Renewable Energy and Power Quality Journal
https://repqj.com/index.php/repqj/index
RE&PQJ, Vol. 23, No. 5, 2025

Extracting Hidden Features of Load Fluctuation in Photovoltaic-Energy Storage System by Variational Autoencoder

Shilong Liu*

Qinghai Huanghe Hydropower Development Co., Ltd., Xining, 810000, Qinghai, China
*Corresponding author's email: goajjian665@yeah.net

Abstract. Aiming at the nonlinearity, noise interference, and periodicity problems of load fluctuation in photovoltaic-energy storage systems, this study proposes a hidden feature extraction method based on VAE. VAE is used to extract key low-dimensional representations to improve the robustness and accuracy of the load forecasting model. This article encodes the collected load data, maps the original high-dimensional data to a low-dimensional latent space using a VAE encoder, and outputs the mean and logarithmic variance vectors at the same time. Eight key latent features are obtained by sampling through reparameterization techniques, including daily cycle amplitude features, peak period features, load fluctuation frequency, etc. The extracted latent features are fused with the original data and used as input to the Informer model to predict future loads. The experiment uses a photovoltaic-energy storage system data set with a data sampling interval of 20 minutes. The results show that VAE-Informer achieves 1.80%, 2.10% and 0.97 in terms of Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE) and R^2, respectively, which is significantly better than Informer (RMSE is 3.50, MAPE is 4.80% and R^2 is 0.91), Sparse Autoencoder (SAE)-Informer (RMSE of 2.50, MAPE of 3.20%, R^2 of 0.95) and Principal Component Analysis (PCA)-Informer (RMSE of 3.00, MAPE of 3.80% and R^2 of 0.93) models. The efficiency of this method in latent feature extraction and load forecasting is fully verified. This method based on VAE to extract latent features and combined with Informer for load forecasting can accurately capture the complex nonlinear and periodic characteristics in load data and improve the prediction accuracy.

Key words. Photovoltaic power generation, Energy storage system, Load fluctuation, Latent features, Variational autoencoder

1. Introduction and Literature Review

With the large-scale access of renewable energy [1,2] and the continuous advancement of smart grid [3,4] technology, photovoltaic power generation and energy storage systems are playing an increasingly important role in modern power systems [5,6]. However, photovoltaic power generation [7,8] is affected by a variety of uncertain factors such as sunshine and weather [9,10]. Its output power and load have obvious volatility, nonlinearity and noise interference, which brings great challenges to system scheduling and safe and stable operation [11,12]. Traditional methods often fail to fully capture the inherent laws of data when processing high-dimensional complex data, and are easily affected by noise, resulting in insufficient prediction accuracy [13,14].

In response to the high-dimensionality, nonlinearity and noise interference of load data, scholars have proposed a variety of feature extraction methods to improve the performance of prediction models. Traditional dimensionality reduction techniques such as principal component analysis [15,16] and linear discriminant analysis [17] can simplify the data to a certain extent, but are clearly insufficient in capturing complex nonlinear dynamic characteristics. With the rise of deep learning technology, autoencoders [18,19] have been widely used in data denoising and feature extraction because they can learn low-dimensional data representation by minimizing reconstruction error. The latent variable space of the autoencoder lacks probability distribution constraints, resulting in insufficient continuity and generalization ability of the latent variables [20]. To solve this problem, researchers proposed a variational autoencoder. It applies KL divergence regularization on the basis of the traditional autoencoder structure to make the latent variable close to the prior distribution and ensure the smoothness and interpretability of the latent space [21,22]. When applying variational autoencoders to feature learning of time series data in fields such as meteorology and power load, good results have been achieved, but their application in load forecasting of photovoltaic-energy storage systems is still in the exploratory stage [23,24].

In the field of long-series time forecasting, the Transformer [27,28] model has attracted widespread attention for its ability to capture global information, but when processing ultra-long time series data, its computational complexity and memory requirements become the main bottleneck. The Informer model [29,30] effectively solves the performance problem in long-series forecasting by introducing sparse attention mechanism and efficient encoding strategy, and shows significant advantages in application scenarios such as power load and traffic flow. Existing research focuses on directly using Informer [31,32] to predict raw data, but pays less attention to the extraction and fusion of latent features in the data. In recent years, some scholars have begun to explore the idea of feature extraction, combining deep generative models with Transformer architecture. By first using generative models such as VAE [33,34] to extract low-dimensional latent features of data, and then using the extracted features as input, data noise and redundant information can be reduced to a certain extent, and the model's sensitivity to abnormal fluctuations and prediction accuracy can be improved.

As an advanced deep learning tool, variational autoencoders (VAEs) have shown unique advantages in feature extraction and data modeling. Unlike traditional autoencoders, VAEs can learn the probability distribution of data by introducing a probabilistic generation model, thereby achieving the ability to generate new samples from the latent space. This feature makes VAEs more flexible and adaptable when dealing with data with uncertainty and complex structures. In addition, the latent space of VAEs is carefully designed to ensure its smoothness and continuity through regularization, effectively avoiding the discontinuity and irregularity problems that may occur in latent variables in traditional autoencoders, which makes VAEs perform well in feature interpolation and extrapolation.

This study constructs a hybrid model that combines VAE latent feature extraction with Informer long sequence prediction, which systematically solves the high dimensionality, nonlinearity and noise interference problems of load data in photovoltaic storage systems. The VAE deep learning framework is used to encode and reconstruct the original load data to achieve data denoising and feature extraction, effectively capturing daily cycles, seasonal changes and nonlinear dynamic characteristics. The extracted low-dimensional latent variables are used as the input of the Informer model, and its efficient sparse attention mechanism and long sequence modeling capabilities are used to accurately predict load trends. This method breaks through the shortcomings of traditional statistics and simple dimensionality reduction methods in feature expression. Experimental verification shows that the model has significant advantages in load forecasting accuracy, provides a solid theoretical basis and practical guidance for smart grid scheduling and energy storage system optimization, and opens up new ideas for the application of deep generative models in the field of power systems.

2. Hidden Feature Extraction and Load Forecasting Methods

A. Data Acquisition and Preprocessing

In this study, data collection is based on a photovoltaic-energy storage system from January 1, 2021 to January 1, 2022. Data collection is carried out every 20 minutes to ensure the continuity and integrity of time series data. The collected data covers core indicators such as system load, photovoltaic power generation output, and energy storage system status, supplemented by meteorological data and other external environmental variables, providing sufficient information support for subsequent model construction and feature extraction. System load data reflects the electricity demand of the entire system in different time periods, while photovoltaic power generation output records the power generation of solar power generation equipment in different time periods. Energy storage status data mainly describes the charging and discharging status of energy storage devices and their remaining capacity. Meteorological data can reveal the impact of weather factors on photovoltaic output and load changes, such as the changing trends of temperature and solar radiation. By performing time-series synchronization on these multi-dimensional data, the consistency and accuracy of each indicator in time are ensured, thus building a high-quality and comprehensive database.The collected data is shown in Table 1.

Table 1. Display of collected data.

Timestamp	System load (MW)	Photovoltaic output (MW)	Storage state (%)	Temperature (°C)	Solar irradiance (W/m²)
2021/1/1 0:00	115.2	20.1	78	4.8	0
2021/1/1 0:20	115.7	20.3	77	4.9	0
2021/1/1 0:40	116.0	20.5	78	5.0	0
…					
2021/1/1 8:20	164.0	21.2	76	8.5	120
2021/1/1 8:40	169.5	21.5	76	9.8	160
2021/1/1 9:00	174.0	21.8	75	10.0	200
2021/1/1 9:20	178.2	22.0	75	12.2	230
…					
2022/1/1 0:00	119.0	22.5	74	6.8	0

The system load, photovoltaic output and energy storage status in the early morning hours remain relatively stable, and the solar radiation is 0, which is consistent with the physical reality that the sun does not rise at night. The data in the late morning show that the temperature, photovoltaic output and solar radiation gradually increase, indicating that as the ambient temperature rises after sunrise, the solar radiation increases, photovoltaic power generation begins to play a role, and the system load also shows an upward trend.

The missing values of the original collected data are processed. For continuous variables, such as system load and photovoltaic output, the mean filling method is used. The observed values of a certain feature are $x_1, x_2, \cdots, x_n$, the missing value is recorded as x_j, and the mean $\overline{x}$ is used to fill it. The formula is:

$$\overline{x} = \frac{1}{n'} \sum_{i=1}^{n'} x_i \quad (1)$$

In formula 1, n' is the number of non-missing samples. If there are too many missing values in a sample, exceeding 55%, the sample can be removed to avoid adverse effects on the overall model training.Outlier detection is an important step in data preprocessing. The Z-score standardization method is used to determine whether it is an outlier by calculating the difference between each data point and the mean. The calculation formula is:

$$z = \frac{x - \mu}{\sigma} \quad (2)$$

In Formula (2), σ represents the standard deviation of the data points and μ is the mean.For outliers, neighboring data interpolation is used to replace them to ensure the rationality of data distribution and reduce the interference of noise on model training. Align each data source according to the timestamp, set a unified sampling interval, standardize the timestamp, and then merge the data. For time alignment, if there is a time gap in a data source, linear interpolation is used to fill it to ensure that each feature has a corresponding value at the same time point.

In the process of data preprocessing, normalization is a key step, especially when training deep learning models, different variables have different value ranges. If normalization is not performed, data of different scales may cause unstable gradient descent, affecting the model convergence speed and prediction performance.

The mathematical formula of Min-Max normalization is:

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (3)$$

Among them, $x_{\min}$ is the minimum value of the data, $x_{\max}$ is the maximum value of the data, and x' is the data after regularization.

B. VAE Model Construction and Latent Feature Extraction

VAE is a generative model based on probabilistic graphical models. Its key goal is to learn the probability distribution of latent variables and generate samples that approximate the original data distribution by sampling latent variables. The core structure of VAE includes encoder, latent variable sampling layer and decoder.

The VAE model is shown in Figure 1.

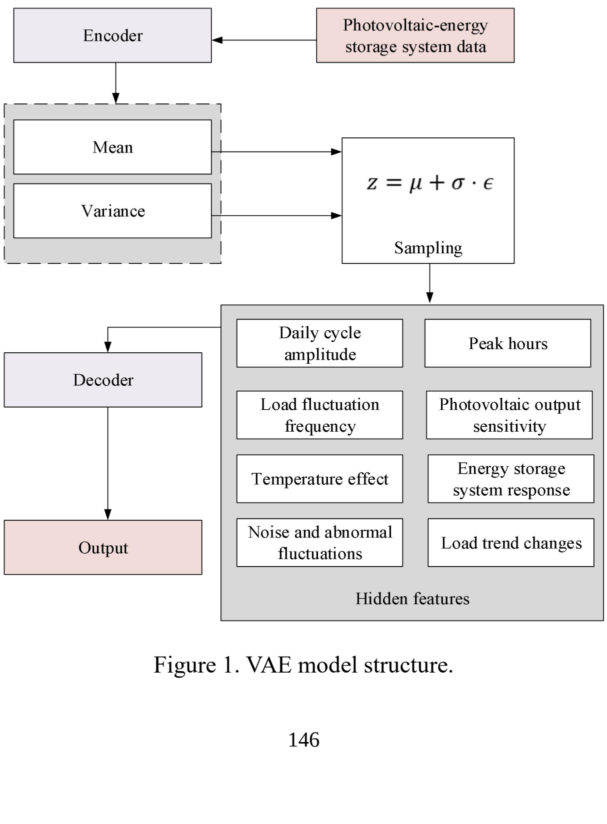

Figure 1. VAE model structure.

The encoder is used to map the data to a low-dimensional latent space to extract key latent features, including daily cycle amplitude features, peak period features, load fluctuation frequency features, photovoltaic output sensitivity features, temperature impact features, energy storage system response features, noise and abnormal fluctuation features, and load trend change features.

The dimension X of the input data is determined by T (time step) and d (number of features per time step). We use a multi-layer fully connected network as the hidden layer, with the number of neurons in each layer gradually decreasing, to achieve feature extraction and dimensionality reduction:

$$h_1 = f\left(W_1 X + b_1\right) \quad (4)$$

$$h_2 = f\left(W_2 h_1 + b_2\right) \quad (5)$$

In formulsa 4-5, W_1 , b_1 , W_2 , b_2 are network parameters. The encoder outputs two vectors, the mean and variance of the latent variable, and the formulas are:

$$\mu = W_\mu h_2 + b_\mu \quad (6)$$

$$\log \sigma^2 = W_\sigma h_2 + b_\sigma \quad (7)$$

In formulas 6-7, W_μ and W_σ are weight matrices mapped to mean and log variance. μ and σ^2 jointly determine the probability distribution of latent variables.In order to enable VAE to be optimized by gradient descent, the reparameterization technique is used for sampling:

$$z = \mu + \sigma \cdot \epsilon \quad (8)$$

In formula 8, ϵ follows the standard normal distribution $\epsilon \sim \mathcal{N}\left(0, I\right)$, which allows the gradient to flow to the mean and variance, thereby performing back-propagation optimization. The role of the decoder is to map the latent variables back to the original data space, thereby achieving reconstruction. The latent variables are used as the input of the decoder and are gradually mapped back to the original data dimension:

$$h_3 = f\left(W_3 z + b_3\right) \quad (9)$$

$$h_4 = f\left(W_4 h_3 + b_4\right) \quad (10)$$

$$\hat{X} = \sigma\left(W_{\text{out}} h_4 + b_{\text{out}}\right) \quad (11)$$

In formula 11, $\sigma(\cdot)$ uses the Sigmoid activation function to keep the output range in [0,1] to ensure numerical stability.The formula for reconstruction loss is:

$$L_{\text{recon}} = \frac{1}{N} \sum_{i=1}^{N} \left\| X_i - \hat{X}_i \right\|^2 \quad (12)$$

In Formula 12, $\hat{X}_i$ is the reconstructed output of the VAE model, and X_i is the real data.KL divergence measures the difference between the posterior distribution $q\left(z \middle| X\right)$ and the standard normal distribution:

$$L_{\text{KL}} = -\frac{1}{2} \sum_{j=1}^{d_z} \left(1 + \log \sigma_j^2 - \mu_j^2 - \sigma_j^2\right) \quad (13)$$

The total loss function of VAE consists of reconstruction loss and KL divergence, and the formula is:

$$L = L_{\text{recon}} + \beta L_{KL} \quad (14)$$

In formula 14, β is a balance coefficient used to control the trade-off between reconstruction ability and regularization strength. When $\beta > 1$, the model is more inclined to learn good latent variable representation, but it may affect the reconstruction ability. When $\beta < 1$, the model pays more attention to the reconstruction quality, but it may cause the latent variable distribution to deviate too much from the standard normal distribution.

In order to effectively train the VAE model, the Adam optimizer is used to achieve stable gradient updates, and the formula is:

$$\theta^* = \arg\min_{\theta} L\left(\theta\right) \quad (15)$$

C. Informer Model Construction and Latent Feature Fusion Prediction

Informer is an improved model based on Transformer, designed for long-sequence time prediction. Due to its global self-attention mechanism, the traditional Transformer has high computational complexity when modeling long sequences, which limits its application in large-scale time series tasks. Informer uses a sparse attention mechanism to focus only on key query points, improving computational efficiency. Informer uses multi-scale convolution filtering for dimensionality reduction and bidirectional residual connections to improve feature expression capabilities, thereby having stronger generalization capabilities in long-term forecasting tasks.

The process of load forecasting by the Informer model combined with latent features is shown in Figure 2.

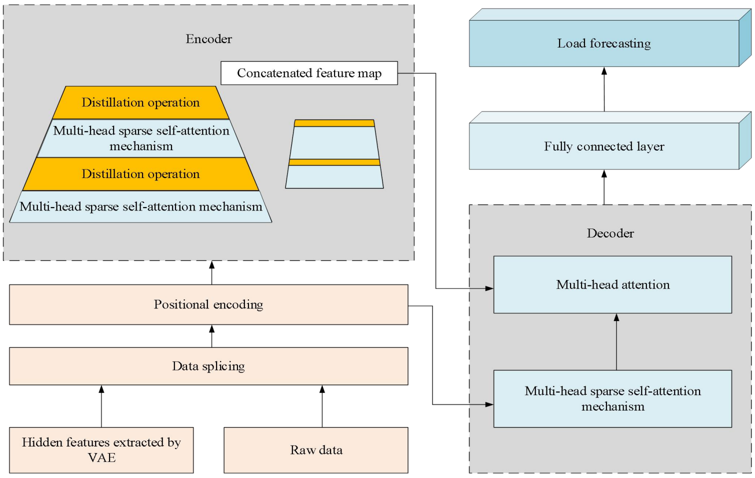

Figure 2. The process of load forecasting by combining the Informer model with latent features.

The latent features are concatenated with the original load data so that the model can use both the key features after dimensionality reduction and the original information. The input sequence is $X \in \mathbb{R}^{L\times d}$, L is the time step, and d is the input feature dimension. The latent variable $Z \in \mathbb{R}^{L\times k}$ after VAE processing, then the complete input is represented as:

$$X_{\text{input}} = [X, Z] \quad (16)$$

Time series data has a strict time order. In order to preserve the time sequence information, Informer uses sine-cosine position encoding:

$$PE_{(t,2i)} = \sin\left(\frac{t}{10000^{\frac{2i}{d}}}\right) \quad (17)$$

$$PE_{(t,2i+1)} = \cos\left(\frac{t}{10000^{\frac{2i}{d}}}\right) \quad (18)$$

In formulas 17-18, t is the time step and i is the feature index. The position encoding vector is added to the input data to enhance the model's time perception.Informer's encoder contains multiple Transformer layers with sparse self-attention:

$$\text{PSA}(\text{Q}, \text{K}, \text{V}) = \text{softmax}\left(\frac{\text{QK}^T}{\sqrt{d}}\right)\text{V} \quad (19)$$

The encoder uses multi-head attention, the formula is:

$$\text{MultiHead}(\text{Q}, \text{K}, \text{V}) = \text{Concat}(\text{head}_1, \cdots, \text{head}_h) W_O \quad (20)$$

Informer uses the Adam optimizer:

$$m_t = \beta_1 m_{t-1} + (1-\beta_1) g_t \quad (21)$$

$$v_t = \beta_2 v_{t-1} + (1-\beta_2) g_t^2 \quad (22)$$

$$\theta_t = \theta_{t-1} - \alpha \frac{m_t}{\sqrt{v_t + \epsilon}} \quad (23)$$

In formulas 21-23, m_t and v_t are the first-order and second-order momentum of the gradient, respectively, and α is the learning rate.

In terms of training strategy, a phased training method is used to fix the VAE part so that it can stably extract potential features and reduce interference with downstream tasks. On this basis, the Informer model is trained with the potential features extracted by VAE so that it can fully learn the temporal pattern of load data and the correlation between multidimensional features. In order to prevent the model from overfitting, L2 regularization and early stopping strategies are introduced during the training process. The performance changes of the model are monitored by setting the number of training rounds and thresholds to ensure that the model maintains a high generalization ability under the premise of stable convergence. In this study, data collected from January 1, 2021 to January 1, 2022 were used for training, totaling approximately 35,040 data points. The model predicts the load data for the next 24 hours and uses an independent test set containing 2,628 data points covering two weeks of load data for verification to ensure the accuracy and reliability of the prediction results.

In this study, in order to optimize the performance of the VAE-Informer model, this study systematically adjusted the hyperparameters. For the VAE part, this study selected the latent dimension as 8, which can effectively extract key features and reduce computational complexity; the hidden dimension is set to 128 to balance the feature expression ability and model complexity; the KL weight is set to 0.01 to balance the reconstruction ability and regularization strength. In the Informer part, the encoder layer is set to 4 layers and the decoder layer is set to 2 layers to efficiently capture long-term dependencies; the feedforward network dimension is set to 512 to improve the model expression ability. We use the Adam optimizer, the learning rate is set to 0.001, and the batch size is set to 64 to ensure training stability and efficiency. The input dimension is 16, the prediction range is 24, and the number of iterations is set to 100. The selection of these hyperparameters is based on multiple experimental verifications and can give full play to the model performance., the parameter initialization settings of the VAE-Informer model are shown in Table 2.

Table 2. Parameter initialization of the VAE-Informer model.

VAE parameters	Value	Informer parameters	Value
Latent dimensions	8	Encoder Layers	4
Hidden dimensions	128	Feedforward network dimensions	512
Optimizer	Adam	Learning rate	0.001
KL weight	0.01	Batch size	64
Epochs	100	Forecast horizon	24
Input dimensions	16	Decoder Layers	2

3. Experimental Environment and Model Evaluation

This study is based on NVIDIA Tesla series GPUs. The specific models include 4 NVIDIA Tesla V100 graphics cards, each with 32GB of video memory, which are interconnected through NVLink to achieve efficient parallel computing. This hardware configuration can accelerate the training and inference process of deep learning models, significantly improve the efficiency of large-scale matrix operations, thereby improving training efficiency and reducing convergence time. This study used a dual-core Intel Xeon Gold 6248 processor with a main frequency of 2.50GHz, a system equipped with 512GB of memory, and ran the Ubuntu 20.04 operating system, providing powerful computing support and a stable operating environment for deep learning tasks.

The experiment was developed and tested based on Python 3.8.10 and above, and the Anaconda environment management system was used to ensure the consistency and reproducibility of the dependent environment. This study used the TensorFlow 2.6.0 deep learning framework for model building and training, and used its efficient automatic differentiation mechanism and GPU acceleration function to simplify the model development process and improve training efficiency. In terms of data preprocessing and analysis, we used Pandas 1.3.5 and NumPy 1.21.5 for data loading, cleaning, feature engineering, and matrix calculation, giving full play to their convenience and efficiency in data operations. For visual analysis, this study used Matplotlib 3.4.3 to draw loss change curves, prediction result comparison charts, and confusion matrices, etc., to intuitively display the model training process and prediction results, and assist in model tuning and result interpretation.

In order to comprehensively evaluate the prediction performance of the VAE-Informer model, this study used three common error indicators. RMSE measures the square root of the mean square error between the predicted value and the true value, reflecting the overall error level of the model. The formula is:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2} \quad (24)$$

In formula 24, y_i is the true value, $\hat{y}_i$ is the model prediction value, and n is the number of samples.MAPE measures the percentage of the prediction error relative to the true value, and the formula is:

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \times 100\% \quad (25)$$

R^2 reflects the model's ability to explain data variance, and the formula is:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2}{\sum_{i=1}^{n}\left(y_i - \overline{y}\right)^2} \quad (26)$$

In formula 26, $\overline{y}$ is the mean of the true value. When R^2 is close to 1, it indicates that the model has a good fitting effect.After the model training is completed, it is evaluated on the test set to measure its generalization ability and performance in real application scenarios. In order to verify the effectiveness of the latent features extracted by VAE in improving the performance of time series prediction, this article designs a comparative experimental analysis, directly uses the original load data to train the Informer model without extracting latent features. Other latent feature extraction methods (PCA, SAE) are used and compared with the features extracted by VAE.

4. Results

A. Hidden Feature Visualization

In this study, the t-distributed stochastic neighbor embedding (t-SNE) method is used to visualize the hidden features extracted by VAE, and the results are shown in Figure 3.

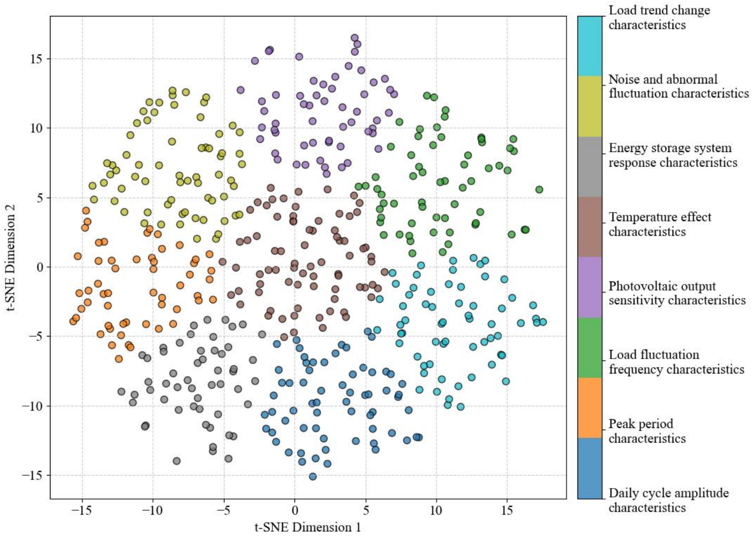

Figure 3. t-SNE latent feature visualization.

In this study, the eight potential features extracted by the VAE model represent the daily cycle amplitude, peak cycle, load fluctuation frequency, photovoltaic output sensitivity, temperature impact, energy storage system response, noise and abnormal fluctuations, and load trend changes. After reducing these high-dimensional potential features to two-dimensional space through t-SNE technology, points of different colors represent samples of different explicit potential features. The cluster corresponding to the load fluctuation frequency feature reflects the stability of the load fluctuation law of the sample at a specific frequency, which is usually related to the operating cycle of industrial equipment or the regularity of residential electricity consumption patterns. The cluster corresponding to the energy storage system response feature reflects the response speed and consistency of the discharge strategy of the energy storage system in the process of regulating load peaks and valleys, which is of great significance for optimizing the charge and discharge management of the energy storage system. The cluster corresponding to the temperature impact feature reflects the comprehensive impact of temperature changes on photovoltaic output and load demand, which helps to predict the balance of power supply and demand under extreme weather. These clustering structures not only show the distribution of data in the latent space, but also directly correspond to the key physical factors and dynamic patterns in the operation of the photovoltaic storage system. The cluster boundaries in the overall image are clear, indicating that the VAE model can effectively remove interference information and compress complex high-dimensional data into low-dimensional representations with practical significance, providing an intuitive visual tool for understanding the inherent mechanism of load fluctuations in photovoltaic storage systems.

The distribution of each cluster in the t-SNE visualization results not only reveals the obvious differences between the latent features, but also reflects the various dynamic modes that exist in the actual operation of the photovoltaic-energy storage system. Each cluster corresponds to a dominant latent feature, and the data points in the same cluster are relatively concentrated, indicating that these samples have a high degree of consistency in the performance of the corresponding features. The cluster corresponding to the load fluctuation frequency feature may reflect the sample with a relatively stable load fluctuation law at a specific frequency, while the cluster of the energy storage system response feature may focus on the response speed and consistency of the discharge strategy of the energy storage system in the process of adjusting the load peak and valley.

B. Load Prediction Accuracy

This article uses the potential features extracted by VAE and combines it with Informer for load forecasting. The test set contains 2628 data points. The comparison between the model prediction and the actual load change is shown in Figure 4.

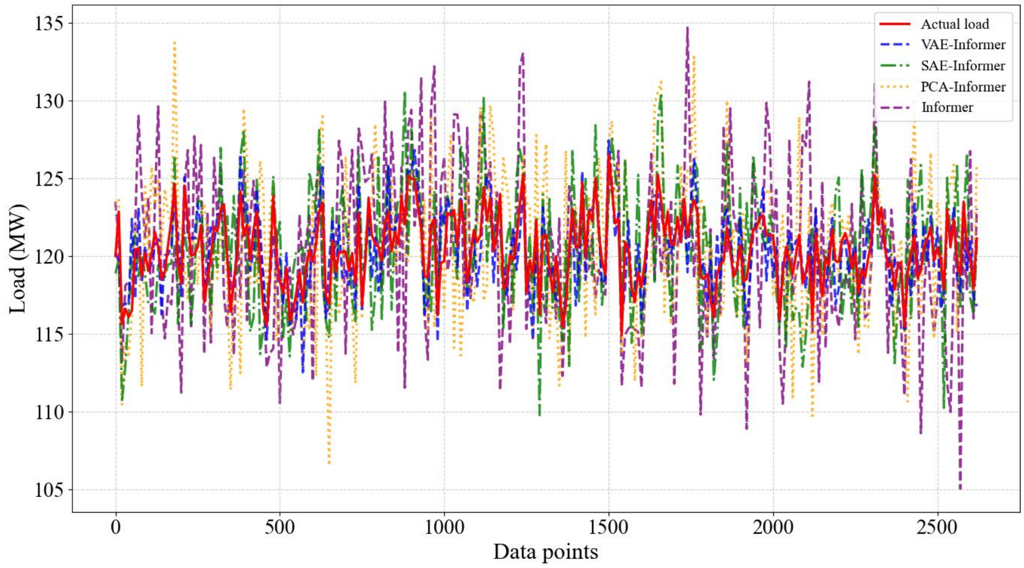

Figure 4. Model prediction and actual load changes.

The VAE-Informer model is closest to the actual load data compared to other models (Informer, SAE-Informer, PCA-Informer), indicating that the model has stronger adaptability in learning time series patterns and extracting deep features. VAE can effectively capture the potential distribution of data, and reduce the dimension, denoise and reconstruct features of the original input data, thereby enhancing the ability to express complex nonlinear load data. Compared with the standard Informer model, VAE-Informer introduces latent variable distribution constraints in the encoding process, which enables the model to have better generalization ability when facing uncertain factors (such as power consumption fluctuations, weather effects, etc.). The prediction results of VAE-Informer are closer to the actual load in terms of trend and value, with the smallest deviation. In comparison, the performance of the SAE-Informer model is inferior. Although SAE can also extract high-dimensional features, its feature extraction method is slightly simpler than that of VAE, and is easily limited by the local optimal solution, resulting in slightly larger prediction errors. PCA-Informer, because its dimensionality reduction method is mainly based on linear projection, cannot well capture the complex nonlinear structure in the load data, so the prediction error is further increased. The worst model is Informer, which only relies on its own attention mechanism to model long-term time dependencies. However, due to the lack of additional feature extraction capabilities, it is easily affected by factors such as data noise and mutation points, resulting in the largest prediction error.

The analysis of error sources is of great significance for optimizing load forecasting models. The error of VAE-Informer mainly comes from the sampling noise and reconstruction error of the latent variable distribution, but due to the regularization characteristics of VAE, these errors are small and have little impact on the prediction results. The error of SAE-Informer is mainly limited by the sparsity constraint of SAE. Although it helps to extract key features, when there are drastic fluctuations in load data, some important features may be ignored, resulting in a certain degree of information loss. In addition, since PCA-Informer uses a linear dimensionality reduction method, it can lose some nonlinear information when processing highly nonlinear time series data. In particular, when the load data presents complex periodic changes, it is difficult to accurately capture the dynamic changes of load peaks and valleys. The error of the Informer model comes from the limitations of its attention mechanism. When modeling long time series, it may be affected by gradient vanishing and computational complexity, thereby reducing the ability to model long-term dependencies. VAE-Informer can improve the accuracy of power load scheduling and reduce energy waste caused by prediction errors due to its lower prediction error, and is suitable for real-time load forecasting of smart grids.

The load forecasting performance is evaluated by RMSE, MAPE, and R^2, and the results are shown in Table 3.

Table 3. Load forecasting performance.

Model	RMSE	MAPE (%)	R2
Informer	3.50	4.80	0.91
VAE-Informer	1.80	2.10	0.97
SAE-Informer	2.50	3.20	0.95
PCA-Informer	3.00	3.80	0.93

Among them, the RMSE of the VAE-Informer model is only 1.80, the MAPE is 2.10%, and the R^2 reaches 0.97, which is significantly better than other models, indicating that the model can show higher accuracy in capturing the complex dynamic characteristics in load data. The reason why VAE-Informer performs well is mainly due to its ability to extract latent features using VAE. VAE maps high-dimensional input data to low-dimensional latent space through a probabilistic generation model, and introduces KL divergence constraints to ensure the continuity and controllability of the latent space, thereby effectively extracting key information from load data (such as daily cycles, peak hours, nonlinear fluctuations, etc.). This feature extraction method not only reduces the impact of data noise, but also enhances the model's robustness to abnormal fluctuations. Therefore, in the prediction process, VAE-Informer can more accurately reproduce the load change trend. Relatively speaking, the Informer model does not combine the implicit feature extraction mechanism, and its prediction results have an RMSE of 3.50, a MAPE of 4.80%, and an R^2 of only 0.91, indicating that its ability to capture complex nonlinear changes is weak. On this basis, the SAE-Informer model uses sparse autoencoders to extract features, which can improve the load forecasting performance to a certain extent. Its RMSE is 2.50, MAPE is 3.20%, and R^2 is 0.95. However, it is still not as good as VAE-Informer in handling randomness and diversity in data. PCA-Informer uses principal component analysis for linear dimensionality reduction. Although it is easy to calculate and reduces the data dimension to a certain extent, since load data usually presents a complex nonlinear structure, its prediction results (RMSE 3.00, MAPE 3.80%, R^2 0.93) are still difficult to fully reflect the actual load fluctuations. Overall, the evaluation indicators show that VAE-Informer has obvious advantages in balancing reconstruction error and model generalization ability. Its low RMSE and MAPE as well as high R^2 value prove the feasibility and superiority of using VAE to extract latent features combined with Informer for load forecasting.

RMSE reflects the average deviation between the model prediction value and the actual load data. The lower the RMSE, the higher the prediction accuracy. MAPE provides a percentage perspective of relative error, which can intuitively show the proportion of prediction error in the actual load. R^2 is used to measure the model's ability to explain data variance. The closer it is to 1, the more comprehensive the model's capture of actual data fluctuations. In this experiment, VAE-Informer fully utilized the advantages of the VAE model in latent feature extraction and effectively characterized the nonlinear features and complex dynamic patterns in the data, which greatly reduced the deviation between the predicted value and the actual load. This is directly reflected in its RMSE and MAPE indicators, which are significantly better than other models. In contrast, the Informer model does not introduce additional feature extraction mechanisms and directly relies on the Transformer architecture to capture temporal dependencies, resulting in large prediction errors when facing noise and random fluctuations in actual data. Although SAE-Informer uses sparse autoencoders to extract some effective features, it is inferior to VAE in terms of data expression ability and feature robustness, which leads to a certain degree of prediction bias. Since PCA-Informer uses a linear dimensionality reduction method, it is limited in processing highly nonlinear and complex fluctuation patterns, and its prediction effect is between Informer and SAE-Informer. The prediction model that uses VAE to extract latent features and combines it with Informer can more accurately reproduce the real load changes when processing actual photovoltaic-energy storage system load data, providing more reliable data support for smart grid scheduling and load management. Its performance indicators perform best in the current comparative experiments.

C. Prediction Loss

This article compares the impact of three latent feature extraction methods, VAE-Informer, SAE-Informer, and PCA-Informer, on load forecasting. The changes in loss values of different models are shown in Figure 5.

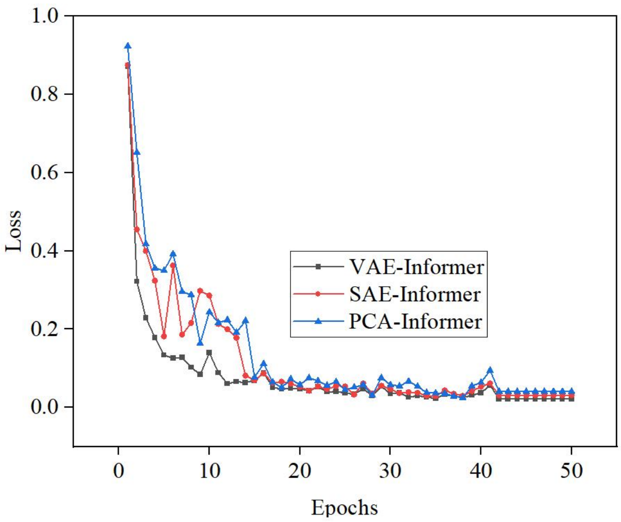

Figure 5. Changes in loss values of different models

In the initial training stage (epoch=1), the loss values of the VAE-Informer, SAE-Informer and PCA-Informer models are 0.872, 0.875 and 0.923 respectively. Each model can capture the basic characteristics of the load data in the early stage, but the difference is not significant. After 41 rounds of training, all three have converged, and the final loss values have dropped to 0.022, 0.031, and 0.041, respectively, with VAE-Informer having the lowest loss. This shows that the method of extracting latent features using VAE is more efficient in expressing complex nonlinear load data. VAE encodes the input data and uses KL divergence to constrain the distribution of latent variables, so that the latent space remains smooth and continuous, thereby more accurately capturing the periodic fluctuations, nonlinear dynamics and noise characteristics in the load data. In contrast, although SAE-Informer uses a sparse autoencoder, it lacks strict regularization of the latent space distribution during feature extraction, resulting in a slightly inferior ability to capture subtle changes in the load. PCA-Informer is limited by the linear dimensionality reduction method and has difficulty in revealing the complex nonlinear relationships in the data, which ultimately shows a higher loss value. The lower loss value directly reflects the model's ability to retain information and the accuracy of feature expression when reconstructing load data. Using VAE to extract latent features and combining it with Informer for effective noise reduction also significantly improves the ability to capture the deep structure of load data, providing a more accurate and robust input representation for load forecasting.

D. *Stability Analysis*

The load characteristics of different time periods are shown in Table 4.

Table 4. Load characteristics of different time periods.

Time period type	Time range	Load level	Average load value (MW)
Peak period	19:00-21:00	Very high	135
High period	8:00-12:00, 18:00-22:00	High	125
Flat period	6:00-8:00, 12:00-18:00	Medium	115
Trough period	22:00-6:00 the next day	Low	105

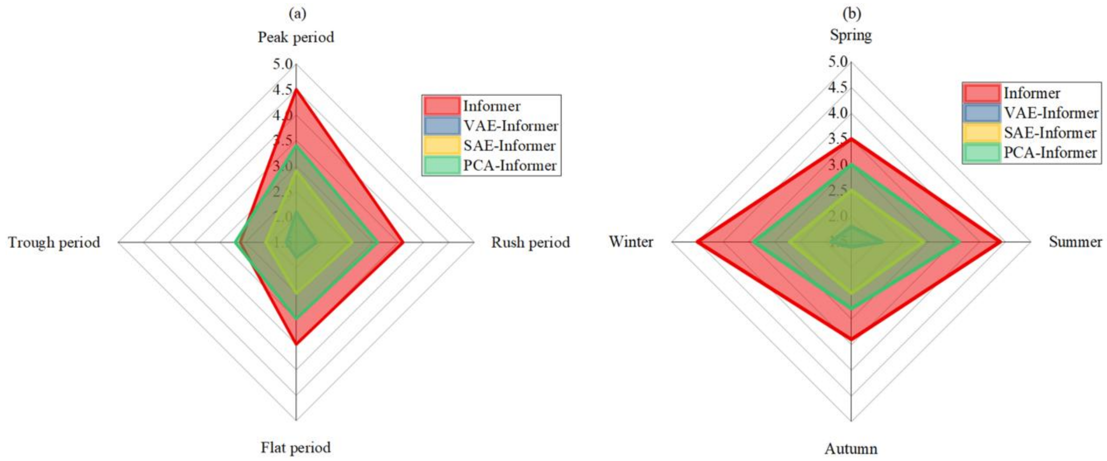

Figure 6. Stability analysis results.Figure 6 (a) Effect of time period; Figure 6 (b) Effect of season.

The load forecasting stability under load fluctuation scenarios is analyzed by using data from different time periods and seasons, and the load forecasting performance is measured by RMSE. The results are shown in Figure 6.

From the RMSE data of time period type, it can be seen that there are obvious differences in the load forecasting performance of each model during peak, high, flat and trough periods. Among them, VAE-Informer always performs best, with a peak RMSE of only 2.1, while Informer is as high as 4.5. This shows that using VAE to extract latent features can more accurately capture subtle changes in load fluctuations, especially during spikes and peak periods when the load fluctuates violently. The model relies on low-dimensional latent variables to effectively remove noise and redundant information, achieving high-fidelity reconstruction of load data. In contrast, although SAE-Informer and PCA-Informer also extract features of load fluctuations to a certain extent, since the former mainly relies on sparsity constraints and the latter relies on linear dimensionality reduction, it is difficult to fully reveal the nonlinear relationship in the data, resulting in a slightly higher prediction error than VAE-Informer. The RMSE of VAE-Informer in the flat and trough periods are 1.8 and 1.7 respectively, which further proves that it can maintain a low error even in the stable load stage. However, the Informer model shows a large error in all periods, which may be related to its failure to fully extract the intrinsic structure of load fluctuations. By comparing the RMSE of data in different time periods, the latent features extracted by VAE have significant advantages in capturing complex dynamic load changes and suppressing noise interference, thereby improving the stability and accuracy of the prediction and providing more refined data support for smart grid scheduling and energy storage management.

In terms of seasonal load forecasting performance, the RMSE indicators of each model also reflect the importance of extracting latent features. Figure 6(b) shows that the RMSE of VAE-Informer in spring, summer, autumn and winter are 1.8, 2.1, 1.6 and 1.9 respectively, which are 3.5, 4.4, 3.4 and 4.5 respectively for the Informer model, and its prediction error is significantly reduced. This advantage is mainly due to the fact that VAE makes the latent space distribution continuous and stable through KL divergence regularization in the latent variable modeling process, thereby capturing the nonlinear and complex fluctuation characteristics caused by seasonal changes in load data. Seasonal loads are greatly affected by climate, electricity usage habits and external environment, and load changes have obvious periodic and trend characteristics. VAE-Informer uses the low-dimensional representation extracted by the deep generative model to effectively integrate these seasonal information, allowing the model to maintain a high prediction accuracy when dealing with factors such as temperature changes, sunshine duration and holidays. In contrast, although SAE-Informer has made some improvements in latent feature extraction, it is limited by sparse constraints when dealing with variable seasonal characteristics, and its ability to capture complex seasonal changes is not as good as VAE. PCA-Informer uses linear dimensionality reduction, which makes it difficult to reflect the nonlinear dynamic characteristics in load data, and its prediction performance is also slightly inferior. The Informer model itself lacks an effective latent feature extraction mechanism, which is particularly evident in summer and winter when seasonal load fluctuations are drastic. The model that uses VAE to extract latent features and combines it with Informer shows a low RMSE in seasonal load forecasting, proving that it has significant advantages in accurately capturing seasonal load characteristics and robustness, which provides a scientific and effective technical means for load forecasting in complex and changing environments.

5. Conclusions

This study combines VAE with Informer to successfully extract the implicit characteristics of load fluctuations in the photovoltaic storage system and build a high-precision load forecasting model. Experimental results show that VAE-Informer is significantly superior to traditional Informer, SAE-Informer and PCA-Informer in terms of RMSE, MAPE and R^2, which fully demonstrates the advantages of VAE in capturing nonlinear dynamic characteristics of data and suppressing noise. This study enriches the theoretical system of deep generative models in time series data analysis and provides scientific and reliable decision support for smart grid scheduling, energy storage management and new energy applications.

However, this study still has some limitations. Future work will focus on integrating multi-source data, including but not limited to energy consumption patterns, to further enhance the performance of the model. Specific expansion directions include: first, expanding the scale of the data set to cover more diverse energy consumption patterns and operating scenarios; second, integrating multiple external variables, such as meteorological data, economic indicators, etc., to more comprehensively capture the factors affecting load fluctuations; third, exploring end-to-end joint training strategies, optimizing model architecture and parameter tuning processes, and improving the robustness and generalization ability of the model. Through these efforts, we hope to gradually promote this method to a wider range of practical engineering applications, and provide more comprehensive and accurate technical support for load forecasting and intelligent control.

Acknowledgment

None

Consent to Publish

The manuscript has neither been previously published nor is under consideration by any other journal. The authors have all approved the content of this article.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author, upon reasonable request.

Funding

None

Conflicts of Interest

The authors declare that they have no financial conflicts of interest.

References

[1] C. Liu, J.K. Zhuo, D.M. Zhao, S.Q. Li, J.S. Chen, J.X. Wang, et al. A review of research on flexible and safe operation of renewable energy microgrids using energy storage systems. Proceedings of the CSEE, 2020, 40(1), 1-18. DOI: 10.13334/j.0258-8013.pcsee.190212

[2] Z. Li, R. Zhang, H.X. Sun, W.D. Zhang, C.X. Mei. A review of key technologies for hydrogen production, storage and transportation based on multi-energy complementation of renewable energy. Transactions of the Chinese Society of Electrical Engineering, 2021, 36(3), 446-462. DOI: 10.19595/j.cnki.1000-6753.tces.200332

[3] L.Y. Jia, S.N. Yun, Z.N. Zhao, H.L. Li, S.Y. Wang, L. Yang. Research progress on application of neural network in short-term photovoltaic power generation prediction. Acta Energiae Solaris Sinica, 2022, 43(12), 88-97. DOI: 10.19912/j.0254-0096.tynxb.2021-0501

[4] C. Dong, Z. Wang, J.Y. Bai, J.D. Jiang, B. Wang, G.H. Liu. A review of ultra-short-term prediction methods for photovoltaic power generation. High Voltage Technology, 2023, 49(7), 2938-2951. DOI: 10.13336/j.1003-6520.hve.20220974

[5] A. Rafique, I. Ferreira, G. Abbas, A.C. Baptista. Recent advances and challenges toward application of fibers and textiles in integrated photovoltaic energy storage devices. Nano-Micro Letters, 2023, 15(1), 40-40. DOI: 10.1007/s40820-022-01008-y

[6] Y.X. Liang, H. Zhang, M.Q. Du, K. Sun. Parallel coordination control of multi-port DC-DC converter for stand-alone photovoltaic-energy storage systems. CPSS Transactions on Power Electronics and Applications, 2020, 5(3), 235-241. DOI:

10.24295/CPSSTPEA.2020.00020
[7] P. Manoharan, U. Subramaniam, T.S. Babu, S. Padmanaban, J.B. Holm-Nielsen, M. Mitolo. Improved perturb and observation maximum power point tracking technique for solar photovoltaic power generation systems. IEEE Systems Journal, 2020, 15(2), 3024-3035. DOI: 10.1109/JSYST.2020.3003255
[8] J.Y. Cho, S.M. Park, A R. Park, O.C. Lee, G. Nam, I.H. Ra. Application of photovoltaic systems for agriculture: A study on the relationship between power generation and farming for the improvement of photovoltaic applications in agriculture. Energies, 2020, 13(18), 4815-4815. DOI: 10.3390/en13184815
[9] S. Maragkaki, K. Savva, E. Stratakis. Advanced photonic processes for photovoltaic, energy storage, and environmental systems. Advanced Sustainable Systems, 2021, 5(3), 2000237-2000237. DOI: 10.1002/adsu.202000237
[10] M. Dada, P. Popoola. Recent advances in solar photovoltaic materials and systems for energy storage applications: a review. Beni-Suef University Journal of Basic and Applied Sciences, 2023, 12(1), 1-15. DOI: 10.1186/s43088-023-00405-5
[11] X. Dong, S. Deng, D. Wang. A short-term power load forecasting method based on k-means and SVM. Journal of Ambient Intelligence and Humanized Computing, 2022, 13(11), 5253-5267. DOI: 10.1007/s12652-021-03444-x
[12] W.L. Guo, L. Che, M. Shahidehpour, X. Wan. Machine-Learning based methods in short-term load forecasting. The Electricity Journal, 2021, 34(1), 106884-106884. DOI: 10.1016/j.tej.2020.106884
[13] H. Habbak, M. Mahmoud, K. Metwally, M.M. Fouda, M.I. Ibrahem. Load forecasting techniques and their applications in smart grids. Energies, 2023, 16(3), 1480-1480. DOI: 10.3390/en16031480
[14] L. Von Krannichfeldt, Y. Wang, G. Hug. Online ensemble learning for load forecasting. IEEE Transactions on Power Systems, 2020, 36(1), 545-548. DOI: 10.1109/TPWRS.2020.3036230
[15] B.M.S. Hasan, A.M. Abdulazeez. A review of principal component analysis algorithm for dimensionality reduction. Journal of Soft Computing and Data Mining, 2021, 2(1), 20-30. DOI: 10.30880/jscdm.2021.02.01.003
[16] D. Huang, F.W. Jiang, K.P. Li, G.S. Tong, G.F. Zhou. Scaled PCA: A new approach to dimension reduction. Management Science, 2022, 68(3), 1678-1695. DOI: 10.1287/mnsc.2021.4020
[17] J.Y. Wang, S.F. Zhang. Abnormal power consumption pattern detection based on linear discriminant analysis and density peak clustering. Automation of Electric Power Systems, 2022, 46(5), 87-98. DOI: 10. 7500/AEPS20210717003
[18] B. Xiao, Z.F. Xiao, Z. Jiang, X. Zhao, Z.F. Kan, X.S. Qi, et al. Spatial power load situation awareness based on denoising autoencoder, singular spectrum analysis and long short-term memory neural network. Proceedings of the CSEE, 2020, 41(14), 4858-4867. DOI: 10.13334/j.0258-8013.pcsee.201260
[19] D.M. Huang, X.X. Lin, A.D. Hu, J.Z. Sun. Deep embedding clustering method of daily load based on convolutional autoencoder. Electric Power Construction, 2021, 42(1), 132-138. DOI: 10.12204/j.issn.1000-7229.2021.01.015
[20] K. Liu, X. Liu, P.H. Zhang, Y. Xue, B. Li, S. Su, et al. Power theft identification method based on LSTM autoencoder with load spike characteristics. Automation of Electric Power Systems, 2023, 47(2), 96-104. DOI: 10.7500/AEPS20220517006
[21] C.Y. Ma, Y. Wang, F. Li, H.Y. Zhang, Y. Zhang, H.Y. Zhang, et al. Constructing Attention-LSTM-VAE Power Load Model Based on Multiple Features. Advances in Mathematical Physics, 2024, 2024(1), 1041791-1041791. DOI: 10.1155/2024/1041791
[22] M.A. Khan, B. Asad, T. Vaimann, A. Kallaste, R. Pomarnacki, V.K. Hyunh, et al. Improved Fault Classification and Localization in Power Transmission Networks Using VAE-Generated Synthetic Data and Machine Learning Algorithms. Machines, 2023, 11(10), 963-963. DOI: 10.3390/machines11100963
[23] Y.L. Yang, C.X. Li, Y.Y. Qin, Y.H. Wang, B. Fang, H. Shu, et al. Research on clustering of daily load curves based on convolutional VAE. Modern Electric Power, 2023, 40(5), 770-778. DOI: 10.19725/j.cnki.1007-2322.2022.0076
[24] X.Y. Wang, Y. Li, X.C. Dong, H.T. Wang, Y.Y. Sun. Multi-source-load scenario generation method for active distribution network based on VAE. Power System Technology, 2020, 45(8), 2962-2969. DOI: 10.13335/j.1000-3673.pst.2020.0903
[25] F.E. Bezerra, F. Grassi, C.G. Dias, F.H. Pereira. A PCA-based variable ranking and selection approach for electric energy load forecasting. International Journal of Energy Sector Management, 2022, 16(6), 1172-1191. DOI: 10.1108/IJESM-12-2019-0009
[26] H.W. Rotib, M.B. Nappu, Z. Tahir, A. Arief, M.Y. A. Shiddiq. Electric load forecasting for Internet of Things smart home using hybrid PCA and ARIMA algorithm. International Journal of Electrical and Electronic Engineering & Telecommunications, 2021, 10(6), 369-376. DOI: 10.18178/ijeetc.10.6.425-430
[27] A. L'Heureux, K. Grolinger, M.A.M. Capretz. Transformer-based model for electrical load forecasting. Energies, 2022, 15(14), 4993-4993. DOI: 10.3390/en15144993
[28] C. Wang, Y. Wang, Z.T. Ding, T. Zheng, J.Y. Hu, K.F. Zhang. A transformer-based method of multienergy load forecasting in integrated energy system. IEEE Transactions on Smart Grid, 2022, 13(4), 2703-2714. DOI: 10.1109/TSG.2022.3166600
[29] H.B. Xu, Q. Peng, Y.H. Wang, Z.W. Zhan. Power-load forecasting model based on informer and its application. Energies, 2023, 16(7), 3086-3086. DOI: 10.3390/en16073086
[30] Z.H. Yang, H.T. Wang, C. Liu. An Informer Model for Very Short-Term Power Load Forecasting. Energies, 2025, 18(5), 1150-1150. DOI: 10.3390/en18051150
[31] X.R. Fan, Y.C. Li. An Improved Informer Network for Short-Term Electric Load Forecasting. Recent Advances in Electrical & Electronic Engineering (Formerly Recent Patents on Electrical & Electronic Engineering), 2023, 16(5), 532-540. DOI: 10.2174/2352096516666230217113610
[32] X.H. Huang, A.H. Jiang. Wind power generation forecast based on multi-step informer network. Energies, 2022, 15(18), 6642-6642. DOI: 10.3390/en15186642
[33] S.K. Yu, Y.Q. Song. Comprehensive evaluation of pumped storage power station performance based on VAE-matter-element analysis. Modern Electric Power, 2022, 39(3), 295-301. DOI: 10.19725/j.cnki.1007-2322.2022.0045
[34] X.J. Gong, O. Tang, R.J. Zhu, W.L. Liao, L. Song. Data augmentation for electricity theft detection using conditional variational auto-encoder. Energies, 2020, 13(17), 4291-4291. DOI: 10.3390/en13174291