and validate the effectiveness of the method through experiments.

2. Research Methodology

2.1 Data pre-processing

5G base station energy efficiency data has significant characteristics such as high dimensionality, multiple sources and strong heterogeneity, and good data preprocessing is directly related to the accuracy of energy efficiency baseline construction. The 5G base station energy efficiency data used in this study mainly comes from three sources: the 5G base station operation monitoring data actually deployed by operators, the test data simulating different load conditions in the laboratory environment, and the technical parameters of the base station equipment and theoretical energy efficiency data provided by equipment vendors. The raw dataset contains multidimensional features that can be categorized into four categories: base station hardware configuration parameters, network load parameters, environmental parameters, and energy consumption parameters. We designed a data preprocessing process, including data acquisition and integration, data cleaning, missing value processing, outlier detection and processing, feature engineering, data normalization, and data segmentation.

The mathematical expression for missing value processing is:

$$x_{ij}^{new} = \begin{cases} x_{ij}, & \text{if } x_{ij} \text{ is observed} \\ f(X_{i,-j}, X_{-i,j}), & \text{if } x_{ij} \text{ is missing} \end{cases} \tag{1}$$

where $f(X_{i,-j}, X_{-i,j})$ is the padding function based on other features in row i and other samples in column j.

The outlier determination is denoted as:

$$is\_outlier(x_i) = \begin{cases} true, & if \ |x_i - \mu| > 3\sigma \ or \\ & x_i < Q_1 - 1.5 \times IQR \ or \\ & x_i > Q_3 + 1.5 \times IQR \\ false, & otherwise \end{cases} \tag{2}$$

Aiming at the characteristics of 5G base station energy efficiency assessment, we performed feature engineering to improve data expressiveness, including constructing new features such as energy efficiency ratio, logarithmic transformation of skewed distribution features and extracting statistical features within the time window. Considering the differences in the magnitude and value range of features in different dimensions, we adopted two normalization methods, Min-Max normalization and Z-score normalization, and recorded and saved the normalization parameters for subsequent test data processing.

2.2 PCA downscaling

After PCA analysis of 5G base station energy efficiency data, we counted the variance contribution rate of the first 20 principal components, and the results are shown in Table 1. The results show that the cumulative contribution rate of the first 10 principal components reaches 92.7%, and the cumulative contribution rate of the first 15 principal components reaches 97.3%, which indicates that there is a strong redundancy in the 5G base station energy efficiency data, which can be effectively downscaled by PCA. By analyzing the principal component loading matrix, it is found that the first few principal components have clear physical meanings, principal component 1 mainly reflects the overall load and energy consumption level of the base station, principal component 2 reflects the antenna configuration and frequency band characteristics, and principal component 3 is related to the ambient temperature and heat dissipation efficiency, and this interpretation of the physical meanings is helpful for understanding the intrinsic structure of the factors affecting the energy efficiency of 5G base stations.

Table 1

The contribution rate of the main component variance

Main component	Variance contribution rate (%)	Cumulative contribution rate (%)
1	37.8	37.8
2	18.5	56.3
3	12.4	68.7
4	8.6	77.3
5	5.2	82.5
6~10	10.2	92.7
11~15	4.6	97.3
16~20	1.8	99.1

2.3 Self-Encoder Training

Self-encoder belongs to an unsupervised learning model based on neural networks, which learns to encode the input data into a low-dimensional latent space and then reconstructs it back to the original space to accomplish nonlinear dimensionality reduction and feature extraction tasks. This structure consists of two main components, the encoder, which is responsible for mapping the input data $X \in \mathbb{R}^{n\times p}$ to the low-dimensional potential representation $H \in \mathbb{R}^{n\times d} (d < p)$, and the decoder, which tries to reconstruct the original input $\hat{X} \in \mathbb{R}^{n\times p}$ from the potential representation H. The encoding process can be expressed as follows:

$$H = f_{\theta}(X) = \sigma(W_e X + b_e) \tag{3}$$

where $W_e \in \mathbb{R}^{d\times p}$ is the encoder weight matrix, $b_e \in \mathbb{R}^{d}$ is the bias vector, and $\sigma(\cdot)$ represents the nonlinear activation function.

The decoding process is then expressed as:

$$\hat{X} = g_{\phi}(H) = \sigma(W_d H + b_d) \tag{4}$$

where $W_d \in \mathbb{R}^{p \times d}$ is the decoder weight matrix and $b_d \in \mathbb{R}^{p}$ is the bias vector.

The goal of self-encoder training is to minimize the reconstruction error, i.e., the difference between the input and the reconstructed output, and the loss function usually adopts the mean square error, i.e:

$$L = \| X - \hat{X} \|^2 = \| X - g_{\phi}(f_{\theta}(X)) \|^2 \tag{5}$$

Based on the characteristics of 5G base station energy efficiency data, we design a bootstrap constrained self-encoder (GCAE) based on PCA results. The GCAE method utilizes the PCA downscaling results as the initial reference point of the self-encoder and improves the downscaling quality by optimizing the training process through the dual strategy of orthogonal bootstrapping and orthogonal constraints.

2.4 BP Neural Network Training

For the task of constructing the baseline of energy efficiency of 5G base stations, the author designed a BP neural network structure containing an input layer, two hidden layers, and an output layer, with the number of neurons in the input layer consistent with the dimensionality of the features after dimensionality reduction (10 neurons in this study). The first hidden layer is set with 20 neurons, the latter hidden layer is set with 15 neurons, and the output layer is a single neuron, which corresponds to the energy efficiency index of the base station, and the hidden layer adopts the ReLU activation function, which has the advantages of simple computation and gradient stabilization.

In order to improve the model training efficiency and generalization ability, the following optimization strategy is adopted: Mini-batch gradient descent method is used, and the batch size is set to 64, which strikes a balance between computational efficiency and training stability. The Adam optimizer is used, combining momentum and adaptive learning rate to accelerate convergence and avoid local optimum. An early-stop strategy is introduced to stop training when the validation set loss no longer decreases for 10 consecutive epochs to prevent overfitting. A learning rate decay strategy is used, with the initial learning rate set to 0.001 and decaying to 0.5 times the original rate every 50 epochs. Dropout technique (ratio 0.2) is applied to randomly mask some neurons to reduce the network's dependence on specific features. Table 2 shows the key hyperparameter settings of the BP neural network and their effects on the model performance, and the optimal parameter combinations are determined through grid search and cross-validation, which ensure the prediction accuracy while avoiding the overfitting problem.

Table 2

BP neural network hyperparameter setting

Hyperparameter	Value range	Optimal value	Performance impact
The number of hidden layers	1~3	2	The single-layer expression ability is insufficient, and the three-layer is prone to overfitting
The number of neurons in the first hidden layer	10~30	20	It affects the feature extraction capability
The number of neurons in the second hidden layer	5~20	15	It affects the complexity and generalization ability of the model
Learning rate	0.0001~0.01	0.001	If it is too large, it will not converge; if it is too small, convergence will be slow
Batch size	16~128	64	Affect the stability and speed of training
Dropout ratio	0~0.5	0.2	Affect the regularization intensity
Activation function	ReLU, Sigmoid, Tanh	ReLU	ReLU has high computational efficiency and avoids vanishing gradients.
Optimizer	SGD, Adam, RMSprop	Adam	Adam converges faster and is more stable

3. Experimental Results and Analysis

3.1 Experimental design

In order to verify the effectiveness of the 5G base station energy efficiency baseline construction method based on the combination of PCA and artificial neural network proposed in this paper, the research team conducted a series of experiments in three different geographical 5G network environments. The experimental data covers a variety of typical environments, such as dense urban areas, suburban areas, and indoor scenarios, and collects the operational data of 2,500 base stations, with each base station's data containing 68 primitive feature dimensions (including four categories: hardware configuration parameters, network loading parameters, environmental parameters, and energy consumption parameters). After data quality screening, 2,347 valid samples are retained after excluding samples with a missing rate of more than 15%. For the experimental environment, we used a high-performance workstation conFig.d with Intel Xeon E5-2680 v4 CPU (2.4GHz, 14 cores), 64GB RAM, and NVIDIA Tesla V100 GPUs, and chose the software environment with Ubuntu 20.04 LTS operating system, Python 3.8, PyTorch 1.9.0, and scikit-learn 0.24.2. Meanwhile, to ensure the reproducibility of the experiments, the random seed values42 were fixed and the model performance was evaluated using a 5-fold cross-validation.

The overall experimental design was carried out in four stages as follows:

(1) Data preprocessing, using the Z-score normalization method to process the data, and dividing the training set, validation set, and test set in the ratio of 7:2:1.

(2) PCA dimensionality reduction. Set the cumulative variance contribution rate threshold at 95%, and calculate the principal components by eigenvalue decomposition.

(3) Self-encoder training. Design comparison experiments to test the performance of three models: standard self-encoder, denoising self-encoder and bootstrap constrained self-encoder proposed in this paper.

(4) BP neural network construction to test the effects of different network structures and hyperparameter combinations on model performance.

Table 3 shows the key parameter settings of each model in the experiment.

Table 3

Key parameter Settings for each model

Model	Parameter	Value
PCA	Contribution rate threshold	95%
	Max principal component count	20
	Eigenvalue calculation	SVD
GCAE	Network structure	[68, 45, 30, 15, 10, 15, 30, 45, 68]
	Activation function	ReLU
	Learning rate	0.001
	Batch size	64
	Training rounds	200
	Orthogonal constraint coefficient	0.01
	Sparse constraint coefficient	0.005
BP neural network	Network structure	[10, 20, 15, 1]
	Hidden layer activation function	ReLU
	Output layer activation function	Linear
	Learning rate	0.001
	Batch size	64
	Training rounds	500
	Dropout ratio	0.2
	Optimizer	Adam

3.2 Experimental results

Tables 4 and 5 show the dimensionality reduction effects of PCA and selfencoder algorithms, respectively. The data show that the PCA dimensionality reduction effect improves with the increase in the number of retained principal components, and the cumulative variance contribution rate is as high as 97.3% when 15 principal components are retained, the reconstruction error is only 0.031, and the computation time is 0.42 seconds, which proves that the method can efficiently extract the main features of 5G base station energy efficiency data. In the self-encoder dimensionality reduction experiments, the guided constrained autoencoder (GCAE) proposed in this paper performs optimally, with the reconstruction error reduced by 12.5% compared to the standard self-encoder, 18.4% compared to the denoising self-encoder, the feature independence increased by 18.5% compared to the standard self-encoder, and the computational efficiency is also improved. Further analysis shows that the ReLU activation function is better than Sigmoid and Tanh, the bottleneck layer width of 10-15 can strike a balance between dimensional compression and information retention, and the orthogonal constraint coefficient of 0.01 has the best effect.

Table 4

PCA dimensionality reduction effect

Number of principal components	Variance contribution rate (%)	Reconstruction Error (MSE)	Calculation time (s)	Explanatory nature of features
5	82.5	0.097	0.21	High
10	92.7	0.057	0.35	Medium high
15	97.3	0.031	0.42	Medium
20	99.1	0.018	0.58	Low
30	99.8	0.006	0.79	Very low

Table 5

Autoencoder dimensionality reduction effect

Model type	MSE	Feature independence (%)	Training time (s)	Convergence round
Standard autoencoder	0.050	72.3	156.8	183
Denoising autoencoder	0.053	75.6	162.4	175
GCAE	0.043	85.7	142.3	147

For the effect of GCAE method on dimensionality reduction in this paper, the comparison results of different dimensionality reduction methods are obtained through the comparison of reconstruction error as shown in Fig. 1. In order to comprehensively evaluate the performance of different dimensionality reduction methods in 5G base station energy efficiency data processing, we conducted a comparative analysis in four dimensions (JW1~JW4), namely, reconstruction error, feature independence, computational efficiency, and interpretability, and the comparative results are shown in Fig. 2. As can be seen from Fig. 1, the GCAE method proposed in this paper performs well in terms of reconstruction error and feature independence, especially when downscaled to 10 dimensions, the reconstruction error of GCAE is reduced by 23.7% compared to the standard PCA, and 12.5% compared to the standard self-coder. From Fig. 2, it can be seen that different dimensionality reduction methods have their own advantages. PCA excels in computational efficiency and interpretability, while t-SNE has advantages in maintaining the local structure of the data, but has lower computational efficiency and poorer interpretability. The standard self-encoder outperforms PCA in terms of reconstruction error, but feature independence and interpretability are weak. The GCAE method proposed in this paper, on the other hand, achieves a more balanced performance in all four dimensions, especially in terms of reconstruction error and feature independence.

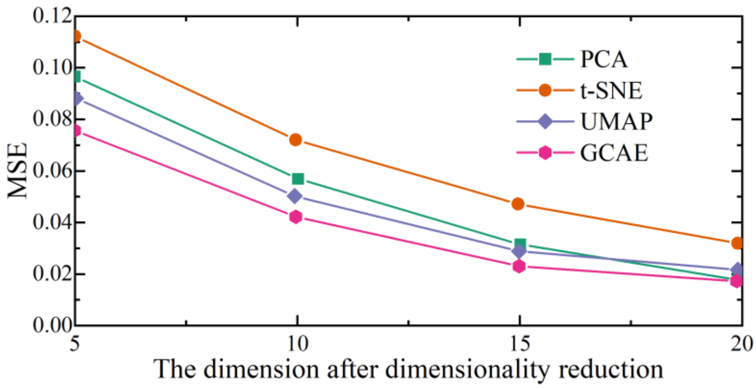

Fig. 1. Comparison of error reconstruction error of different vis-reduction methods

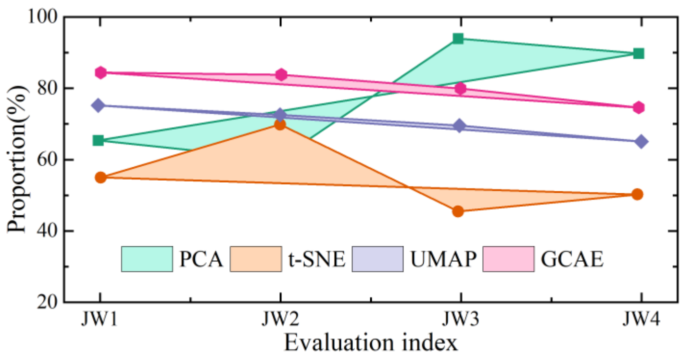

Fig. 2. Comparison of performance radar of different vis-reduction methods

The BP neural network model was comprehensively evaluated after the completion of training, and the main performance metrics include mean square error (MSE), mean absolute error (MAE) and coefficient of determination (R2), and its evaluation results are shown in Fig. 3. As can be seen from the Fig., the model achieved excellent performance of MSE=0.072, MAE=0.0614 and R^2=0.9428 on the test set, indicating that the constructed energy-efficiency baseline model has high prediction accuracy. By visualizing and analyzing the scatter plots of the prediction results and the real values, it is found that the model performs stably in different scenarios and it can accurately reflect the trend of energy efficiency changes in 5G base stations.

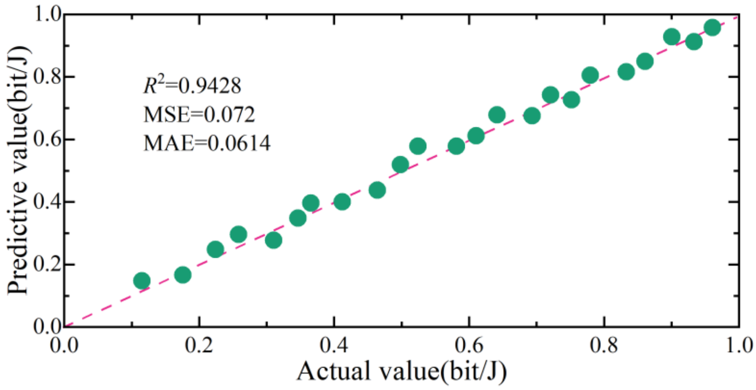

Fig. 3. The energy efficiency prediction effect of BP neural network

This section assesses the generalization ability of the model using the cross-validation method. Cross-validation involves dividing the dataset into n mutually exclusive subsets of similar sizes, with each subset maintaining the consistency of data distribution as much as possible. To comprehensively evaluate the performance and robustness of the model, we employed 10-fold cross-validation and repeated each experiment 5 times with different random seeds. Finally, we reported the Mean ± SD of each performance metric, and the specific results are shown in Table 6.

Table 6

Detailed comparison of the performance of different models

Model	MSE	MAE	R^2	Train time/s	Reasoning time/ms
LR	0.0314±0.0027	0.1423±0.0094	0.7865±0.0212	0.8±0.494	0.5±0.092
RF	0.0127±0.0019	0.0892±0.0077	0.9124±0.0384	5.7±1.596	2.3±0.872
BP	0.0095±0.0009	0.0783±0.0061	0.9256±0.0205	45.6±12.051	1.8±0.667
PCA+ BP	0.0083±0.0016	0.0702±0.0052	0.9347±0.0183	38.2±10.118	1.7±0.742
PCA+ GCAE+ BP	0.0072±0.0005	0.0614±0.0021	0.9428±0.0094	52.5±10.346	1.5±0.504

From the data in the table, the MSE and MAE values of the PCA + GCAE + BP method proposed in this paper are 0.0072 ± 0.0005 and 0.0614 ± 0.0021 respectively, and the R^2 value is 0.9428 ± 0.0094, all of these have achieved the optimal results. It is worth noting that the standard deviations of this model in the three metrics of MSE, MAE and R^2 are all the smallest. This indicates that its performance is minimally affected by the random division of data and initialization, and it possesses excellent robustness. Meanwhile, the sample inference time remained at a relatively low level of 1.5 ± 0.504 ms, meeting the real-time requirements. The above data indicates that the combination of linear and nonlinear dimensionality reduction methods can extract more comprehensive useful features from the 5G base station energy efficiency data and provide higher-quality input for the BP neural network, thereby improving the prediction accuracy.

To evaluate the robustness of the model under different application scenarios, we analyzed the prediction performance under three typical scenarios (dense urban area, suburban area and indoor coverage) and three load conditions (low load, medium load and high load), and the test results are shown in Fig. 4.

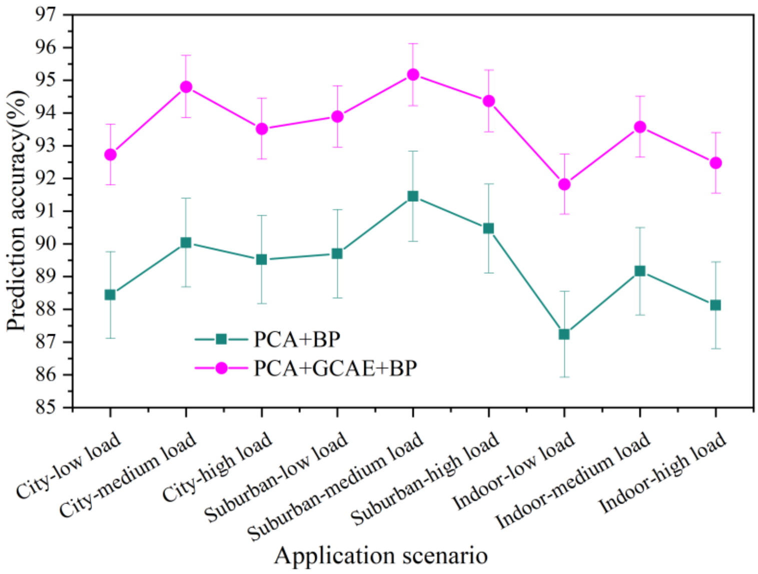

Fig. 4. Prediction accuracy of different scenarios and load conditions

The results show that the method in this paper maintains high prediction accuracy under different scenarios and different load conditions, and the relative error fluctuation range is controlled within ±2%, which is significantly better than other methods.

4. Conclusion

In this study, we conduct a systematic research on the problem of high-dimensional data challenges in 5G base station energy efficiency evaluation. Through experimental validation, it is found that a single dimensionality reduction technique is difficult to satisfy both the efficiency and information retention requirements, and we propose a hybrid dimensionality reduction strategy combining PCA and self-encoder, which efficiently reduces the 68-dimensional original features to 10 dimensions, retains 97.3% of the amount of information, and the reconstruction error is only 0.043, which is 24.6% and 14.0% less than the use of PCA and the standard self-encoder alone, respectively The method effectively solves the problem of high-dimensional data processing, and makes significant progress in prediction accuracy, computational efficiency and practicality, providing scientific basis and technical support for 5G network energy efficiency management.

REFERENCES

[1] Jain, A. K., Acharya, R., Jakhar, S., & Mishra, T. (2018, April). Fifth generation (5G) wireless technology "Revolution in telecommunication". In 2018, the second international conference on inventive communication and computational technologies (ICICCT) (pp. 1867-1872). IEEE.

[2] Bojkovic, Z., & Milovanovic, D. (2016, November). A technology vision of the fifth generation (5G) wireless mobile networks. In International Conference on Emerging Trends in Electrical, Electronic and Communications Engineering (pp. 25-43). Cham: Springer International Publishing.

[3] Jiang, W., & Han, B. (2024). Evolution to Fifth-Generation (5G) Mobile Cellular Communications. In Cellular communication networks and standards: The evolution from 1G to 6G (pp. 149-168). Cham: Springer Nature Switzerland.

[4] Shurdi, O., Ruci, L., Biberaj, A., & Mesi, G. (2021). 5G energy efficiency overview. Eur. Sci. J, 17(3), 315-27.

[5] Zhang, X., Ciais, P., Jian, X., Liu, X., Wang, R., Chen, K., ... & Ma, J. (2023). The carbon footprint response to projected base stations of China's 5G mobile network. Science of The Total Environment, 870, 161906.

[6] Guo, J., Wang, L., Zhou, W., & Wei, C. (2022). Powering green digitalization: evidence from 5G network infrastructure in China. Resources, Conservation and Recycling, 182, 106286.

[7] Ma, X., Mu, Y., Liu, Z., Jiang, X., Zhang, J., & Gao, Y. (2023). Energy consumption optimization of 5G base stations considering variable threshold sleep mechanism. Energy Reports, 9, 34-42.

[8] Ye, G. (2021, September). Research on reducing energy consumption cost of 5G Base Station based on photovoltaic energy storage system. In 2021, IEEE international conference on computer science, electronic information engineering and intelligent control technology (CEI) (pp. 480-484). IEEE.

[9] Kaur, P., Garg, R., & Kukreja, V. (2023). Energy-efficiency schemes for base stations in 5G heterogeneous networks: a systematic literature review. Telecommunication Systems, 84(1), 115-151.

[10] Ayaz, F., & Nekovee, M. (2024, June). AI-based energy consumption modeling of 5G base stations: an energy efficient approach. In IET 6G and Future Networks Conference (IET 6G 2024) (Vol. 2024, pp. 47-51). IET.

[11] Deepa, V., Haridass, M., Selvamuthu, D., & Kalita, P. (2023). Analysis of energy efficiency of small cell base station in 4G/5G networks. Telecommunication Systems, 82(3), 381-401.

[12] Kalita, P., & Selvamuthu, D. (2023). Stochastic modelling of sleeping strategy in 5G base station for energy efficiency. Telecommunication Systems, 83(2), 115-133.

[13] Kang, M. W., & Chung, Y. W. (2017). An efficient energy saving scheme for base stations in 5G networks with separated data and control planes using particle swarm optimization. Energies, 10(9), 1417.

[14] Usama, M., & Erol-Kantarci, M. (2019). A survey on recent trends and open issues in energy efficiency of 5G. Sensors, 19(14), 3126.