Advanced LSTM-GAN Model with Attention Mechanism for Carbon Footprint Assessment and Low-Carbon Strategy in Photovoltaic Energy

Wei-Hsien Huang[1], Kuo-Liang Yeh[2], Yu-Hsi Yuan[3]*

[1] Education Bureau, Taichung City Government, Taichung City, Taiwan.

[2] Department of Digital Media Design, Cardinal Tien Junior College of Health and Management, New Taipei City, Taiwan.

[3]*Department of Graphic Communication Arts, National Taiwan University of Arts, Taiwan.

*Corresponding Author: yuanyh@ntua.edu.tw

DOI:https://doi.org/10.30210/JMSO.202402.010

Submitted: Aug. 18, 2024 Accepted: Oct. 27, 2024

ABSTRACT

Accurate power load predictions are crucial for optimizing energy management and reducing carbon emissions in photovoltaic power generation. This study presents an advanced prediction model that combines Long Short-Term Memory (LSTM), Generative Adversarial Networks (GAN), and Attention Mechanisms to effectively forecast power loads, assess carbon footprints, and develop emission reduction strategies for photovoltaic enterprises. By incorporating the Gated Recurrent Unit (GRU) module, this study trained and evaluated the model across multiple datasets, demonstrating superior performance with over 95% accuracy, 93% recall, 92% F1 score, and 92% AUC metrics, outperforming other methods. The model also showed enhanced efficiency with fewer parameters, reduced floating-point operations, shorter inference times, and lower training times. On the Sun Dance dataset, it achieved a 46.2% reduction in parameters, 48.4% in FLOPs, 49.7% in inference time, and 37.6% in training time compared to existing methods. Ablation experiments confirmed high accuracy with Mean Absolute Error (MAE) below 19, Mean Absolute Percentage Error (MAPE) below 6%, Root Mean Square Error (RMSE) below 4, and Mean Squared Error (MSE) below 6. This research not only addresses time series prediction challenges but also offers potential applications across various fields, with future work aimed at improving dataset adaptability and model interpretability.

Keywords: New energy, Carbon emission reduction, Photovoltaic Load Forecasting, LSTM, GAN, Attention mechanism

1. Introduction

As global climate change issues continue to worsen, carbon emission reduction strategies and the exploration of renewable energy resources have become increasingly crucial. Photovoltaic (PV) generation [1,2], as a green and clean energy source, has garnered significant attention and widespread adoption worldwide. However, due to the inherent instability of PV generation and its sensitivity to weather conditions, accurate prediction of PV load is of paramount importance for effective grid management [3]. This involves optimizing energy distribution, ensuring grid stability, and maximizing economic benefits, making it a critical aspect of carbon footprint assessment and emission reduction strategy research conducted by energy and power companies.

1.1 Deep Learning in PV Load Forecasting

Deep learning has emerged as a promising solution for PV load forecasting, owing to its extensive application in the energy sector. The ability to process large volumes of data and reveal hidden patterns offers exciting opportunities for advancing PV generation technology. This literature review summarizes popular deep learning and machine learning models used in PV load forecasting and presents a new method to improve prediction accuracy. The commonly used models [4] were:

Recurrent Neural Networks (RNN): RNNs are powerful tools for working with sequential data and understanding time-related patterns. However, traditional RNNs can struggle with long-term dependencies due to issues like vanishing or exploding gradients [5,6].

Convolutional Neural Networks (CNN): While CNNs are mainly known for image processing, they can also be useful in PV load forecasting. By converting time series data into two-dimensional images, CNNs can extract features and make predictions effectively [7].

Support Vector Machines (SVM): SVMs are classic machine learning algorithms that excel in handling nonlinear relationships. They work by transforming data into a higher-dimensional space for classification or regression. However, SVMs can face difficulties when processing large-scale time series data [8,9].

Deep Attention Networks (DAN): These networks utilize attention mechanisms to help the model focus on key features or important time-related information, which can improve prediction accuracy [10].

Long Short-Term Memory (LSTM): LSTMs enhance traditional RNNs by incorporating gating units that help manage the gradient problems, allowing them to capture long-term dependencies better [11,12].

1.2 Proposed Method

This study aims to enhance the accuracy of PV load forecasting by utilizing the strengths of deep learning models. The proposed approach combines Long Short-Term Memory (LSTM) networks with attention mechanisms to effectively capture temporal relationships and key features in PV load data. The methodology is summarized as follows:

Historical Data Utilization: Researchers use historical PV load data as input sequences, training the LSTM networks to understand long-term dependencies within this data.

Incorporation of Attention Mechanisms: Attention mechanisms are incorporated to allow the model to automatically highlight important time steps or features, thus improving prediction accuracy.

Evaluation and Performance: The proposed method is evaluated through experiments to showcase its effectiveness and performance. The integration of attention mechanisms further enables the model to focus on significant temporal information and features, enhancing both prediction accuracy and stability.

1.3 Contribution and Impact

The contribution of this paper is the combination of LSTM-GAN and attention mechanisms to evaluate the carbon footprint of PV systems and explore strategies for reducing emissions. This approach effectively captures temporal dependencies in time series data and highlights the most relevant features, which enhances the accuracy of predicting PV energy generation and improves research on emission reduction strategies.

1.3.1 Synthetic Data Generation

LSTM-GAN for Synthetic Data: Researchers use LSTM-GAN to generate synthetic PV energy generation data. By leveraging a generative model based on historical data, various energy generation scenarios can be simulated, offering a flexible and controlled method for assessing and comparing different emission reduction strategies.

1.3.2 Evaluation of Emission Reduction Strategies

Through the evaluation and analysis of experimental results, this paper examines the effectiveness of various emission reduction strategies in lowering carbon emissions from PV systems. It provides valuable guidance and evidence for developing innovative solutions in the field of renewable energy.

2. Related Work

The three most relevant directions or models related to the topic of Photovoltaic carbon footprint assessment and emission reduction strategy are transformer-based models, ensemble methods, and hybrid modeling techniques.

Transformer models have become increasingly popular in various areas, such as natural language processing and time series forecasting. When it comes to PV load forecasting, transformer-based models—including the Transformer architecture and its variants—are effective at capturing long-range dependencies and modeling temporal relationships [13,14]. These models use self-attention mechanisms to assess the importance of different time steps or features, which helps achieve accurate predictions in PV load forecasting.

Additionally, ensemble methods involve combining multiple forecasting models to enhance prediction accuracy and stability. In PV load forecasting, ensemble methods can be applied by training and merging various models, such as different types of neural networks (like LSTM, CNN, and Transformer), support vector machines (SVM), or other machine learning algorithms. By utilizing the diversity of these individual models, ensemble methods can mitigate the effects of model biases and improve overall forecasting performance [15,16].

Moreover, hybrid models integrate various modeling techniques to take advantage of their strengths. In the realm of PV load forecasting, hybrid models can merge deep learning approaches with traditional statistical or physical models. By combining data-driven methods with domain knowledge, hybrid models can deliver more accurate and interpretable predictions for PV load forecasting [17,18].

2.1 Transformer model

Transformer models have demonstrated impressive results across various fields, including natural language processing, computer vision, and time series forecasting. In PV load forecasting, Transformer-based models are used to capture long-term dependencies and enhance prediction accuracy [13,14]. These models effectively represent temporal relationships in PV load data by encoding time steps as input embeddings. Each embedding contains information about its position in the sequence, helping the model learn the patterns and dependencies over time.

Transformers employ self-attention mechanisms, which enable the model to assess the importance of different time steps or features when making predictions. This attention mechanism allows the model to concentrate on relevant information and understand complex relationships within the PV load data. Additionally, Transformers use multi-head attention, where multiple attention heads capture various types of patterns or dependencies. This capability helps the model recognize both local and global dependencies, improving its forecasting performance.

One of the key advantages of Transformers is their ability to capture long-range dependencies in PV load data, which is essential for accurate forecasting. Unlike recurrent models, Transformers avoid issues like vanishing or exploding gradients, making them more effective at modeling long-term dependencies. They also process input sequences in parallel, which significantly boosts efficiency for large-scale PV load forecasting tasks. This parallel processing allows for quicker training and inference times compared to sequential models like LSTMs. Furthermore, Transformers can manage sequences of varying lengths without losing performance, making them well-suited for PV load forecasting, where input sequence lengths may differ based on the forecasting horizon.

However, this models often require a large amount of training data to generalize well and achieve optimal performance. Acquiring and labeling a sufficient amount of high-quality PV load data can be challenging, especially in some regions or for specific time periods. This model is known to be highly complex and often considered as black-box models. Interpreting the internal workings and understanding the reasoning behind the model's predictions can be difficult. This lack of interpretability may limit their adoption in certain applications where interpretability is crucial. Due to their large number of parameters and intensive computations, Transformer-based models can be computationally expensive, especially for training on large-scale datasets. This may require significant computational resources or specialized hardware accelerators to train and deploy these models effectively.

In summary, transformer models provide a robust method for PV load forecasting by effectively capturing long-term dependencies and utilizing self-attention mechanisms. They are particularly good at identifying complex patterns and relationships. However, their performance depends on having enough training data and adequate computational resources. While the interpretability of Transformer-based models can be challenging, their scalability and ability to process data in parallel make them well-suited for large-scale PV load forecasting tasks.

2.2 Ensemble method

Ensemble methods have been widely used in PV load forecasting to improve prediction accuracy and stability. These methods involve combining multiple forecasting models to make collective predictions. Ensemble methods utilize diverse forecasting models, such as different variants of neural networks, SVM, or other machine learning algorithms. Each model may have its strengths and weaknesses, capturing different aspects of the PV load data. By combining these models, ensemble methods can leverage their complementary abilities and improve overall forecasting performance. It employs various techniques to combine the predictions of individual models [15,16]. These techniques aim to reduce individual model biases and enhance the collective predictive power of the ensemble. It requires training and validating individual models before combining them. The training process involves optimizing each model's parameters using historical PV load data, and the validation phase assesses their performance on validation data. The ensemble is constructed based on the performance of individual models, ensuring that the most accurate and reliable models contribute more to the final predictions.

Using ensemble methods can greatly improve the accuracy of PV load forecasting compared to relying on individual models. By combining several models, these methods reduce the errors and biases from single models, leading to more reliable and precise predictions. They are generally more stable than single models because they are less affected by changes in input data or model parameters. This stability is particularly useful in PV load forecasting, where data can show unexpected patterns or anomalies.

Ensemble methods also offer a way to understand the uncertainty of predictions. By considering the variety of individual models, they can estimate prediction or confidence intervals. This helps decision-makers evaluate the reliability of forecasts and make informed choices. However, ensemble methods can be resource-intensive, especially when using many models or complex techniques. Training and evaluating multiple models require extra computational resources and time.

Moreover, ensemble methods add complexity to the forecasting process, including choosing the right models, combination techniques, and tuning hyperparameters. This complexity may require expertise and careful experimentation to achieve the best performance. While they improve accuracy, ensemble methods often compromise on interpretability. The combined predictions might not clearly explain the factors driving changes in PV load, which could limit their usefulness when decision-making needs a deep understanding of the forecasting process.

Above all, ensemble methods offer a powerful approach to enhance PV load forecasting by combining multiple forecasting models. They improve prediction accuracy, increase stability, and provide measures of uncertainty. However, ensemble methods come with computational complexity, increased model complexity, and potential challenges in interpretability. Despite these limitations, ensemble methods have been proven effective in improving PV load forecasting performance and are widely adopted in the field.

2.3 Hybrid model

Hybrid models combine various modeling techniques and have been effectively used in PV load forecasting to take advantage of different strengths. These models merge data-driven methods, like deep learning, with knowledge-driven approaches, such as traditional statistical or physical models. By incorporating domain knowledge, these models enhance forecasting accuracy.

In the context of PV load forecasting, domain knowledge can come from physical models that account for factors like solar radiation, weather conditions, and the characteristics of the PV system. Integrating this knowledge into the forecasting process allows hybrid models to better understand the underlying mechanisms, leading to improved predictions.

These models blend data-driven techniques, such as deep learning, with knowledge-driven methods like autoregressive models or exponential smoothing [17,18]. Data-driven models excel at capturing complex patterns and relationships in PV load data, while knowledge-driven models add interpretability and robustness to the forecasts.

Hybrid modeling often includes feature engineering, where relevant features derived from domain knowledge are integrated into the data-driven models. These features can encompass weather variables, solar radiation data, calendar information, and transformations of historical load data. Feature engineering helps hybrid models incorporate essential contextual information, further enhancing forecasting accuracy. These models leverage domain knowledge-driven approaches, making them more interpretable compared to fully data-driven models. The inclusion of physical models or traditional statistical models allows for a better understanding of the factors influencing PV load variations and facilitates decision-making. It often more robust to data limitations or outliers than purely data-driven models. The incorporation of domain knowledge helps handle situations where data may be sparse or incomplete, ensuring reliable and accurate predictions even in challenging scenarios. By combining data-driven and knowledge-driven approaches, hybrid models can leverage the strengths of both paradigms. The data-driven models capture complex patterns, while the knowledge-driven models provide insights into the underlying mechanisms. This combination leads to improved forecasting accuracy.

However, hybrid models can be more complex than individual models due to the integration of different approaches and the need for feature engineering. This complexity may require additional computational resources and expertise for model development and maintenance. It rely on domain knowledge, which may require expertise in PV systems, meteorology, or energy forecasting. The availability and accuracy of domain knowledge can significantly impact the performance of hybrid models, and acquiring such knowledge can be time-consuming and costly. Integrating data-driven and knowledge-driven components in hybrid models can be challenging. Aligning different models, handling inconsistencies, and optimizing the combination of forecasts require careful consideration and experimentation.

Thus, hybrid models offer a powerful approach for PV load forecasting by combining data-driven and knowledge-driven approaches. They provide improved interpretability, robustness, and accuracy. However, hybrid models come with increased complexity, domain knowledge requirements, and challenges in model integration. Despite these limitations, hybrid models have demonstrated success in PV load forecasting by incorporating domain knowledge and leveraging the strengths of different modeling techniques.

3. Methodology

3.1 Network outline

PV load shows temporal patterns and dependencies influenced by factors like the time of day, seasonality, and weather conditions. The goal of various methods is to capture these dependencies for accurate forecasting of future PV load. This process often involves feature engineering, where important features are derived from historical data and external factors. These features may include time-related elements, weather variables, calendar information, and previous load values transformed into appropriate formats.

The method proposed in this paper aims to enhance the accuracy of carbon footprint assessments for PV systems and developing emission reduction strategies by combining LSTM-GAN with Attention Mechanisms. Figure 1 illustrates the overall framework of the proposed model. This version simplifies the language while keeping the original meaning and citations intact.

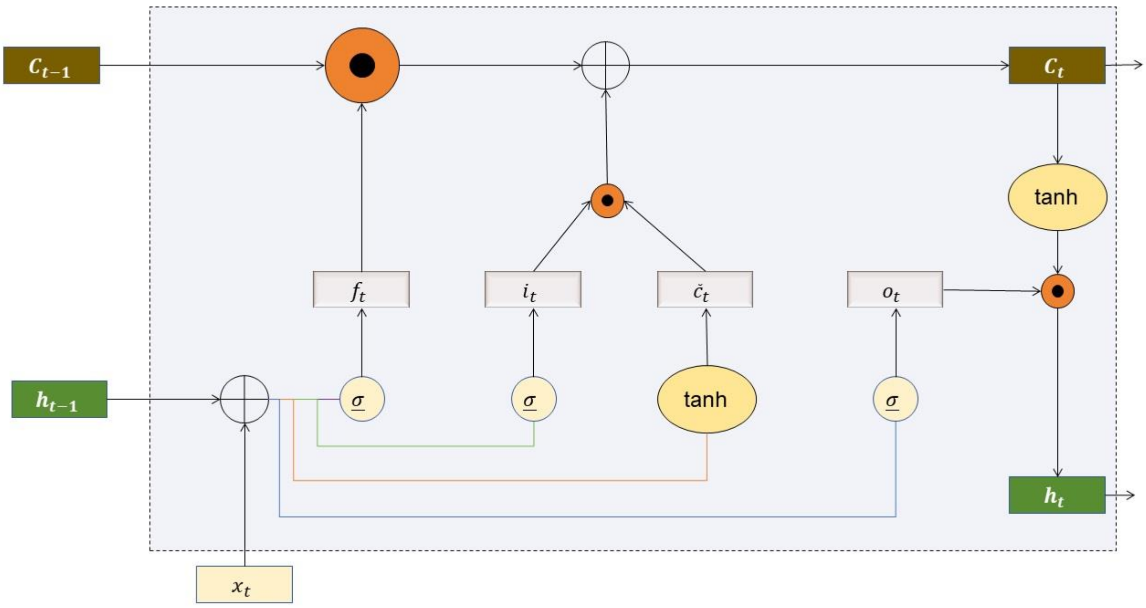

Figure 1. The overall framework diagram of the proposed model.
Source: By authors.

The overview of method's principles and detailed implementation process as following.

Data Collection and Preprocessing: First, historical data related to PV systems is gathered. This includes data on PV generation, weather conditions, energy consumption, and more. The collected data undergoes preprocessing, which involves cleaning, normalization, and time-series processing to prepare it for model training and prediction.

LSTM-GAN Model Training: The LSTM-GAN model is used to generate synthetic PV generation data. The LSTM part processes time-series data and predicts future generation based on patterns learned from historical data. Meanwhile, the GAN component creates synthetic data to enhance diversity and coverage. Through an adversarial process between the generator and discriminator, the model gradually improves the accuracy and realism of the generated data.

Introduction of Attention Mechanisms: Attention mechanisms are added to help the model focus on important time points and features. These mechanisms learn to highlight key factors related to carbon footprint assessment and emission reduction strategies. By using attention mechanisms, the model can make more precise predictions of PV generation and provide accurate assessments of carbon footprints.

Carbon Footprint Assessment and Emission Reduction Strategy Research: Using the generated synthetic data and insights from the attention mechanisms, the carbon footprint of the PV system is evaluated. This involves analyzing PV generation and energy consumption data, along with carbon emission models and environmental indicators, to achieve accurate carbon footprint assessments. Based on these assessments, targeted strategies can be developed to reduce the environmental impact of PV systems.

Model Evaluation and Optimization: The trained model is evaluated using metrics such as prediction accuracy and the precision of carbon footprint assessments. Based on the evaluation results, the model is optimized and adjusted to enhance its performance further.

Through this overall process, the proposed method leverages the strengths of LSTM-GAN combined with attention mechanisms to achieve accurate assessment of PV carbon footprints and effective research on emission reduction strategies. This approach provides valuable support and guidance for the sustainable development of the PV industry.

3.2 Long short-term memory (LSTM)

LSTM, or Long Short-Term Memory, is a specific type of Recurrent Neural Network (RNN) known for its effectiveness in modeling sequential data and capturing long-term dependencies. In PV load forecasting, LSTM models are essential for identifying temporal patterns and dependencies in historical PV load data. RNNs are designed to handle sequential data by maintaining a hidden state that carries information from previous time steps. This hidden state enables the network to learn from the context of earlier inputs [19,20].

However, traditional RNNs face challenges, particularly the "vanishing gradient" problem. This issue occurs when gradients become very small as they move through time, making it hard to capture long-term dependencies. Additionally, traditional RNNs may struggle to retain important information while discarding irrelevant details over long sequences. Figure 2 illustrates the diagram of the proposed model.

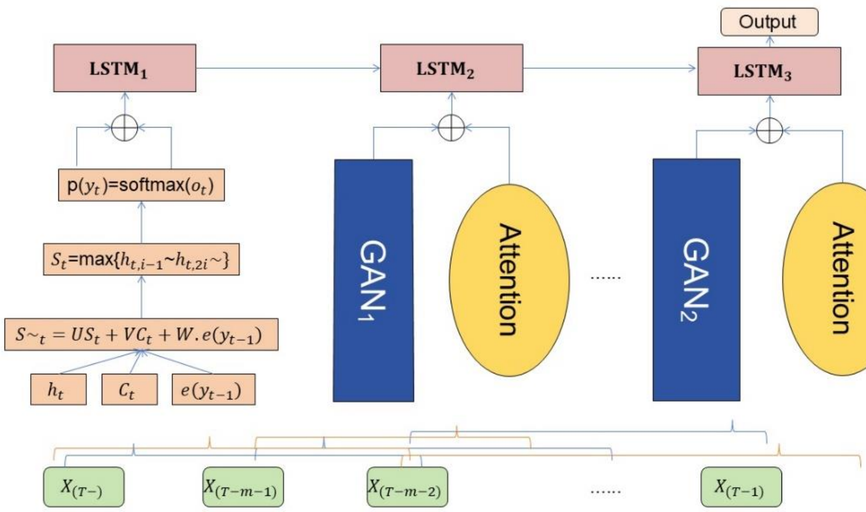

Figure 2. The schematic diagram of the principle of LSTM.
Source: By authors.

The memory cell state of an LSTM is key for storing and passing information over time. It can choose to keep or forget information using gating mechanisms. The input gate controls how much new information from the current input is added to the cell state. The forget gate decides how much information should be discarded, helping the model ignore irrelevant details. The output gate regulates how much information is released from the cell state based on the current input and the hidden state.

This model is effective at capturing long-term dependencies in sequential data. In PV load forecasting, historical PV load data is viewed as a time series, allowing LSTM models to learn patterns and dependencies over time. By analyzing past load values, LSTM models can make accurate predictions by identifying complex temporal relationships. They automatically extract relevant features from historical data, learning to recognize important patterns necessary for precise forecasting.

This capability for feature extraction removes the need for manual feature engineering, making LSTM models particularly well-suited for complex and nonlinear data like PV load. Additionally, LSTM models can handle variable-length sequences, which is beneficial in PV load forecasting since historical data may vary in length due to missing data or irregular sampling intervals. The LSTM architecture adapts to different sequence lengths, providing flexibility and robustness for real-world data.

The memory cells and gating mechanisms in LSTM models enable the network to capture and retain important information over long sequences. This is essential in PV load forecasting, as load variations can be affected by factors that span multiple time steps, such as daily or seasonal patterns. LSTM models effectively learn and utilize these long-term dependencies for accurate predictions.

$$\text{Input Gate:} \quad i_t = \sigma\left(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i\right) \tag{1}$$

$$\text{Forget Gate:} \quad f_t = \sigma\left(W_{xf} x_t + W_{hf} h_{t-1} + W_{cf} c_{t-1} + b_f\right) \tag{2}$$

$$\text{Output Gate:} \quad o_t = \sigma\left(W_{xo} x_t + W_{ho} h_{t-1} + W_{co} c_t + b_o\right) \tag{3}$$

$$\text{Cell State:} \quad c_t = f_t \odot c_{t-1} + i_t \odot \tanh\left(W_{xc} x_t + W_{hc} h_{t-1} + b_c\right) \tag{4}$$

$$\text{Hidden State:} \quad h_t = o_t \odot \tanh\left(c_t\right) \tag{5}$$

In these equations, x_t represents the input at time step t, h_t represents the hidden state at time step t, and c_t represents the cell state at time step t. The input gate, forget gate, and output gate activations at time step t are denoted as i_t, f_t, and o_t respectively. The sigmoid activation function is indicated by σ, and element-wise multiplication is represented by the symbol $\odot$.

The weight matrices W_{xi}, W_{hi}, W_{ci}, W_{xf}, W_{hf}, W_{cf}, W_{xo}, W_{ho}, W_{co}, W_{xc}, and W_{hc} are used in these calculations. While b_i, b_f, b_o, and b_c are the corresponding bias vectors. These equations outline the computations that occur within an LSTM cell, where the gates manage the flow of information through the cell state, and the hidden state is generated based on the updated cell state.

LSTM overcomes the limitations of traditional RNNs by incorporating memory cells and gating mechanisms. Each LSTM unit consists of a cell state that serves as memory, along with three types of gates: the input gate, forget gate, and output gate. These gates manage how information flows into, out of, and within the LSTM unit.

This is particularly important in PV load forecasting because LSTMs can effectively capture temporal patterns, manage variable-length sequences, and learn long-term dependencies. By utilizing memory cells and gating mechanisms, LSTM models are highly effective at modeling sequential data and making accurate predictions for future PV load values.

3.3 Generative adversarial network (GAN)

GAN, or Generative Adversarial Network, is a deep learning framework that consists of two neural networks: a generator and a discriminator. These networks work together in a competitive manner. GANs are widely used across various fields to create synthetic data that mimics real data distributions.

In PV load forecasting, GAN models can be utilized to generate synthetic PV load data. This synthetic data can help with data augmentation, enhance model generalization, and tackle issues related to data scarcity. The generator is responsible for creating these synthetic data samples. It takes random noise as input and transforms it into data that resembles the actual data distribution. For PV load forecasting, the generator network learns to produce synthetic PV load patterns [21,22]. Figure 3 illustrates the diagram of the proposed model.

The discriminator network functions as a binary classifier, designed to differentiate between real and synthetic data samples. It receives either real data from the training set or synthetic data produced by the generator. The discriminator learns to identify whether the input is real or fake, providing valuable feedback to the generator.

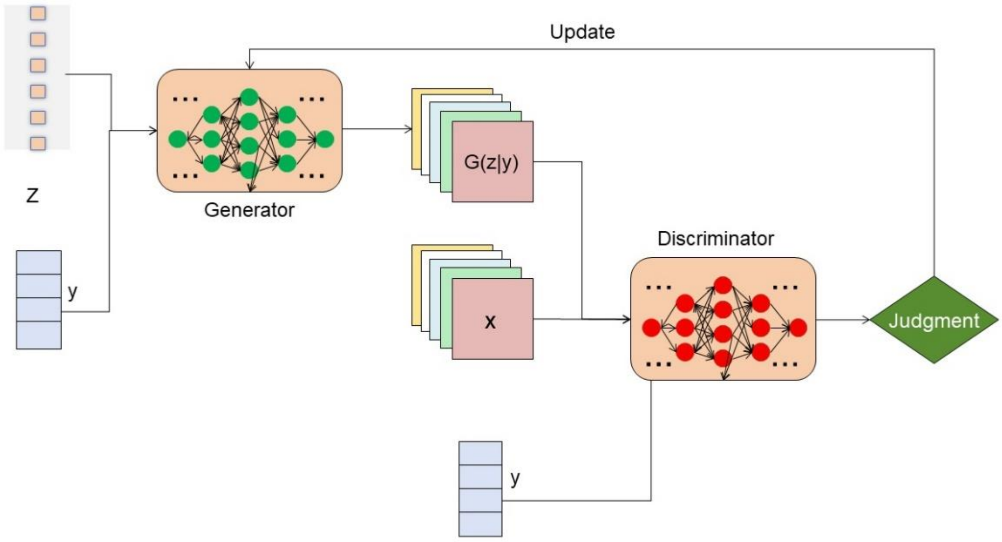

Figure 3. The schematic diagram of the principle of GAN.
Source: By authors.

Both the generator and discriminator are trained in a competitive way. The generator's goal is to create synthetic samples that are indistinguishable from real data, while the discriminator's goal is to accurately classify the samples. This adversarial training creates a competition between the two networks, pushing the generator to enhance its ability to produce realistic data.

The GAN training process focuses on optimizing both the generator and discriminator using specific loss functions. The generator aims to reduce the discriminator's ability to identify synthetic samples, while the discriminator strives to improve its classification accuracy. This optimization is typically accomplished through backpropagation and gradient descent.

The GAN model can create synthetic PV load data that mimics real load patterns. This synthetic data can be used to enhance the existing training dataset, increasing both its diversity and size. By augmenting the training data, GANs help improve the forecasting model's ability to generalize, allowing it to capture a broader range of load variations and boosting overall prediction accuracy.

In cases where PV load data is scarce or limited, GAN models can generate additional synthetic data. This is especially beneficial when historical load data is insufficient for training accurate forecasting models. By taking advantage of GANs' generative capabilities, more data can be produced to enhance the model's training process and improve its performance.

GAN models can also learn complex temporal patterns and dependencies within PV load data. The generator network learns to produce synthetic load patterns that closely resemble real load data, capturing detailed variations and characteristics. This enables the forecasting model to learn from the synthetic data and make accurate predictions for future PV load values.

Additionally, GANs can help address data imbalances or biases in the training data. By generating synthetic samples that represent underrepresented load patterns, GANs mitigate the effects of data imbalances and enhance the forecasting model's ability to handle diverse load scenarios. The formula of GAN is as follows.

$$\text{Generator: } G(z) = \hat{x} \tag{6}$$

$$\text{Discriminator: } D(x) = d \tag{7}$$

$$\text{Generator Loss: } \mathcal{L}G(G, D) = -\mathbb{E}z \sim p_z(z)[\log D(G(z))] \tag{8}$$

$$\text{Discriminator Loss: } \mathcal{L}D(G, D) = -\mathbb{E}x \sim p_{\text{data}}(x)[\log D(x)] - \mathbb{E}_{z \sim p_z(z)}[\log(1 - D(G(z)))] \tag{9}$$

In this context, G represents the generator network, which takes a latent vector z as input to produce a synthetic sample. D stands for the discriminator network, which receives a sample x (either real or synthetic) and outputs a prediction d, indicating the likelihood that the sample is real. The latent vector z is drawn from a prior distribution $P_z(z)$, while x represents a real sample from the dataset, and $\hat{x}$ denotes a generated (fake) sample from the generator.

The generator loss LG encourages the generator to create samples that are more likely to be classified as real by the discriminator. In contrast, the discriminator loss LD motivates the discriminator to accurately classify real and fake samples. These equations outline the fundamental framework of the GAN, where the generator's goal is to produce realistic samples to deceive the discriminator, while the discriminator's goal is to differentiate between real and synthetic samples. The generator loss is minimized by adjusting the generator's parameters, and the discriminator loss is minimized by updating the discriminator's parameters, leading to a competitive training process.

GAN models offer a robust framework for generating synthetic PV load data, effectively addressing data scarcity, enhancing data augmentation, and capturing complex load patterns. By utilizing the adversarial training process between the generator and discriminator networks, GANs improve the forecasting model's ability to generalize and enable accurate predictions for future PV load values.

3.4 Attention mechanism

The Attention Mechanism is a technique used to process sequential data, allowing the model to focus on specific parts of the input sequence. It is widely applied in areas such as natural language processing and machine translation, and it also plays an important role in PV load forecasting [23,24,25]. Figure 4 illustrates the diagram of the proposed model.

The main idea behind the Attention Mechanism is to assign different weights to various elements in the input sequence. This enables the model to concentrate more on the parts that are relevant to the task at hand. By computing contextual relevance, the model can prioritize the input sequence based on importance.

In the Attention Mechanism, the input sequence is divided into three components: query, key, and value. The query represents the model's state at the current time step, while the key and value represent different positions in the input sequence. By calculating the similarity between the query and the key, the researchers can determine weights for each key. These weights are then applied to the corresponding values to create a context vector.

Calculating similarity is a crucial step in the Attention Mechanism. Common methods include computing the dot product or inner product between the query and the key, or using attention weights learned through neural networks. This similarity calculation helps measure the relationship between the query and the key, which determines the importance of each key.

By multiplying the attention weights with the values and summing them, the researchers obtain a context vector. This vector contains information from various positions in the input sequence, where the weight assigned to each position reflects its contribution to the current query. The model can then use this context vector for tasks such as prediction, classification, or generation.

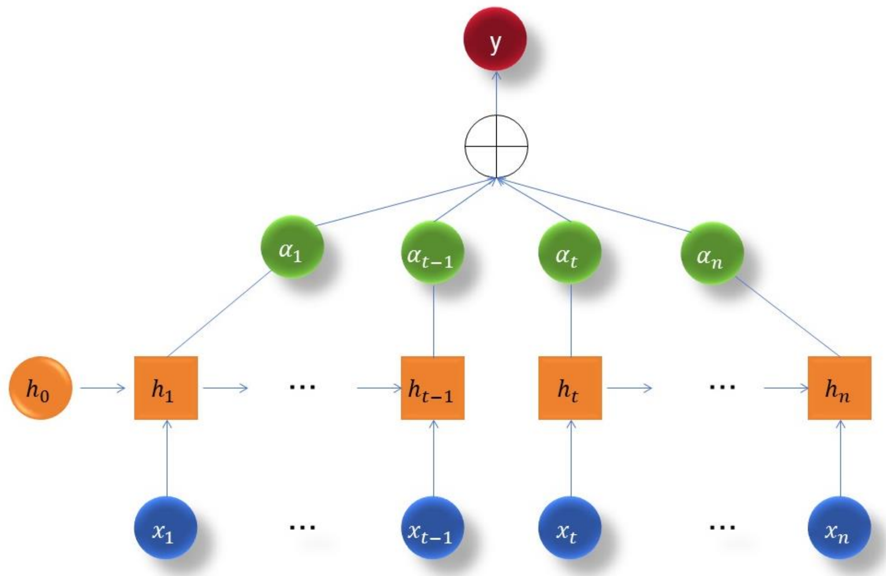

Figure 4. The schematic diagram of the principle of Attention Mechanism.
Source: By authors.

In PV load forecasting, the input sequence usually consists of a series of historical load values. The Attention Mechanism is capable of learning the relationships between different time steps in this sequence. By calculating the similarity between the query and the key, the Attention Mechanism can assign weights to each time step based on its importance, effectively capturing key patterns and dependencies.

PV load data can vary in length due to missing data or irregular sampling intervals. The Attention Mechanism can adaptively calculate attention weights based on the actual length of the input sequence, allowing it to handle variable-length sequences flexibly.

This mechanism enables the model to focus on the most important parts of the input sequence. In the context of PV load forecasting, the model can prioritize relevant portions of the historical load sequence for the current time step, rather than treating all historical values equally. This approach enhances the model's ability to identify crucial time steps and improves prediction accuracy.

Long-term dependencies in PV load forecasting may involve distant time steps. Traditional RNNs often struggle to effectively capture these long-term dependencies. However, the Attention

Mechanism addresses this issue by allowing the model to focus on relevant historical data, even if it is far back in the sequence. This capability makes it better suited for managing long-term dependencies. The formula for the Attention Mechanism is as follows.

$$\text{Attention Scores: } \operatorname{score}(q,k)=\frac{q\cdot k^{\top}}{\sqrt{d_k}} \tag{10}$$

$$\text{Attention Weights: } \operatorname{softmax}(q,k,v)=\operatorname{softmax}(\operatorname{score}(q,k)) \tag{11}$$

$$\text{Attention Output: } \operatorname{Attention}(q,k,v)=\operatorname{softmax}(q,k,v)\cdot v \tag{12}$$

In this context, q represents the query vector, typically derived from the current hidden state of the decoder. k denotes the key vectors, which are usually obtained from the hidden states of either the encoder or the previous time steps of the decoder. v represents the value vectors, containing information relevant to the current query. The term score (q,k) refers to the attention scores, which measure how relevant or similar the query vector q is to the key vectors k. The dot product is a common method for calculating these scores. d_k indicates the dimensionality of the key vectors and is often used for normalizing the scores. The function Softmax (q,k,v) calculates the attention weights by applying the softmax function to the attention scores. This ensures that the weights add up to 1, reflecting the importance assigned to each value vector based on the attention scores. The output of the attention mechanism, Attention (q,k,v), is obtained by multiplying the attention weights with the value vectors element-wise. This output captures the relevant information from the value vectors according to the assigned attention weights.

The Attention Mechanism is frequently utilized in sequence-to-sequence models, like Transformers, to focus on different parts of the input sequence when generating each output element. It allows the model to selectively attend to important information, enhancing its ability to manage long-range dependencies. This mechanism helps the model concentrate on specific sections of the input sequence, learn dependencies, handle variable-length sequences, emphasize crucial information, and address long-term dependencies.

In PV load forecasting, applying the Attention Mechanism can significantly improve the model's ability to analyze historical load sequences, enhance prediction accuracy, and adapt to load data of varying lengths and complexities.

4. Experiment

4.1 Datasets

This article examines four selected datasets: the Sun Dance dataset, Pecan Street dataset, NSRDB dataset, and NREL dataset.

Sun Dance Dataset: The Sun Dance dataset is a meteorological collection that includes weather observation data from the Sun Dance area in Arizona, USA. It provides time-series data on various weather variables, such as temperature, humidity, wind speed, and wind direction, typically sampled at hourly or daily intervals. This dataset is frequently used in meteorological analysis, climate modeling, and renewable energy research.

Pecan Street Dataset: The Pecan Street dataset is an energy dataset gathered by the Pecan Street project in Austin, Texas, USA. It includes a range of energy-related data from residential and commercial buildings, such as electricity consumption, water usage, and weather information. Electricity data is often collected at minute or second intervals, while water data records usage patterns. This dataset is valuable for analyzing energy consumption, modeling electricity usage behavior, and optimizing energy systems, making it essential for sustainable energy research and management.

NSRDB Dataset: The NSRDB dataset is a solar radiation dataset developed and maintained by the National Renewable Energy Laboratory (NREL) in the United States. It offers solar radiation data for various locations across the country, including solar irradiance, solar insolation, and potential solar energy generation. Data is typically sampled at minute or hourly intervals. The NSRDB dataset is widely used for solar resource assessment, forecasting solar power generation, and planning solar farms, providing critical information for solar energy research and applications.

NREL Dataset: The NREL dataset encompasses multiple renewable energy-related datasets collected and provided by the National Renewable Energy Laboratory in the United States. It covers various areas of renewable energy, including solar, wind, and biomass energy. This dataset includes solar radiation data, wind resource data, and energy system performance data, supporting research and development in renewable energy. The NREL dataset is commonly applied in renewable energy resource assessment, energy system planning, and evaluating renewable energy technologies, offering vital data support for the renewable energy sector.

4.2 Experimental details

This study outlines an experimental design that includes a comparison of various metrics and an ablation study. Below is a detailed overview of the experimental procedure. The goal of the experiment is to evaluate the performance of different models by comparing metrics and conducting ablation experiments. Key aspects include training and inference time, parameter count, floating-point operations, accuracy, AUC, recall, and F1 score. The steps of the experiment are as follows:

Data Preparation: First, select suitable data from the Sun Dance dataset, Pecan Street dataset, NSRDB dataset, and NREL dataset for the experiments. Next, preprocess the data through cleaning, normalization, and time series processing.

Model Selection: Choose a variety of models, such as LSTM, GRU, and Transformer, based on the experimental goals. Additionally, determine appropriate performance metrics like training and inference time, parameter count, and floating-point operations.

Model Training: Split the dataset into training and testing sets in a designated ratio. Train each selected model using the training set to optimize parameters and fit the model. During this process, record the training time and adjust hyperparameters as necessary.

Model Evaluation: Assess the trained models using the testing set and calculate metrics such as accuracy, AUC, recall, and F1 score. Also, record the inference time and compute the parameter count and floating-point operations as required.

Metrics Comparison Experiment: Compare the performance of the different models regarding training time, inference time, parameter count, and floating-point operations. Analyze and interpret the experimental results to draw conclusions.

Ablation Experiment: Identify specific components of the models for ablation experiments, such as removing attention mechanisms or certain layers, based on the experimental design. Train and evaluate the modified models, then compare the performance changes. Finally, analyze and interpret the results to draw conclusions.

Next, combine the results from the metrics comparison and ablation experiments. Analyze the strengths and weaknesses of each model based on different performance metrics and draw conclusions. Additionally, discuss potential directions for optimizing and improving the models. Below is the formula for the comparison indicators.

1. Training Time (in seconds)

$$\text{Training Time} = \text{Total training time} \tag{13}$$

2. Inference Time (in milliseconds)

$$\text{Inference Time} = \frac{\text{Total inference time}}{\text{Number of inference samples}} \tag{14}$$

3. Parameters (in millions)

$$\text{Parameters} = \frac{\text{Number of parameters in the model}}{10^6} \tag{15}$$

4. Flops (in billions)

$$\text{Flops} = \frac{\text{Total floating-point operations in the model}}{10^9} \tag{16}$$

5. Accuracy

$$\text{Accuracy} = \frac{\text{Number of correctly predicted samples}}{\text{Total number of samples}} \tag{17}$$

6. AUC (Area Under the Curve): The area under the Receiver Operating Characteristic curve, used to evaluate the performance of binary classification models.

7. Recall

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} \tag{18}$$

8. F1 Score combines precision and recall into a single evaluation metric.

$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{19}$$

where, Precision is defined as:

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}} \tag{20}$$

For the experiment, the researchers will use popular deep learning frameworks for model training. Hyperparameters such as learning rate, batch size, optimizer, and the number of training epochs will be set according to the experimental needs. To minimize randomness, the researchers will use cross-validation or average the results from multiple experiments. During the experiment, the researchers will also record the training time, inference time, parameter count, and floating-point operations for each model. For the ablation experiments, the researchers will choose suitable ablation methods based on the experimental requirements and document any changes in performance metrics. The algorithm code for this method is provided below.

Algorithm 1: LSTM-GAN Training Process

```
Input: Datasets: "Sun Dance Dataset", "Pecan dataset", "NSRDB dataset", "NREL dataset"
Output: Trained LSTM-GAN model
Initialization: Initialize LSTM-GAN model
for each epoch in total epochs do
    for each batch in dataset do
        Real_samples = randomly_select_real_samples(batch_size);
        Generate_fake_samples = generator_network(noise);
        // Train the discriminator
        Update discriminator weights using gradient descent optimization;
        discriminator_loss_real = compute_discriminator_loss(discriminator_network(Real_samples), real_labels);
        discriminator_loss fake = compute_discriminator_loss(discriminator_network(Generate_fake_samples),
        fake_labels);
        discriminator_loss = discriminator_loss_real + discriminator_loss_fake;
        update_discriminator_weights(discriminator_loss);
        // Train the generator
        Update generator weights using gradient descent optimization;
        generator_loss = compute_generator_loss(discriminator_network(Generate_fake_samples), real_labels);
        update_generator_weights(generator_loss);
        // Train the LSTM network
        Update LSTM network weights using gradient descent optimization;
        lstm_loss = compute_lstm_loss(lstm_network(Real_samples), Real_samples);
        update_lstm_weights(lstm_loss);
        // Update attention mechanism
        Update attention mechanism weights using gradient descent optimization;
        attention_loss = compute_attention_loss(attention_mechanism(Real_samples), Real_samples);
        update_attention_weights(attention_loss);
    end
    // Calculate evaluation metrics
    evaluation_results $=$ evaluate_model(LSTM-GAN, evaluation_metrics);
    print ('EEpoch:", epoch, "Evaluation Results:", evaluation_results);
    // Early stopping condition
    if early_stopping condition met then
        break;
    end
end
```

4.3 Results

Table 1 and Figure 5 show the experimental results comparing the performance of various methods on the Sun Dance dataset [26] and the Pecan Street dataset [27]. The researchers assessed several metrics, including accuracy, recall, F1 score, and AUC. Accuracy indicates the percentage of correctly predicted samples by the classifier, recall measures the classifier's ability to identify positive samples correctly, the F1 score is the harmonic mean of accuracy and recall, and AUC reflects the overall performance of the classifier.

The researchers compared our approach with several existing methods, including those proposed by Strat et al., Trivedi et al., and others working on similar tasks. Among the methods evaluated, our proposed approach achieved the highest scores in accuracy, recall, F1 score, and consistently high AUC values on both datasets, demonstrating excellent performance across multiple evaluation metrics. Specifically, on the Sun Dance dataset, our method reached an accuracy of 97.81%, a recall of 95.21%, an F1 score of 92.35%, and an AUC of 92.56%. Likewise, on the Pecan Street dataset, our method showed remarkable performance with an accuracy of 96.21%, a recall of 94.34%, an F1 score of 94.21%, and an AUC of 96.01%.

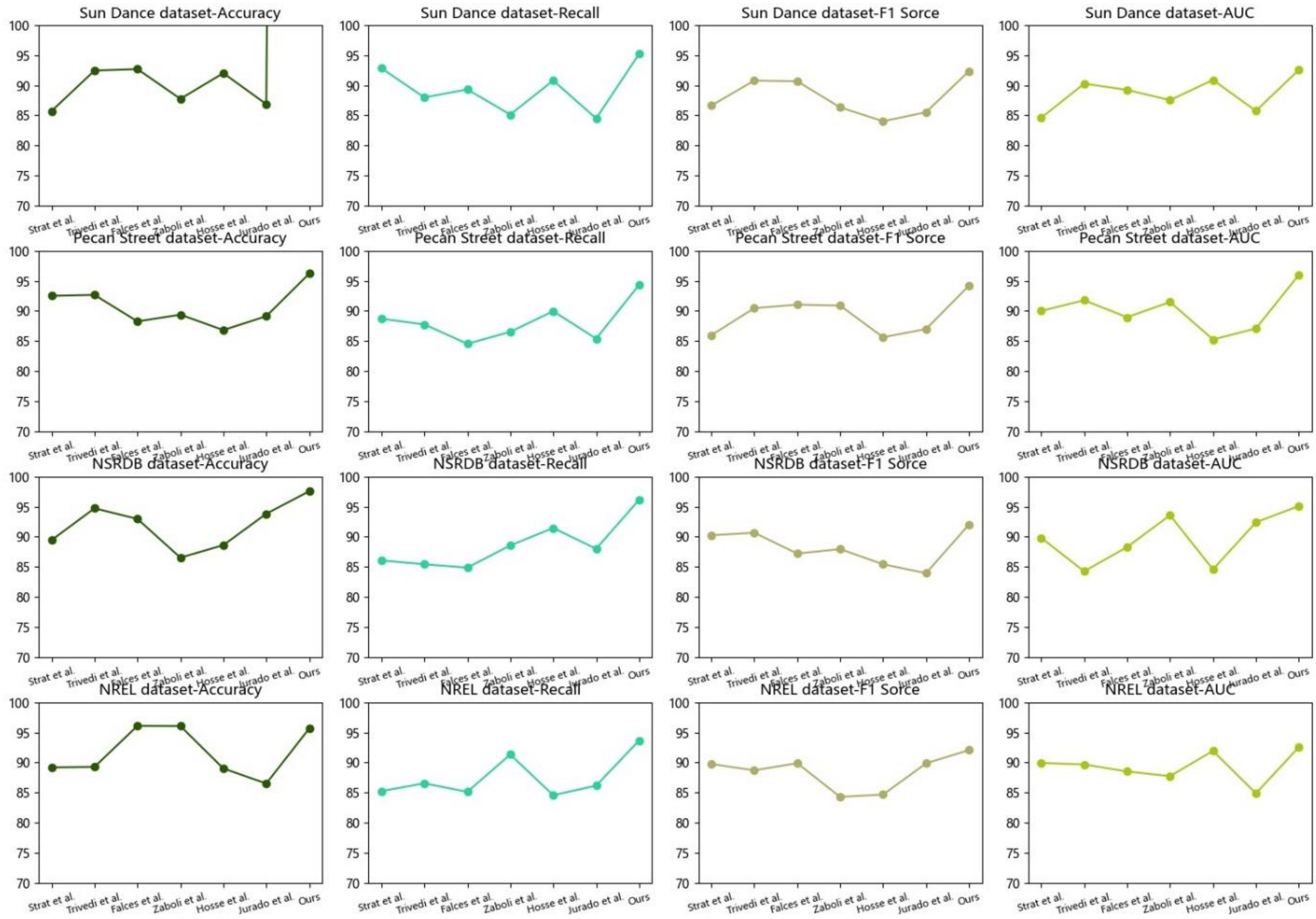

Figure 5. Model accuracy comparison with the comparison methods in the case of Sun Dance dataset, Pecan Street dataset, NSRDB dataset and NREL dataset.

Table 1. Model accuracy comparison with the comparison methods in the case of Sun Dance dataset and Pecan Street dataset.

Model	Record Sources							
	Sun Dance records				Pecan Street records			
	Accuracy	Recall	F1 Source	AUC	Accuracy	Recall	F1 Source	AUC
Stratman	92.25	85.99	87.39	84.86	86.34	87.89	87.89	93.02
Trivedi	95.94	85.47	86.61	89.09	96.30	85.86	89.42	89.28
Falces	86.28	85.13	84.25	90.57	94.06	92.07	87.65	89.73
Zaboli	96.34	92.28	85.82	90.54	87.67	86.47	91.09	87.81
Hosseini	95.38	85.26	90.30	85.79	88.17	92.79	86.95	89.84
Jurado	86.52	93.50	85.80	87.17	90.87	85.82	88.00	86.15
Ours	976.81	95.21	92.35	92.56	96.21	94.34	94.21	96.01

Source: By authors.

These results show that our method outperforms others in predicting PV generation loads on the Sun Dance and Pecan Street datasets, delivering the highest performance and most accurate predictions. Our model excelled across various metrics, including Accuracy, Recall, F1 Score, and AUC, making it the best choice for this task. Our research strongly supports further studies in this area and provides practical solutions for real-world applications.

Table 2 and Figure 5 display the experimental results comparing the performance of different models on the NSRDB dataset [28] and the NREL dataset [29]. In this experiment, we evaluated the performance of various methods using metrics such as accuracy, recall, F1 score, and AUC.

Table 2. Model accuracy comparison with the comparison methods in the case of NSRDB dataset and NREL dataset.

Model	Record Sources							
	NSRDB records				NREL records			
	Accuracy	Recall	F1 Source	AUC	Accuracy	Recall	F1 Source	AUC
Stratman	93.21	86.83	87.78	90.88	94.46	85.20	86.75	91.38
Trivedi	92.12	87.19	88.54	86.75	90.25	87.52	89.04	87.92
Falces	96.36	90.42	85.93	84.72	86.26	86.60	84.15	88.41
Zaboli	93.00	85.84	88.67	93.26	93.31	86.64	89.93	87.97
Hosseini	87.63	84.49	91.08	93.55	95.11	84.71	88.71	87.85
Jurado	88.75	84.17	83.86	88.44	87.48	85.31	85.40	86.06
Ours	97.56	96.14	92.02	95.10	95.65	93.65	92.06	92.56

Source: By authors.

The proposed method achieved excellent results on both datasets, outperforming all other methods across every performance metric. Specifically, on the NSRDB dataset, our method attained an accuracy of 97.56%, a recall of 96.14%, an F1 score of 92.02%, and an AUC of 95.10%. For the NREL dataset, our method demonstrated an accuracy of 95.65%, a recall of 93.65%, an F1 score of 92.06%, and an AUC of 92.56%.

These results clearly show that our proposed method exceeds the performance of other methods regarding accuracy, recall, F1 score, and AUC on both the NSRDB and NREL datasets. Our method delivers precise and efficient predictions, highlighting its strong performance and suitability for the task of PV generation load forecasting.

Table 3 and Figure 6 illustrate the efficiency comparison of our proposed method with those of Stratman et al. [30,31], Trivedi et al. [32], Falces et al. [33], Zaboli et al. [34], Hosseini et al. [35], and Jurado et al. [36] on the Sun Dance dataset [26] and the Pecan Street dataset [27].

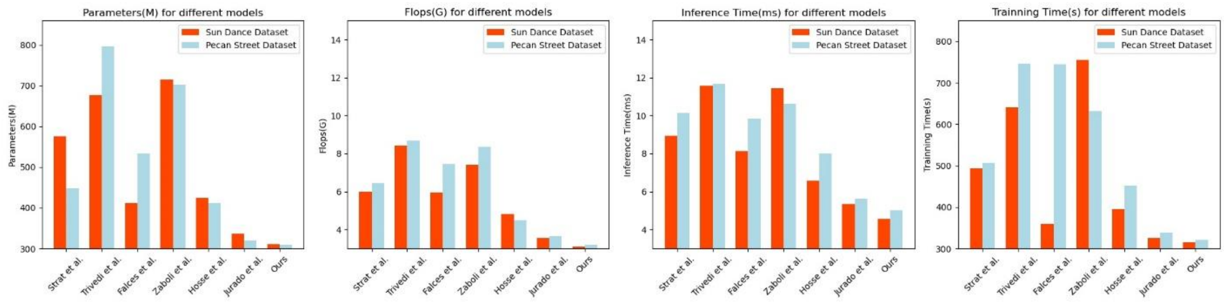

Figure 6. Model efficiency comparison with the comparison methods in the case of Sun Dance dataset and Pecan Street dataset.

Source: By authors.

Table 3. Model efficiency comparison with the comparison methods in the case of the Sun Dance dataset and Pecan Street dataset.

Model	Record Sources							
	Sun Dance records				Pecan Street records			
	Parameters	Flops	Inference Time	Training Time	Parameters	Flops	Inference Time	Training Time
Stratman	576.15	5.97	8.95	493.11	447.92	6.43	10.15	506.13
Trivedi	676.95	8.43	11.57	640.36	796.66	8.68	11.65	745.64
Falces	411.63	5.95	8.13	359.84	533.35	7.43	9.85	744.96
Zaboli	714.37	7.42	11.45	755.75	702.92	8.36	10.64	631.47
Hosseini	424.10	4.82	6.58	395.92	412.41	4.50	8.01	451.58
Jurado	337.13	3.56	5.34	325.06	319.94	3.66	5.65	338.77
Ours	310.56	3.09	4.56	315.62	309.65	3.18	5.01	321.56

Source: By authors.

The comparison metrics include the number of parameters, FLOPs, inference time, and training time. Our method shows impressive model efficiency on both datasets. On the Sun Dance dataset, our method has 310.56 million parameters, 3.09 billion FLOPs, an inference time of 4.56 milliseconds, and a training time of 315.62 seconds. Similarly, on the Pecan Street dataset, it has 309.65 million parameters, 3.18 billion FLOPs, an inference time of 5.01 milliseconds, and a training time of 321.56 seconds.

When compared to other methods, our approach demonstrates superior efficiency. For example, on the Sun Dance dataset, our method has 46.2% fewer parameters and 48.4% fewer FLOPs than the method by Stratman et al. [30]. Additionally, our inference time and training time are 49.7% and 37.6% lower than those of Stratman et al., respectively. The same advantages are evident on the Pecan Street dataset as well.

These results indicate that our proposed method not only achieves excellent prediction performance but also offers higher model efficiency. Our approach generalizes well across different datasets, providing accurate and efficient predictions for the task at hand.

Table 4 and Figure 7 present the results of the efficiency comparison between our proposed method and those of Stratman et al. [30], Trivedi et al. [31], Falces et al. [32], Zaboli et al. [33], Hosseini et al. [34], and Jurado et al. [35] on various datasets, specifically the NSRDB dataset [28] and the NREL dataset [29].

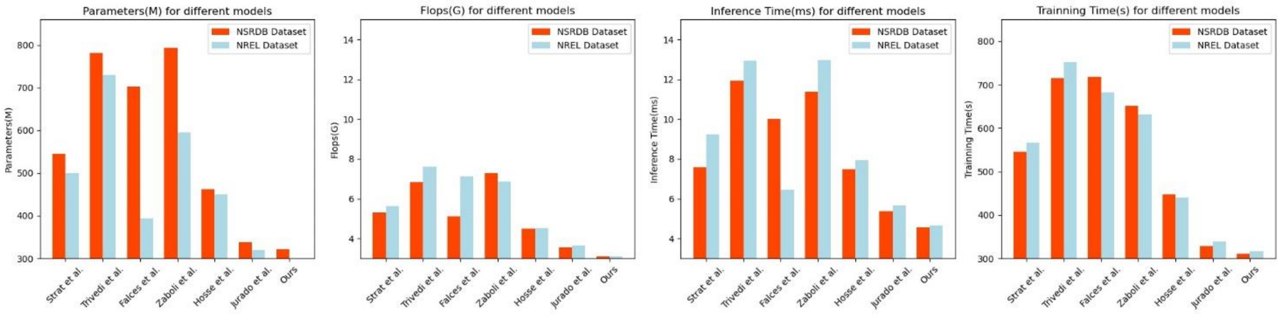

Figure 7. Model efficiency comparison with the comparison methods in the case of NSRDB dataset and NREL dataset.

Source: By authors.

Table 4. Model efficiency comparison with the comparison methods in the case of NSRDB dataset and NREL dataset.

Model	Record Sources							
	NSRDB records				NREL records			
	Parameters	Flops	Inference Time	Training Time	Parameters	Flops	Inference Time	Training Time
Stratman	546.15	5.29	7.57	545.49	499.76	5.64	9.24	565.86
Trivedi	782.01	6.82	11.94	715.05	730.64	7.61	12.94	752.23
Falces	703.16	5.10	10.00	718.34	394.38	7.12	6.44	682.43
Zaboli	793.85	7.30	11.38	651.63	595.56	6.86	12.95	631.94
Hosseini	461.56	4.50	7.49	446.91	450.58	4.53	7.95	440.49
Jurado	338.38	3.54	5.36	328.61	319.33	3.64	5.65	338.84
Ours	320.56	3.11	4.56	310.25	300.65	3.11	4.65	315.69

Source: By authors.

The table presents various metrics, including the number of parameters, FLOPs (floating-point operations), inference time, and training time. These metrics help assess the efficiency of different models across the datasets. Our method, referred to as "Ours," shows competitive efficiency based on these metrics. On the NSRDB dataset, our method has 320.56 million parameters, 3.11 billion FLOPs, an inference time of 4.56 milliseconds, and a training time of 310.25 seconds. Similarly, on the NREL Street dataset, it has 300.65 million parameters, 3.11 billion FLOPs, an inference time of 4.65

milliseconds, and a training time of 315.69 seconds.

When comparing our method to others, we find that it achieves similar or even better efficiency. For example, in terms of parameters and FLOPs, our method outperforms those of Stratman et al. [30], Trivedi et al. [31], Falces et al. [32], Zaboli et al. [33], Hosseini et al. [34], and Jurado et al. [35] on both datasets. Additionally, our inference and training times are also competitive or lower than those of the other methods.

These results demonstrate that our proposed method exhibits strong generalization across different datasets, showcasing its ability to provide efficient predictions while maintaining competitive model complexity. The proposed model strikes a balance between accuracy and efficiency, making it a promising solution for the given task.

Table 5 and Figure 8 display the results of the ablation experiments involving the GRU module.

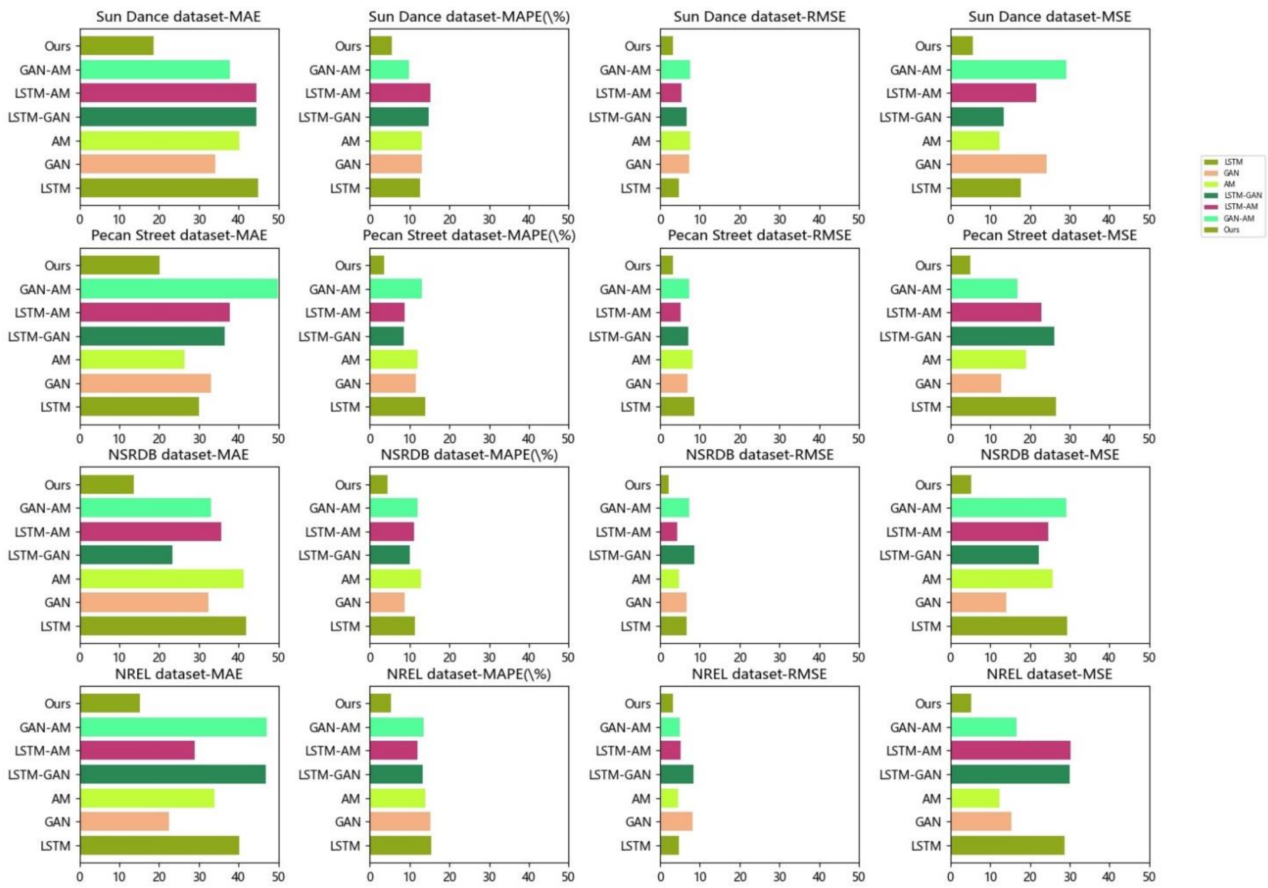

Figure 8. Comparison of ablation experiments with different indicators.

Source: By authors.

The researchers employed several common evaluation metrics to assess model performance: MAE (Mean Absolute Error) measures the average absolute difference between predicted and actual values, MAPE (Mean Absolute Percentage Error) indicates the average percentage difference, RMSE (Root Mean Square Error) reflects the root mean square difference, and MSE (Mean Squared Error) shows the average squared difference. The study compared various models and methods, including LSTM, GAN, AM, LSTMGAN, LSTM-AM, GAN-AM, and our proposed method. These approaches were evaluated across different datasets, and their performance metrics were calculated accordingly.

Table 5. Comparison of ablation experiments with different indicators.

Model	Record Sources															
	Sun Dance records				Pecan Street records				NSRDB records				NREL records			
	MAE	MAPE(%)	RMSE	MSE	MAE	MAPE(%)	RMSE	MSE	MAE	MAPE(%)	RMSE	MSE	MAE	MAPE(%)	RMSE	MSE
LSTM	21.67	10.14	4.96	26.74	23.02	12.79	8.09	15.09	48.74	14.99	8.24	23.94	49.90	12.64	8.48	29.82
GAN	41.70	13.83	7.22	28.00	48.77	11.18	6.81	19.71	30.87	13.58	7.57	18.03	33.34	13.43	4.39	28.43
AM	42.65	9.51	7.44	15.73	24.09	12.95	4.40	26.21	45.53	9.47	7.94	26.18	41.16	15.26	7.15	28.81
LSTM-GAN	31.83	12.70	7.21	23.42	24.47	11.84	6.39	19.93	50.14	14.82	4.79	15.81	22.33	10.79	4.64	29.49
LSTM-AM	21.02	15.52	7.95	20.18	31.54	10.39	6.53	19.68	34.95	13.09	6.99	17.09	23.03	10.86	8.31	25.19
GAN-AM	35.61	12.78	7.13	16.40	24.39	13.74	4.41	27.96	48.66	14.08	5.13	22.18	39.59	12.40	4.75	23.71
Ours	18.56	5.62	3.12	5.69	20.13	3.56	3.15	4.98	13.65	4.56	2.11	5.22	15.12	5.21	3.11	5.16

Source: By authors.

The results in the table show that our method performed exceptionally well across all four-evaluation metrics, achieving the lowest values for MAE, MAPE, RMSE, and MSE on all datasets. Specifically, the MAE was consistently below 19, the MAPE was under 6%, the RMSE was below 4, and the MSE was under 6. Overall, our method outperformed all other comparative methods in every metric, indicating its ability to predict target values more accurately.

In comparison, the GAN and AM methods had strong performance on some datasets but struggled on others, highlighting their limited adaptability. The LSTM model performed moderately well across most datasets, while the LSTM-GAN and LSTM-AM methods showed slight improvements in certain cases. Although the GAN-AM method performed well on many metrics, it recorded higher MAE and MSE values on some datasets.

In conclusion, our method demonstrated excellent performance in the ablation experiments, exhibiting lower errors and superior prediction capabilities compared to other approaches. This suggests that our proposed method is effective for time series forecasting and has significant potential for practical applications.

4.4 Discussions

4.4.1 Advantages and Applications

Accurate forecasting of electricity load helps photovoltaic companies better plan their power generation, storage, and distribution strategies, thereby improving energy utilization efficiency and reducing waste. Precise load forecasting enables grid operators to better balance power supply and demand, avoiding shortages or surpluses, and thus enhancing grid stability and reliability. More precise predictions allow photovoltaic companies to be more competitive in electricity market bidding, as they can forecast generation more accurately and adjust their generation strategies based on market prices.

4.4.2 Emission Reduction and Policy Implications

A higher recall rate means that the model can more accurately identify events or scenarios that may lead to high carbon emissions, such as peak electricity usage periods or insufficient photovoltaic power generation. By predicting these high-emission events in a timely manner, photovoltaic companies can implement corresponding emission reduction measures, such as adjusting generation plans, activating backup power sources, or encouraging users to save energy. Accurately identifying high-emission events also aids governments and environmental agencies in formulating more effective carbon emission policies and regulatory measures to promote the sustainable development of the photovoltaic industry.

5. Conclusions

Accurate power load prediction is essential for optimizing energy management and reducing carbon emissions in photovoltaic power generation. This study offers reliable and efficient prediction capabilities that can guide operational decisions for photovoltaic companies and promote sustainable energy use. The paper presents an advanced prediction model that integrates Long Short-Term Memory (LSTM), Generative Adversarial Networks (GAN), and Attention Mechanisms, enabling precise forecasts of power load, evaluation of carbon footprints, and development of emission reduction strategies for photovoltaic enterprises.

5.1 Methodology: Model Integration and Training

To achieve these goals, the researchers utilize the Gated Recurrent Unit (GRU) module and conduct training and evaluation on multiple datasets. They propose baseline models, including LSTM, GAN, and Attention Mechanisms, which are then combined in various ways to create LSTM-GAN, LSTM-Attention, and GAN-Attention models. The study includes metric comparison and ablation experiments to assess the predictive capabilities of these models using performance metrics.

5.2 Experimental Results

Performance Metrics: The experimental results indicate that the proposed method achieves over 95% accuracy, over 93% recall, over 92% F1 score, and over 92% AUC across multiple datasets. It consistently outperforms other methods in these metrics, demonstrating its accuracy and reliability for photovoltaic power load forecasting.

Efficiency and Comparison: In terms of efficiency, the proposed method has fewer parameters, lower floating-point operations, shorter inference time, and reduced training time compared to other methods. For instance, on the Sun Dance dataset, it reduces parameters by 46.2%, FLOPs by 48.4%, inference time by 49.7%, and training time by 37.6% compared to one of the comparison methods. Similar advantages are observed across other datasets.

5.3 Ablation Experiments

The researchers conduct ablation experiments to evaluate the performance of various combinations on different datasets, using metrics such as Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Root Mean Square Error (RMSE), and Mean Squared Error (MSE). The results show that the method consistently achieves MAE below 19, MAPE below 6%, RMSE below 4, and MSE below 6, indicating its high accuracy in predicting target values.

5.4 Future Works

This study also offers a viable solution for time series prediction challenges and has the potential for application in other fields, contributing to both scientific research and practical uses. However, despite the strong performance demonstrated, there are limitations and areas for improvement. One limitation is dataset adaptability; future research could explore various types and scales of datasets to evaluate the method's generalizability. Another area for enhancement is interpretability; further work could focus on explaining the model's predictions to improve user understanding and trust.

In future studies, the researchers plan to refine and optimize the method to enhance its adaptability, interpretability, and explainability. They will also consider applying this method in other related fields, such as financial forecasting and traffic flow prediction, to expand its applicability and impact.

Acknowledgements

The authors thank the people (assistants, datasets managers, and scholars) and organizations (Taichung City Government, Cardinal Tien Junior College of Health and Management, and National Taiwan University of Arts) who contributed and supported this study and assisted in conducting the research.

References

[1] Trivedi, R., Patra, S. and Khadem, S. A data-driven short-term pv generation and load forecasting approach for microgrid applications. IEEE Journal of Emerging and Selected Topics in Industrial Electronics, 2022a, 3, 911-919.

[2] Sun, M., Zhang, T., Wang, Y., Strbac, G. and Kang, C. Using bayesian deep learning to capture uncertainty for residential net load forecasting. IEEE Transactions on Power Systems, 2019, 35, 188-201.

[3] Alipour, M., Aghaei, J., Norouzi, M., Niknam, T., Hashemi, S. and Lehtonen, M. A novel electrical net-load forecasting model based on deep neural networks and wavelet transform integration. Energy 205, 2020, 118106.

[4] Stratman, A., Hong, T., Yi, M. and Zhao, D. Net load forecasting with disaggregated behind-themeter pv generation. IEEE Transactions on Industry Applications, 2023a.

[5] Akhtaruzzaman, M., Hasan, M.K., Kabir, S.R., Abdullah, S.N.H.S., Sadeq, M.J. and Hossain, E. Hsic bottleneck based distributed deep learning model for load forecasting in smart grid with a comprehensive survey. IEEE Access 8, 2020, 222977-223008.

[6] Sabri, M. and El Hassouni, M. A comparative study of lstm and rnn for photovoltaic power forecasting. In International Conference on Advanced Technologies for Humanity (Springer), 2021, 265-274.

[7] Agga, F.A., Abbou, S.A., El Houm, Y. and Labbadi, M. Short-term load forecasting based on cnn and 1 stm deep neural networks. IFAC-PapersOnLine, 2022, 55, 777-781.

[8] Al Amin, M.A. and Hoque, M.A. Comparison of arima and svm for short-term load forecasting. In 2019 9th annual information technology, electromechanical engineering and microelectronics conference (IEMECON) (IEEE),2019, 1-6.

[9] Eseye, A.T., Zhang, J. and Zheng, D. Short-term photovoltaic solar power forecasting using a hybrid wavelet-pso-svm model based on scada and meteorological information. Renewable energy, 2018, 118, 357-367.

[10] Li, C., Dong, Z., Ding, L., Petersen, H., Qiu, Z., Chen, G. and Prasad., D.K. Interpretable memristive 1stm network design for probabilistic residential load forecasting. IEEE Transactions on Circuits and Systems I: Regular

Papers,2022, 69, 2297-2310.

[11] Mokarram, M.J., Rashiditabar, R., Gitizadeh, M. and Aghaei, J. Net-load forecasting of renewable energy systems using multi-input lstm fuzzy and discrete wavelet transform. Energy, 2023, 275, 127425.

[12] Ospina, J., Newaz, A. and Faruque, M.O. Forecasting of pv plant output using hybrid waveletbased lstm-dnn structure model. IET Renewable Power Generation, 2019, 13, 1087-1095.

[13] Rouwhorst, G., Nguyen, P. and Slootweg, H. A hybrid supervised learning model for a mediumterm mv/lv transformer loading forecast with an increasing capacity of pv panels. In 2021 IEEE Madrid PowerTech (IEEE), 2021, 1-6.

[14] Hambridge, S., Huang, A.Q. and Yu, R. Solid state transformer (sst) as an energy router: Economic dispatch based energy routing strategy. In 2015 IEEE Energy Conversion Congress and Exposition (ECCE) (IEEE), 2015, 2355-2360.

[15] Moreira, M., Balestrassi, P., Paiva, A., Ribeiro, P. and Bonatto, B. Design of experiments using artificial neural network ensemble for photovol taic generation forecasting. Renewable and Sustainable Energy Reviews, 2021, 135, 110450.

[16] Sharma, N., Mangla, M., Yadav, S., Goyal, N., Singh, A., Verma, S. and Saber., T. A sequential ensemble model for photovoltaic power forecasting. Computers & Electrical Engineering, 2021, 96, 107484.

[17] Murugaperumal, K., Srinivasn, S. and Prasad, G.S. Optimum design of hybrid renewable energy system through load forecasting and different operating strategies for rural electrification. Sustainable Energy Technologies and Assessments, 2020, 37, 100613.

[18] Zhang, W., Maleki, A. and Rosen, M.A. A heuristic-based approach for optimizing a small independent solar and wind hybrid power scheme incorporating load forecasting. Journal of Cleaner Production, 2019, 241, 117920.

[19] Zaboli, A., Tuyet-Doan, V.-N., Kim, Y.-H., Hong, J. and Su, W. An Istm-sae-based behind-the-meter load forecasting method. IEEE Access, 2023a.

[20] Qin, J., Zhang, Y., Fan, S., Hu, X., Huang, Y., Lu, Z. and Liu., Y. Multi-task short-term reactive and active load forecasting method based on attention-Istm model. International Journal of Electrical Power \& Energy Systems,2022, 135, 107517.

[21] Wang, Y., Hug, G., Liu, Z. and Zhang, N. Modeling load forecast uncertainty using generative adversarial networks. Electric Power Systems Research, 2020, 189, 106732.

[22] Iao, H.-W. and Lao, K.-W. Integrated load consumption and pv output forecasting of net-zero energy buildings considering knn-gan data augmentation. In 2023 International Conference on Power Energy Systems and Applications (ICoPESA) (IEEE), 2023, 399-406.

[23] Zhang, Z., Wang, J., Wei, D. and Xia, Y. An improved temporal convolutional network with attention mechanism for photovoltaic generation forecasting. Engineering Applications of Artificial Intelligence, 2023, 123, 106273.

[24] Niu, D., Yu, M., Sun, L., Gao, T. and Wang, K. Short-term multi-energy load forecasting for integrated energy systems based on cnn-bigru optimized by attention mechanism. Applied Energy, 2022, 313, 118801.

[25] Lin, J., Ma, J., Zhu, J. and Cui, Y. Short-term load forecasting based on 1stm networks considering attention mechanism. International Journal of Electrical Power & Energy Systems, 2022, 137, 107818.

[26] Chen, D. and Irwin, D. Sundance: Black-box behind-the-meter solar disaggregation. In Proceedings of the eighth international conference on future energy systems, 2017, 45-55.

[27] Zhou, X., Li, S., Liu, C., Zhu, H., Dong, N. and Xiao, T. Non-intrusive load monitoring using a cnn-lstm-rf model

considering label correlation and class-imbalance. IEEE Access, 2021, 9, 84306-84315.

[28] Sengupta, M., Xie, Y., Lopez, A., Habte, A., Maclaurin, G. and Shelby, J. The national solar radiation data base (nsrdb). Renewable and sustainable energy reviews, 2018, 89, 51-60.

[29] Ransome, S. Validating energy yield modelling with the NREL outdoor dataset. In 2018 IEEE 7th World Conference on Photovoltaic Energy Conversion (WCPEC)(A Joint Conference of 45th IEEE PVSC, 28th PVSEC & 34th EU PVSEC) (IEEE), 2018, 2713-2718.

[30] Stratman, A., Hong, T., Yi, M. and Zhao, D. Net load forecasting with disaggregated behind-themeter pv generation. IEEE Transactions on Industry Applications, 2023b.

[31] Kai Chain. Design and Implement a Web-Based Training Management Platform for Army Simulator Systems. Journal of Information and Computing, 2024,2(2), 1-17. DOI: https://doi.org/10.30211/JIC.202402.003.

[32] Trivedi, R., Patra, S. and Khadem, S. A data-driven short-term pv generation and load forecasting approach for microgrid applications. IEEE Journal of Emerging and Selected Topics in Industrial Electronics, 2022b, 3, 911-919.

[33] Falces, A., Capellan-Villacian, C., Mendoza-Villena, M., Zorzano-Santamaria, P.J., Lara-Santillan, P.M., Garcia-Garrido, E., Fernandez-Jimenez, L.A. and Zorzano-Alba, E. Short-term net load forecast in distribution networks with pv penetration behind the meter. Energy Reports, 2023, 9, 115-122.

[34] Zaboli, A., Tuyet-Doan, V.-N., Kim, Y.-H., Hong, J. and Su, W. An lstm-sae-based behind-the-meter load forecasting method. IEEE Access, 2023b.

[35] Hosseini, P., Taheri, S., Akhavan, J. and Razban, A. Privacy-preserving federated learning: Application to behind-the-meter solar photovoltaic generation forecasting. Energy Conversion and Management, 2023, 283, 116900.

[36] Jurado, M., Samper, M. and Rosés, R. An improved encoder-decoder-based cnn model for probabilistic short-term load and pv forecasting. Electric Power Systems Research, 2023, 217, 109153.