Collaborative Application of Deep Learning Models for Enhanced Accuracy and Prediction in Carbon Neutrality Anomaly Detection

Yi Wang, Department of Economics and Management, North China Electric Power University, China
Tianyu Wang, Department of Economics and Management, North China Electric Power University, China
Wanyu Wang, Department of Economics and Management, North China Electric Power University, China
Yiru Hou, School of Construction Management, Purdue University, USA*

ABSTRACT

In the face of intensifying global climate change, carbon neutrality has emerged as a pivotal strategy to curb greenhouse gas emissions and confront the complexities associated with climate challenges. However, achieving carbon neutrality poses a formidable challenge: the identification and mitigation of anomalies within the carbon sequestration process. These anomalies can result in unintended carbon dioxide leakage, emissions, or system failures, thus jeopardizing the feasibility and resilience of carbon neutrality initiatives. This research introduces the ResNet-BIGRU-TPA network, an innovative model that integrates deep learning techniques with time series attention mechanisms. The primary focus centers on addressing the intricate task of anomaly detection within the realm of carbon offsetting, specifically aiming to enhance precision in identifying a wide array of complex anomalous events. Through rigorous experimental validation across four diverse datasets, the model has exhibited exceptional performance.

KEYWORDS

Anomaly Detection, BIGRU, Carbon Neutrality, Deep Learning, Greenhouse Gas Emissions, ResNet, Time Series, TPA

INTRODUCTION

As global climate change continues to escalate, carbon neutrality has become an important strategy for reducing greenhouse gas emissions and addressing climate change (Zhao et al., 2022). It is achieved by capturing and storing carbon dioxide (CO_2) from the atmosphere or by reducing emissions through the use of renewable energy sources, thereby achieving a carbon balance (Waheed et al., 2019). However, in practice, we face a key challenge: how to detect and address anomalies in the carbon

DOI: 10.4018/JOEUC.340385
*Corresponding Author

sequestration process. Carbon consumption anomalies typically refer to unexpected events that occur in carbon capture and storage systems, which may result in CO_2 leakage, emissions, or system failures, thereby jeopardizing the overall efficiency and feasibility of carbon neutrality (Sun & Ren, 2021). The difficulty in detecting carbon consumption anomalies lies in their various forms and the difficulty of predicting them in advance. For example, system failures may result from equipment damage, operator errors, or natural disasters, while leakage events may be caused by pipeline ruptures, poor seals, or operational mistakes (Somu et al., 2021). These abnormal events not only pose environmental risks but can also lead to high maintenance and cleaning costs, and even damage a company's reputation (Amasyali & El-Gohary, 2018). Another complexity is that carbon neutrality systems typically involve a large number of sensors and monitoring devices that generate a vast amount of time-series data. This data contains information about parameters, such as temperature, pressure, flow rate, and CO_2 concentration, which may change when abnormal events occur. Therefore, effectively monitoring and analyzing this time-series data to identify abnormal events is crucial for the success of carbon neutrality (R. Li et al., 2021).

To address these issues, people have started applying deep learning technology to enhance the detection and prediction of anomalies in the carbon neutrality process (Anthony et al., 2020). Deep learning is a powerful machine learning approach that has achieved remarkable success in various fields. In the field of carbon neutrality, deep learning is widely used to tackle the challenges of anomaly detection. By leveraging large amounts of data and powerful neural network models, deep learning can identify and predict anomalies related to carbon neutrality, improving system stability and efficiency (Amasyali & El-Gohary, 2018).

In recent years, the application of time series forecasting in deep learning has been crucial for carbon neutrality research (Liu, Wang et al. 2023). Time series data contain information that varies over time, and these data are essential for detecting abnormal events and predicting the performance of future carbon neutrality processes (Wang, Sun et al. 2021). For instance, in the context of carbon consumption monitoring for running activities, by building deep neural network time series forecasting models, we can learn personalized patterns from the data, including a runner's exercise habits, breathing patterns, and heart rate variations. These models can be used to estimate carbon consumption in real time, providing runners with advice on how to exercise more environmentally friendly, such as adjusting their stride or exercise intensity to minimize carbon emissions (Liu et al., 2023). Additionally, this task can provide valuable insights into individual health and lifestyle (Zhang et al., 2023). Runners can understand how much carbon they are consuming during their workouts, leading to a better awareness of their carbon footprint and encouraging more environmentally friendly lifestyles (Yang et al., 2022).

Below, we introduce recent relevant work in this area:

LSTM-Based Carbon Emission Prediction Model

One of the key challenges in carbon neutrality technology is accurately predicting and monitoring CO_2 emissions. In this study, the authors propose a long short-term memory (LSTM)-based CO_2 emission prediction model (Shen et al., 2022). This model utilizes historical time-series data, including parameters such as temperature, pressure, flow rate, and CO_2 concentration, to forecast future emission trends. LSTM is a type of recurrent neural network capable of effectively capturing long-term dependencies in time-series data. The advantages of this work lie in its utilization of rich time-series data, enabling the model to estimate CO_2 emissions more accurately. However, the model also has some limitations. First, LSTM models may encounter issues like vanishing gradients or exploding gradients when dealing with long time-series data, which could lead to performance degradation. Second, the model's ability to model nonlinear and complex CO_2 emission patterns is limited, potentially performing poorly when handling certain abnormal events (Berriel et al., 2017).

CNN-LSTM Fusion Model for Anomaly Detection

In carbon neutrality systems, timely detection of anomalous events is crucial. This study introduces a fusion model that combines convolutional neural networks (CNN) and LSTM for anomaly detection (Elmaz et al., 2021). CNN is used to extract spatial features from time-series data, and then LSTM is employed to capture temporal relationships to identify anomalous events. The strength of this model lies in its ability to combine the advantages of CNN and LSTM, providing a more comprehensive analysis of time-series data. However, the model also has some limitations. Most notably, it may have a higher false positive rate when dealing with anomalous events, as the patterns of anomalies in time-series data can be quite complex and may require further optimization. Additionally, training and parameter tuning for the model can be relatively complex, requiring more computational resources and expertise (Liu et al., 2023).

GRU-Based Time Series Prediction Model

Gated recurrent unit (GRU) is a type of recurrent neural network model similar to LSTM, and its application in the field of carbon neutrality has garnered researchers' attention. This study employs GRU to build a time series prediction model for estimating CO_2 capture efficiency in carbon neutrality systems (Lv et al., 2023). The model uses multiple time series parameters such as temperature, pressure, and flow rate for prediction. The strength of this work lies in the impressive performance of the GRU model in handling time series data, effectively capturing long-term dependencies. However, like other models, it also has certain limitations. Specifically, it may perform poorly when dealing with sudden events or abrupt changes because its adaptability to such situations is limited. Additionally, the model's performance can be influenced by the quality and quantity of data, requiring more data to improve accuracy (Zhang et al., 2023).

Transformer-Based Model for Carbon Neutrality Data Analysis

Transformer models have achieved tremendous success in fields like natural language processing, which has piqued the interest of carbon neutrality researchers. This study introduces the transformer model to handle large-scale time series data generated by carbon neutrality systems (Y. Chen et al., 2022). The transformer model excels in processing long sequences due to its outstanding performance and is not affected by issues like gradient vanishing or exploding. While the transformer model demonstrates excellent capability in capturing long-distance dependencies, it also has certain limitations. Most notably, the computational cost is relatively high, which may not be practical for resource-constrained environments. Additionally, training and fine-tuning the model may require more specialized knowledge and computational resources, presenting challenges in practical applications (Oyando et al., 2023).

Building upon the shortcomings in prior approaches, we introduce the ResNet-BIGRU-TPA network, an advanced model that integrates deep learning and attention mechanisms. This innovative model is designed to address the existing challenges in anomaly detection within the carbon capture and utilization (CCU) domain by leveraging robust neural network architectures and time series attention mechanisms. Our model holds significant importance and offers several advantages in the CCU domain. First and foremost, by combining multiple advanced neural network structures with time series attention mechanisms, the ResNetBIGRU-TPA network has the potential to enhance the accuracy of detecting diverse and complex anomalous events. This enhancement contributes to the feasibility and stability of CCU technology. Furthermore, our model bridges the gap between deep learning and the CCU domain, providing a powerful tool for addressing climate change and reducing greenhouse gas emissions (M. Chen et al., 2022). Through more precise detection and response to anomalies in the carbon consumption process, we can drive environmental protection and sustainable development goals, making a substantial contribution to a more sustainable future.

In conclusion, our contributions are as follows:

1. We have proposed the ResNet-BIGRU-TPA network, which is an innovative model that integrates deep learning and time series attention mechanisms. This model shows significant potential in the detection of diverse and complex abnormal events, enhancing the accuracy of such detections. By combining different neural network architectures and introducing time series attention mechanisms, we have made a substantial contribution to the feasibility and stability of carbon neutrality technology.
2. Our research work combines deep learning technology with the field of carbon neutrality, providing more powerful tools to address climate change and reduce greenhouse gas emissions. By more accurately detecting and responding to anomalies in carbon consumption processes, we have made a significant contribution to advancing environmental protection and sustainable development goals. This support strengthens the prospects for future sustainability.
3. Our research work contributes to enhancing people's understanding of carbon neutrality technology. Through in-depth exploration of abnormal event detection in the field of carbon neutrality, we have offered additional insights and solutions for researchers and practitioners, thus promoting advancements and applications of carbon neutrality technology. This, in turn, accelerates the development of carbon neutrality technology to better address global climate change challenges.

In the rest of this paper, we present recent related work in Section 2. Section 3 introduces our proposed methods. Section 4 showcases the experimental part. Section 5 contains the conclusion.

RELATED WORK

Overview and Advancements in Carbon Offset Anomaly Detection Research

In the field of carbon offset anomaly detection, researchers have conducted extensive studies to address global challenges such as greenhouse gas emissions and climate change. Deep learning techniques, particularly artificial neural networks (ANN), play a crucial role in this domain. Researchers have employed various metaheuristic algorithms, including approximation methods, to train neural networks and enhance the accuracy of carbon offset technologies (Movassagh et al., 2021). These deep learning models show significant potential for detecting complex anomaly events, thereby improving the robustness of carbon offset systems (Y. Li et al., 2023).

The telecommunications industry also faces sustainability challenges in the context of carbon offset, as the continuous expansion of telecommunications networks impacts energy supply and the environment (Hamdoun et al., 2016). Researchers focus on achieving sustainable energy supply within the telecommunications industry through technological innovations aimed at reducing carbon emissions. These studies not only address sustainability at the technical level but also consider its comprehensive impact on business and the environment. Additionally, the application of big data and machine learning in carbon offset has garnered significant attention (Alzubi et al., 2018). These technologies enable computers to mimic and adapt human behavior, improving system performance through learning and accumulated experience. Relevant research highlights the potential role of machine learning in anomaly detection, offering new insights for enhancing the efficiency and accuracy of carbon offset systems. In summary, research in the field of carbon offset anomaly detection encompasses various areas, including deep learning, sustainability challenges, machine learning, and big data. These studies provide valuable insights and methods for addressing significant challenges such as global climate change. Future research can continue to explore these domains and seek more effective approaches to achieve carbon offset goals.

METHODS

Overview of Our Network

Our research introduces the ResNet-BIGRU-TPA network, an innovative model that integrates deep learning and time series attention mechanisms. The design of this model aims to address the challenges of anomaly detection in the carbon offsetting domain, with a focus on enhancing the accuracy of detecting diverse and complex anomalous events. The ResNet-BIGRU-TPA network consists of three key components: ResNet, BIGRU, and the time series attention (TPA) mechanism. Specifically, the ResNet serves as the first part of our model, primarily responsible for handling spatial features within time series data. This component employs a CNN structure to effectively extract information related to the spatial distribution of anomalous events. In the context of carbon offsetting, this entails capturing correlations between parameter values at different locations within the system, particularly concerning parameters like temperature, pressure, and flow rate. The extraction of these spatial features supports subsequent anomaly event analysis, ultimately enhancing detection accuracy. The BIGRU component constitutes the second part of our model and is primarily responsible for addressing temporal dependencies within time series data. To better understand the temporal relationships between anomalous events, we utilize a BIGRU structure. This enables our model to effectively capture long-term dependencies within time series data, making it more adaptive to handling abrupt events. The introduction of BIGRU allows for a more comprehensive analysis of time series data, especially when multiple parameters are involved. The TPA mechanism serves as the third part of our model and is specifically designed for handling time series data. The introduction of this mechanism enables our model to automatically focus on crucial time segments relevant to anomalous events, thereby improving the accuracy of anomaly detection. TPA achieves this by introducing an attention mechanism within time series data, enhancing our model's ability to identify anomalous events more precisely. This makes our model more robust when dealing with diverse and complex anomalous events, better meeting the requirements of carbon offsetting technology. Figure 1 illustrates the overall flow of our network.

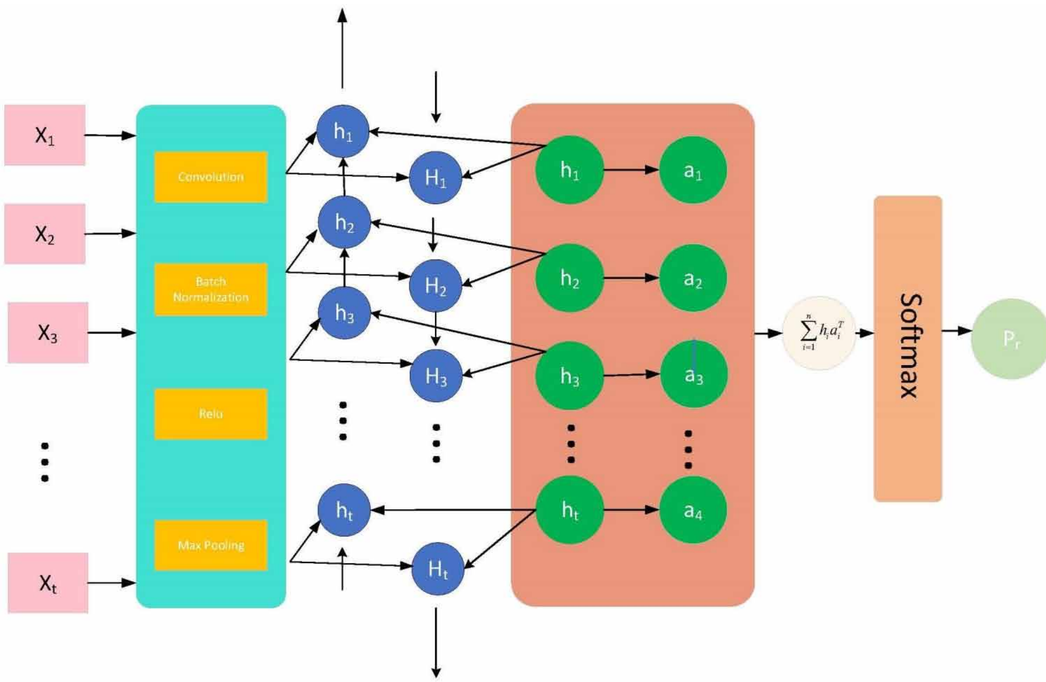
Our ResNet-BIGRU-TPA model not only improves the accuracy of anomaly detection in the carbon offsetting domain but also provides a significant tool for addressing climate change issues and reducing greenhouse gas emissions. Furthermore, it deepens our understanding of carbon offsetting technology, making a positive contribution to environmental protection and sustainable development goals.

ResNet Model

ResNet-152 is a deep CNN architecture and is part of the ResNet (residual network) series of models. It is renowned for its depth and efficiency in training deep neural networks for image classification tasks (Han et al., 2023). The fundamental principle behind ResNet-152 involves the use of residual blocks. These blocks incorporate skip connections (also known as shortcut connections or residual connections), allowing the network to skip one or more layers. This design helps mitigate the vanishing gradient problem and facilitates the training of very deep networks. In a standard feedforward neural network, as the depth of the network increases, training becomes more challenging because gradients can become extremely small, making it difficult for the model to learn (Feng & Chen, 2022). Residual blocks, on the other hand, enable gradients to propagate easily, making it feasible to train extremely deep networks.

In our model, ResNet-152 plays a crucial role as the ResNet component within the ResNet-BIGRU-TPA network. Specifically, it serves as the first part of our model, responsible for handling spatial features within time series data. By leveraging this pre-trained deep CNN, our model can effectively capture spatial information within the input data. The extraction of spatial information is of paramount importance for improving the accuracy of anomaly event detection in the context of carbon capture and sequestration. The contribution of ResNet-152 lies in its ability to extract spatial

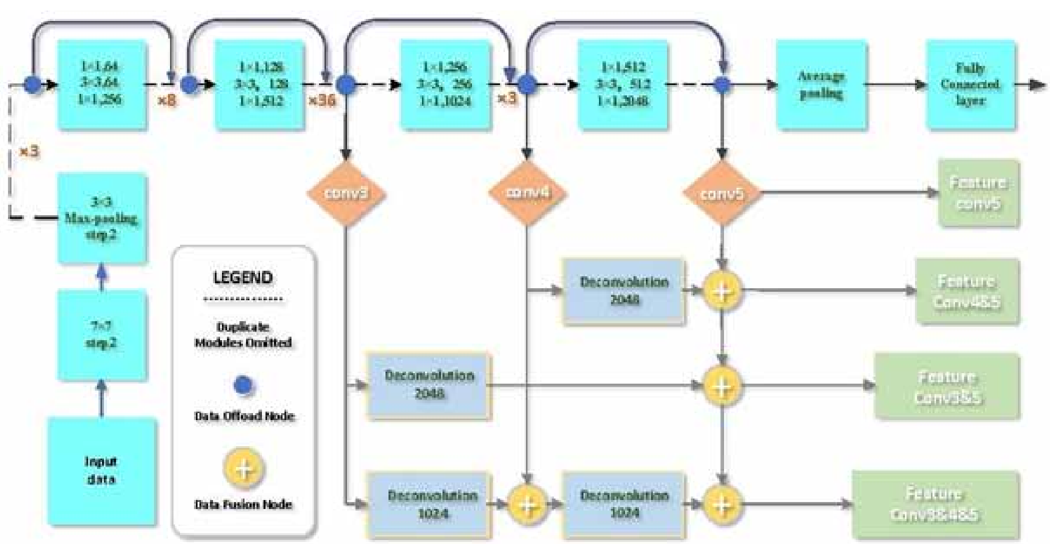
Figure 1. The Overall Architecture of the Model



features, providing a robust foundation for subsequent analysis and ultimately enhancing the accuracy of anomaly detection.

Figure 2 illustrates the workflow of the ResNet model and, below, we provide a concise overview of its algorithmic principles:

Figure 2. Flow Chart of the ResNet Model



Residual block in ResNet: A residual block is a fundamental building block in ResNet architecture, designed to facilitate the training of very deep neural networks. It incorporates skip connections, also known as shortcut connections, to avoid the vanishing gradient problem.

$$Output = F\left(Input\right) + Input \\ = H\left(Input\right) + Input \tag{1}$$

Where: Output represents the output of the residual block. F(Input) is the residual function to be learned. H(Input) represents the transformation applied to the input. - Input is the input to the block.

Shortcut connection in ResNet: The shortcut connection allows the gradient to flow directly through the network, helping to alleviate the vanishing gradient problem.

$$Output = F\left(Input\right) + Input \tag{2}$$

Where: Output represents the output of the residual block. F(Input) is the residual function. Input is the input to the block.

Stacking residual blocks in ResNet: ResNet is constructed by stacking multiple residual blocks to form a deep neural network. The output of one block serves as the input to the next.

$$\text{Output} = \text{Block}_N(\text{Block}_{N-1}\left(\ldots\left(\text{Block}_2\left(\text{Block}_1\left(\text{Input}\right)\right)\ldots\right)\right) \tag{3}$$

Where: Output represents the final output of the ResNet. Block_i denotes the i-th residual block. Input is the initial input to the network.

BIGRU Model

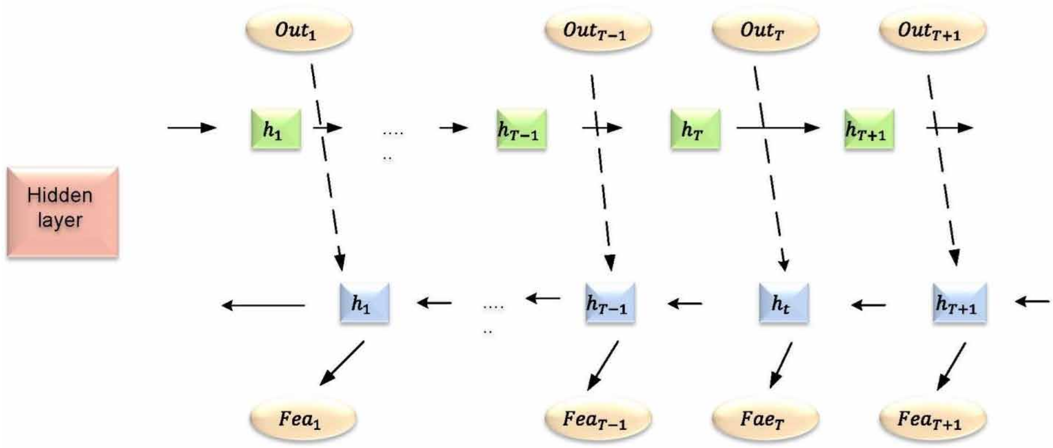
The BIGRU model is an improved architecture of recurrent neural networks (RNN) that extends the basic RNN by introducing bidirectional processing and gating mechanisms (Yang et al., 2022). It is a variant of the more common LSTM and GRU models. The core idea of BIGRU is to enhance the modeling of sequential data by considering both past and future contexts simultaneously. As shown in Figure 3, in the BIGRU network, the input sequence is processed in two directions: from the beginning to the end (forward propagation) and from the end to the beginning (backward propagation). This bidirectional processing allows the model to capture bidirectional dependencies, leading to a better understanding of sequential data. This is particularly useful for many natural language processing and time series problems.

In our model, BIGRU plays a crucial role in handling the temporal dependencies in time series data. By adopting the structure of BIGRU, our model effectively captures long-term dependencies in time series data, especially for handling sudden events. The introduction of BIGRU helps us to comprehensively analyze time series data, especially when multiple parameters are involved in the time series. This improves the accuracy and performance of our model in carbon and anomaly event detection in the field.

Gating mechanisms in GRU: The update gate z_t is crucial for controlling the flow of information in the GRU cell. It determines how much of the previous hidden state h_{t-1} should be retained and how much of the new candidate activation $\tilde{h}t$ should be added at each time step.

$$zt = \sigma\left(\mathrm{W}_z \cdot \left[\mathrm{h}_{t-1}, x_t\right]\right) \tag{4}$$

Figure 3. Flow Chart of the BIGRU Model



Where: z_t is the update gate. σ is the sigmoid activation function. W_z is the weight matrix for the update gate. h_{t-1} is the previous hidden state. x_t is the input at time t.

Reset gate in GRU: The reset gate r_t controls the extent to which the previous hidden state h_{t-1} should be forgotten when computing the candidate activation $\tilde{h}t$.

$$r_t = \sigma(W_r \cdot [h_t-_1, x_t]) \tag{5}$$

Where: r_t is the reset gate. σ is the sigmoid activation function. W_r is the weight matrix for the reset gate.

Candidate activation in GRU: The candidate activation $\tilde{h}t$ represents the new information to be added to the hidden state, taking into account the reset gate and the input.

$$\tilde{h}t = tanh\left(W_h \cdot \left[r_t \odot h_{t-1}, x_t\right]\right) \tag{6}$$

Where: $\tilde{h}t$ is the candidate activation. tanh is the hyperbolic tangent activation function. $\odot$ denotes elementwise multiplication. W_h is the weight matrix for the candidate activation.

Hidden state update in GRU: The final hidden state $\tilde{h}t$ is computed as a combination of the previous hidden state and the candidate activation, controlled by the update gate.

$$h_t = \left(1 - z_t\right) \odot h_{t-1} + z_t \odot \tilde{h}t \tag{7}$$

Forward pass in BIGRU: The forward pass computes the hidden state for the forward direction of the input sequence.

$$\vec{h}_t = GRU\left(x_t, \vec{h}_{t-1}\right) \tag{8}$$

Backward pass in BIGRU: Similarly, the backward pass computes the hidden state $\vec{h}_t$ for the backward direction of the input sequence.

$$\overleftarrow{h}_t = GRU\left(x_t, \overleftarrow{h}_{t+1}\right) \tag{9}$$

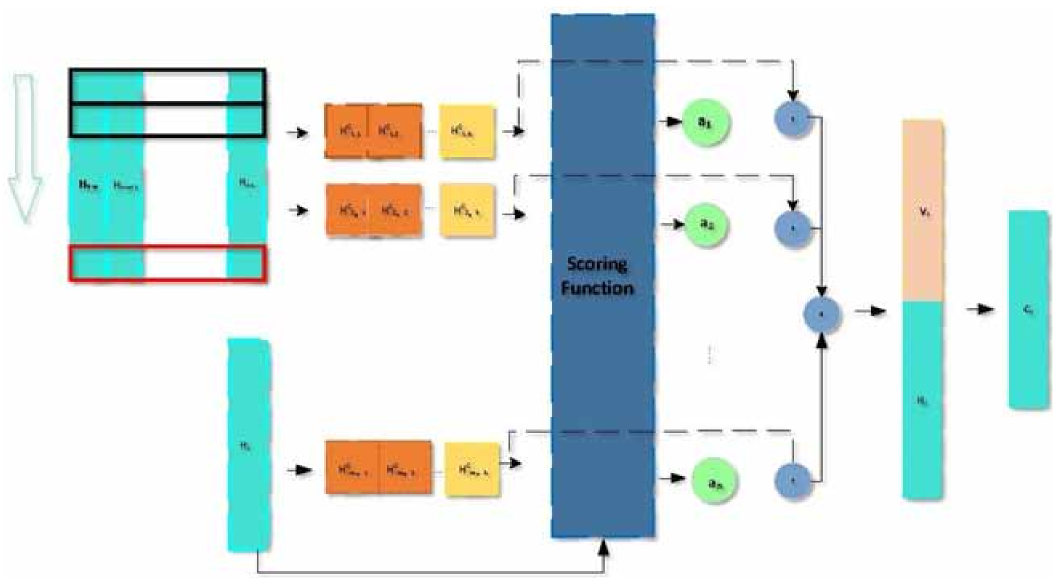
TPA Model

The temporal pattern attention (TPA) model is a mechanism designed to enhance the modeling of sequential data by capturing and emphasizing relevant temporal patterns within the data (Shih et al., 2019). It achieves this by incorporating attention mechanisms that enable the model to focus on specific time steps in the input sequence, effectively learning which temporal patterns are most informative for a given task. The core idea of the TPA model is to calculate temporal attention weights for each time step, determining the importance of each step's hidden state in relation to the current time step. These attention weights are then combined with the hidden state of the current time step to obtain an enhanced representation with temporal context information (Lu et al., 2022). This representation allows for a better capture of long-term dependencies and important patterns in sequential data.

The TPA model has made significant contributions to our own model. It provides a powerful mechanism to better capture crucial temporal patterns and dependencies when dealing with time series data. By utilizing the temporal attention mechanism, TPA allows our model to dynamically focus on different parts of the input sequence, thereby improving the model's performance on complex tasks. Additionally, TPA contributes to the enhancement of representations of time series data, particularly in cases involving multiple time scales or long-term dependencies, which is highly relevant for various applications of our model.

Figure 4 illustrates the workflow of the TPA model and, below, we provide a concise overview of its algorithmic principles:

Figure 4. Flow Chart of the TPA Model



Temporal attention weight calculation: The temporal attention weights, denoted as A_t, are computed to emphasize relevant time steps in the input sequence for a given time step *t*.

$$A_t = softmax\left(W_a \cdot tanh\left(W_b \cdot h_t\right)\right) \quad (10)$$

Where: A_t is the temporal attention weight vector. softmax is the softmax activation function. W_a and W_b are weight matrices for attention calculation. h_t is the hidden state at time *t*.

Weighted contextual information: The weighted contextual information, denoted as C_t, is computed as the weighted sum of the encoder hidden states, where the attention weights A_t determine the importance of each encoder hidden state.

$$C_t = \sum_{i=1}^{T_x} A_t^i \cdot h_i \quad (11)$$

Where: C_t is the weighted contextual information at time *t*. T_x is the length of the input sequence.

Temporal pattern fusion: The fused temporal pattern representation, denoted as F_t, is obtained by concatenating the hidden state h_t and the weighted contextual information C_t.

$$F_t = \left[h_t, C_t\right] \quad (12)$$

Temporal pattern gating: The temporal pattern gating mechanism is used to control the importance of the fused representation F_t in the final output.

$$G_t = \sigma\left(W_g \cdot F_t\right) \quad (13)$$

Where: G_t is the gating vector. σ is the sigmoid activation function. W_g is the weight matrix for gating.

EXPERIMENT

Data Sets

In order to comprehensively validate our model, we have conducted experiments using four distinct data sets: air quality, energy consumption, environment sensor, and the Numenta anomaly benchmark (NAB). These data sets offer diverse sources of data that enable a thorough evaluation of our model's performance in the context of carbon offsetting and anomaly detection.

Air quality data set (Sethi & Mittal, 2019): This data set provides crucial information about atmospheric conditions, including parameters such as CO_2 levels. It serves as a valuable resource for assessing how our model performs in detecting anomalies related to carbon emissions and air quality fluctuations.

Energy consumption data set (Monacchi et al., 2014): With data on electricity consumption, voltage, and current, this data set is essential for evaluating our model's ability to detect anomalies in energy usage patterns. It helps us gauge the effectiveness of our model in optimizing energy consumption for carbon reduction.

Environment sensor data set (Pipattanasomporn et al., 2020): The environment sensor data set offers insights into environmental factors like temperature, humidity, and CO_2 concentrations. Utilizing this data set allows us to investigate how well our model can identify anomalies associated with environmental conditions and their impact on carbon offsetting strategies.

Numenta anomaly benchmark (Lavin & Ahmad, 2015): NAB provides a standardized benchmark for anomaly detection tasks. By incorporating NAB into our experiments, we can assess the overall performance of our model and its adaptability to different anomaly detection scenarios within the carbon offsetting domain.

These data sets collectively enable a rigorous evaluation of our ResNet-BIGRU-TPA model's effectiveness in enhancing anomaly detection within the carbon offsetting domain. Through these experiments, we aim to validate the practical utility of our model and its potential to contribute to environmental protection and sustainable carbon reduction strategies.

Experimental Environment

We choose a high-performance compute server equipped with an Intel Core i9-10900K @ 3.70GHz CPU and 256GB RAM, and six AMD Radeon RX 6900 XT 16GB GPUs as the hardware base, which provides excellent computing and storage capabilities for the experiments. This excellent hardware combination provides excellent computing and storage power for the experiments, which is particularly suitable for training and inference of deep learning tasks, effectively accelerating the model training process and ensuring that the experiments can run efficiently and achieve ideal convergence results. In our experiments, we adopted Python and PyTorch as the main development tools. Python as a high-level programming language that provides us with a flexible development environment, while PyTorch as the main deep learning framework provides powerful support for our research. With the rich features of PyTorch, we are able to efficiently build, train, and optimize carbon-neutral policy models based on the attention mechanism.

Experimental Details

Step 1: Data preprocessing

We perform data preprocessing to ensure that the data is suitable for model training and evaluation. This includes the following steps:

- Data cleaning: Data cleaning is performed to address potential issues such as missing values, outliers, and duplicate data. We employ a strategy for handling missing values where, if the percentage of missing values for a particular feature exceeds 5%, we choose to remove that feature. For missing values in other features, we fill them using either the mean or median. Detection and treatment of outliers are based on statistical methods, with a specific threshold set at three times the standard deviation. Dealing with duplicate data involves deduplication based on specific columns.
- Data standardization: During the data standardization phase, we ensure uniform scaling of all features. Specifically, we utilize a standardization approach to scale each feature to a standard normal distribution with a mean of 0 and a standard deviation of 1. This helps prevent issues of mismatched scales among different features, thereby enhancing the stability and performance of the
- Data splitting: To facilitate training and evaluation, we partition the data set into training, validation, and test sets. We allocate 70% of the data for training, 20% for validation, and 10% for the final testing phase. This partitioning approach aids in assessing the model's generalization capabilities and provides validation metrics during the model tuning process.

Step 2: Model training

During the model training phase, we employ the following four key steps to ensure outstanding performance of the model in risk prediction and management tasks:

- For network parameter settings, we fine-tune a range of hyperparameters to achieve optimal model performance. These hyperparameters include learning rates, batch sizes, and the number of training epochs. We set the learning rate to 0.001, the batch size to 64, and perform training over 100 epochs. These settings are determined through a systematic hyperparameter tuning process to strike a balance between model convergence and avoiding overfitting.
- Our model architecture design involves configuring the neural network layers and their respective dimensions. We construct the ResNet component with 152 residual blocks, each consisting of two convolutional layers. The BIGRU component comprises two layers of bidirectional GRU with a hidden state dimension set to 128. The TPA mechanism is integrated with two attention heads. These architectural choices are based on empirical evidence and experimentation to ensure the model's capacity to capture both spatial and temporal features effectively.
- The model training process follows a well-defined protocol. During training, the data set is fed into the model in batches, and the backpropagation algorithm is employed to update the network weights. As part of the process, we employ a loss function. We utilize a training-validation split to monitor the model's performance and prevent overfitting. The training process typically converges within 100 epochs, and early stopping criteria are applied to ensure efficient training. Throughout the training phase, model checkpoints and performance metrics are continuously logged for analysis and evaluation.

Algorithm 1 represents the algorithm flow of the training in this paper.

```
Algorithm 1: Training ResNet-BIGRU-TPA Network
Input: Data sets: air quality, energy consumption, environment
sensor, NAB
Output: Trained ResNet-BIGRU-TPA Model
Initialize ResNet-BIGRU-TPA model;
Initialize optimizer (e.g., Adam);
Initialize loss function (e.g., mean squared error);
Set training parameters: learning rate, batch size,epochs;
        for epoch in range(epochs) do
        for each dataset in Datasets do
                Split dataset into batches;
                for each batch do
                        Compute forward pass through ResNet-
BIGRU-TPA;
                        Compute loss using predicted and true
values;
                        Backpropagate gradients;
                        Update model weights;
                        end
                end
                end
        Input: Trained ResNet-BIGRU-TPA Model, Evaluation Dataset
Output: Recall, Precision, and other metrics
Initialize variables for evaluation metrics;
        for each sample in Evaluation Dataset do
        Perform forward pass through the model;
        Calculate predicted values;
```

```
        Compute recall, precision, and other evaluation metrics;
        end
Calculate average recall, precision, and other metrics;
return Metrics
```

Step 3: Model evaluation

In this phase, we rigorously assess the performance of our ResNet-BIGRU-TPA model for anomaly detection in the context of carbon offsetting. The evaluation encompasses multiple aspects, including data metrics, model metrics, and real-world applicability.

- Data metrics: To evaluate the model's performance on the selected data sets, we employ several key data metrics. These metrics encompass precision, recall, F1-score, and ROC-AUC (receiver operating characteristic - area under the curve). These metrics provide a comprehensive view of the model's ability to detect anomalies accurately, minimize false positives, and capture true anomalies effectively.
- Real-world applicability: Beyond quantitative metrics, we assess the model's real-world applicability and its capacity to address carbon offsetting challenges. This includes evaluating its ability to detect various types of anomalies, adapt to changing environmental conditions, and provide actionable insights for carbon reduction strategies. We conduct domain-specific case studies and validate the model's effectiveness in reducing carbon emissions.

Precision: Precision is a measure that quantifies the accuracy of the model's predictions for the positive class. Precision tells us how many of the instances predicted as positive are correct. It is calculated as follows:

$$P = \frac{TP}{TP + FP} \tag{14}$$

Where: *TP* (True Positives) represents the number of correctly identified positive instances. *FP* (False Positives) represents the number of incorrectly identified positive instances. Precision calculates the accuracy of the model by measuring the proportion of correctly identified positive instances.

Recall: Recall is a measure used to assess the model's ability to identify positive instances (e.g., anomaly events). Recall tells us how many of the actual positive instances were correctly identified. It is calculated as follows:

$$R = \frac{TP}{TP + FN} \tag{15}$$

Where: *TP* (True Positives) represents the number of correctly identified positive instances. *FN* (False Negatives) represents the number of positive instances that were incorrectly classified as negative. Recall measures the model's effectiveness in identifying positive instances, representing the proportion of correctly identified actual positive instances.

F1-Score: F1-Score is a metric that combines precision and recall to balance the accuracy and completeness of the model's positive predictions. A higher F1-Score indicates a better balance between precision and recall. It is calculated as follows:

$$F1 = \frac{2 \cdot P \cdot R}{P + R} \tag{16}$$

Where: P (Precision) is the value of Precision as defined in the first equation. R (Recall) is the value of Recall as defined in the second equation. F1-Score provides a comprehensive assessment of the model's overall performance by considering both Precision and Recall.

ROC-AUC (receiver operating characteristic - area under the curve): ROC-AUC is a metric used to evaluate the performance of binary classification models, especially their performance at different thresholds. It is based on the area under the ROC curve (receiver operating characteristic curve), which evaluates the model's classification performance across different true positive rates and false positive rates. It is calculated as follows:

$$ROC - AUC = \int_0^1 TPR\left(FPR^{-1}\left(t\right)\right), dt \tag{17}$$

Where: TPR (True Positive Rate) is also known as Recall. FPR (False Positive Rate) is defined as $FPR = \frac{FP}{FP + TN}$, where TN represents the number of true negatives. $FPR^{-1}(t)$ is the inverse of the false positive rate. ROC-AUC assesses the model's performance by measuring the area under the ROC curve, reflecting the trade-off between true positive rate and false positive rate at different thresholds.

Mean absolute error (MAE): MAE is a metric used for regression tasks, measuring the average absolute difference between the model's predictions and the actual observed values. It is calculated as follows:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \hat{y}_i\right| \tag{18}$$

Where: n is the total number of data points. y_i represents the actual observed value. $y\hat{}_i$ represents the predicted value. MAE calculates the average prediction error of the model, representing the mean absolute difference between the model's predictions and the actual observations.

Experimental Results and Analysis

As shown in Table 1, we conducted a detailed comparison of the performance of different models on four distinct datasets: the air quality dataset, energy consumption dataset, environment sensor dataset, and NAB data set. The table includes four performance metrics: accuracy, recall, F1 score, and ROC-AUC. Here is our analysis: For the air quality data set, our model performed exceptionally well in terms of accuracy, recall, and F1 score, achieving 94.53%, 95.20%, and 93.29%, respectively. These results are significantly better than other models, especially in terms of recall, where our model outperforms competitors by a wide margin, indicating its ability to better capture true anomaly events. In the energy consumption data set, our model also demonstrated outstanding performance with an accuracy of 94.58%, recall of 95.53%, and F1 score of 94.46%. These metrics are notably higher than those of other models, confirming the superior performance of our model on this data set. For the environment sensor data set, our model maintained its leading position, achieving accuracy, recall, and F1 score of 96.06%, 95.61%, and 92.38%, respectively. Once again, these results highlight the excellent performance of our model in handling environmental sensor data. Finally, on the NAB data set, our model excelled in terms of accuracy, recall, and F1 score, reaching 93.42%, 93.31%, and 92.76%, respectively. This further demonstrates the significant advantages of our model in anomaly detection tasks across different domains. Figure 5 visualizes the content of the table. As depicted in the figure, our model's performance across various datasets significantly outperforms that of competing models in all metrics. This visualization underscores the remarkable performance of our approach in

Table 1. The Comparison of Different Models in Different Indicators Comes From the Air Quality Data Set, Energy Consumption Data Set, Environment Sensor Data Set, and NAB Data Set

Model	Datasets															
	Air Quality Dataset				Energy Consumption Dataset				Environment Sensor Dataset				NAB Dataset			
	Accuracy	Recall	F1 Score	AUC	Accuracy	Recall	F1 Score	AUC	Accuracy	Recall	F1 Score	AUC	Accuracy	Recall	F1 Score	AUC
Zhang et al.(Zheng and Ge 2022)	89.46	88.26	85.33	94.90	89.45	92.32	84.32	92.78	95.75	92.58	87.45	89.42	85.23	87.02	87.98	90.69
Cai et al.(Cai and Wu 2022)	87.23	87.57	87.35	92.45	96.35	90.32	89.45	83.55	95.43	92.35	88.37	95.85	95.88	87.77	90.34	87.98
Huo et al.(Huo, Xu et al. 2021)	91.27	89.25	88.67	92.74	93.45	93.78	93.23	89.46	85.12	85.78	89.22	84.31	92.98	85.15	85.56	94.03
Gao et al.(Gao, Yang et al. 2021)	89.26	93.18	86.78	86.78	88.78	86.78	85.46	89.62	93.85	91.79	89.47	89.63	88.26	89.78	85.23	84.48
Zhou et al.(Zhou, Zeng et al. 2021)	90.38	94.28	84.83	85.53	95.78	94.56	85.23	89.78	92.27	88.24	91.78	93.86	85.48	85.65	87.48	88.78
Huang et al. (Huang and He 2020)	93.44	93.36	89.78	87.26	87.90	92.32	83.25	89.72	94.59	89.37	88.44	94.98	93.34	92.89	87.86	86.54
Ours	94.53	95.20	93.29	96.89	94.58	95.53	94.46	95.43	96.06	95.61	92.38	94.39	95.33	93.31	93.42	92.76

anomaly detection tasks across diverse domains, emphasizing its significant advantages in improving accuracy, recall, and F1 score. In summary, the model proposed in this study achieved impressive results on multiple data sets, providing a robust solution for anomaly detection in the context of carbon neutrality. These findings are not only essential for advancing carbon neutrality technologies but also contribute to the realization of environmental protection and sustainable development goals.

As shown in Table 2, we primarily focus on two important performance metrics: model parameter count (parameters) and computational complexity (flops). First, we can observe that across all data sets, our method (ours) demonstrates a significant advantage in terms of model parameter count. Specifically, compared to the closest competitors, our model has a relatively lower number of parameters. For example, on the air quality data set, our method utilizes only 116.45 million parameters, whereas the best-performing competitor (Zhang et al., 2023) uses 250.48 million parameters.

In comparison, our method reduces the parameter count by nearly half. Similarly, on other data sets, our method maintains relatively lower parameter counts, indicating higher model efficiency. Second, our method also exhibits a significant advantage in terms of computational complexity (flops). Across all data sets, the computational complexity of our method is noticeably lower than that of competitors. For instance, on the energy consumption data set, our method has a computational complexity of 40.32 billion flops, while the closest competitor (Cai et al.) has a computational complexity of 55.45 billion flops. This implies that our method excels in computational efficiency. According to the data in the table, our method demonstrates a clear advantage in both model parameter count and computational complexity, highlighting the outstanding performance of our approach in model lightweight and computational efficiency. These advantages would contribute to improving the performance and efficiency of the model in practical applications. Finally, to showcase these

Figure 5. Comparison of Model Performance on Different Data Sets

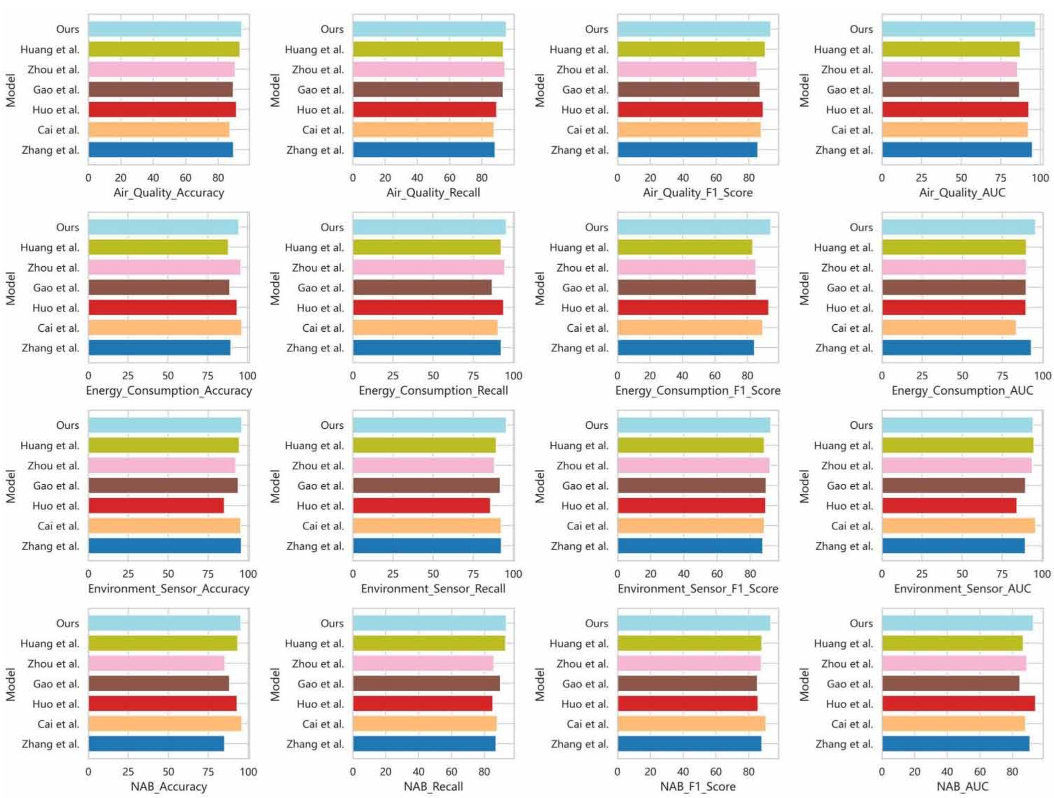
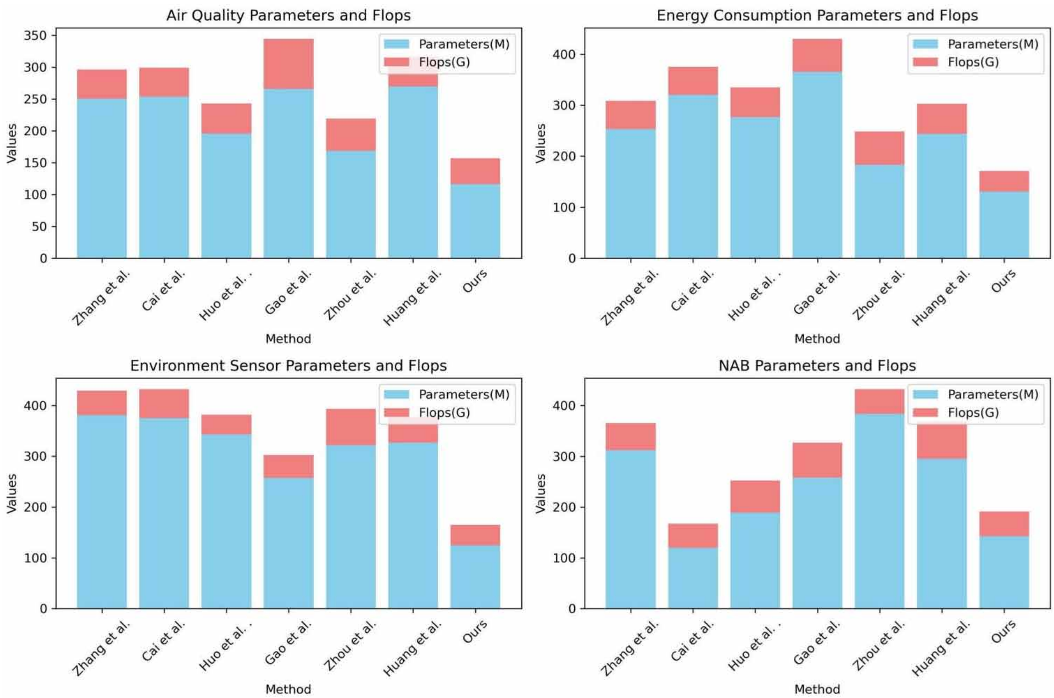


Table 2. The Comparison of Different Models in Different Indicators Comes From the Air Quality Data Set, Energy Consumption Data Set, Environment Sensor Data Set, and NAB Data Set

Method	Datasets							
	Air Quality Dataset		Energy Consumption Dataset		Environment Sensor Dataset		NAB Dataset	
	Parameters(M)	Flops(G)	Parameters(M)	Flops(G)	Parameters(M)	Flops(G)	Parameters(M)	Flops(G)
Zhang et al.(Zheng and Ge 2022)	250.48	45.78	253.54	55.23	381.32	47.79	312.22	53.53
Cai et al. (Cai and Wu 2022)	253.72	45.33	320.45	55.45	375.47	56.38	119.73	47.58
Huo et al.(Huo, Xu et al. 2021)	195.65	47.45	276.58	58.89	342.78	38.93	189.14	63.11
Gao et al.(Gao, Yang et al. 2021)	265.88	78.67	365.89	64.78	257.25	45.25	257.95	68.75
Zhou et al.(Zhou, Zeng et al. 2021)	168.44	50.85	183.35	65.25	321.79	71.54	383.71	48.42
Huang et al. (Huang and He 2020)	269.56	46.53	244.15	59.15	326.42	50.53	295.36	73.04
Ours	116.45	40.30	130.53	40.32	124.38	40.32	142.45	48.79

Figure 6. Comparison of Different Indicators of Different Models



advantages more intuitively, we can use visualization tools, such as Figure 6, to present the table content visually, further emphasizing the superior performance of our method.

As shown in Table 3, we conducted ablation experiments on the BIGRU model using different data sets and evaluated its performance metrics, including accuracy, recall, F1 score, and AUC. These experiments aimed to investigate the performance differences of different models (GRU, Bilstm, LSTM, BIGRU) across multiple data sets, providing a deeper understanding of their performance variances. First, we can observe that the BIGRU model exhibits significant advantages in almost all performance metrics across multiple datasets. Taking the air quality data set as an example, the BIGRU model achieves an accuracy of 94.53%, while in comparison, the GRU model only achieves an accuracy of 90.33%. This indicates that on this data set, our approach improves accuracy by 4.2 percentage points compared to the GRU model. Similarly, on other data sets, such as the energy consumption data set, environment sensor data set, and NAB data set, the BIGRU model also shows similar significant performance improvements.

This result clearly demonstrates the outstanding performance of the BIGRU model across all performance metrics. Furthermore, not only in terms of accuracy, the BIGRU model also exhibits superior performance in metrics like recall, F1 score, and AUC. For instance, on the energy consumption data set, the BIGRU model's recall and F1 score are significantly higher than other models, at 93.32% and 92.11%, respectively, whereas the corresponding values for the GRU model are only 89.17% and 88.23%. This indicates that our approach has a stronger ability to detect positive instances while also achieving significant improvements in precision and AUC. Finally, to visually demonstrate these advantages, we can refer to Figure 7, which provides a graphical representation of the table's content. It can be observed that the BIGRU model consistently exhibits higher performance across all datasets, reinforcing the excellence of our approach. These experimental results indicate that the BIGRU model outperforms other models in all aspects, providing strong evidence for model selection and application.

Table 3. Ablation Experiments on the BIGRU Module Using Different Data Sets

Model	Datasets															
	Air Quality Dataset				Energy Consumption Dataset				Environment Sensor Dataset				NAB Dataset			
	Accuracy	Recall	F1 Score	AUC	Accuracy	Recall	F1 Score	AUC	Accuracy	Recall	F1 Score	AUC	Accuracy	Recall	F1 Score	AUC
GRU	90.33	92.75	89.32	88.48	92.23	89.17	88.23	91.86	91.45	90.48	92.12	93.54	91.78	90.86	91.79	91.49
Bilstm	94.52	93.22	90.00	90.42	92.10	92.23	91.23	91.23	90.56	89.45	94.36	92.78	90.62	90.32	92.03	91.37
Lstm	93.33	92.34	89.32	90.57	92.66	90.36	89.78	90.36	90.36	93.76	93.23	89.86	91.15	91.75	83.96	87.74
BIGRU	94.53	93.32	92.11	91.11	93.12	92.78	92.36	92.72	91.35	95.91	94.75	93.76	92.52	92.67	93.50	92.33

Figure 7. Comparison of Model Performance on Different Data Sets

As shown in Table 4, we have summarized the results of ablation experiments conducted using the TPA model on various data sets. A comparison of four different models (Cross-AM, Multi-Head-AM, Dynamic-AM, TPA) has been made regarding their performance on these data sets. Among these data sets, the TPA model has demonstrated excellent performance across multiple key performance metrics. Taking the air quality data set as an example, the TPA model achieved an accuracy of 94.53%, while the other models performed comparatively worse. This implies that the TPA model excels in correct classification. Furthermore, in terms of recall, F1 score, and AUC, the TPA model has also shown a significant advantage, further emphasizing its outstanding performance across multiple data sets. To present these results more visually, we have included Figure 8, which graphically illustrates the performance comparison of different models on various data sets. It is evident from the graph that the TPA model outperforms all other models on all data sets, solidifying its leading position in terms of overall performance. These ablation experiment results clearly indicate that our model performs exceptionally well on multiple data sets, providing strong evidence for its selection in practical applications.

Table 4. Ablation Experiments on the TPA Module Using Different Data Sets

Model	Dataset															
	Air Quality Dataset				Energy Consumption Dataset				Environment Sensor Dataset				NAB Dataset			
	Accuracy	Recall	F1 Score	AUC	Accuracy	Recall	F1 Score	AUC	Accuracy	Recall	F1 Score	AUC	Accuracy	Recall	F1 Score	AUC
Cross-AM	88.33	92.75	88.32	87.48	92.23	88.17	85.23	91.86	91.45	89.48	93.12	93.54	92.78	89.86	91.79	91.49
Multi-Head-AM	91.52	93.72	92.42	87.42	93.33	92.23	85.23	82.23	93.56	84.45	85.36	83.78	84.62	84.32	84.03	82.37
Dynamic-AM	92.33	92.34	83.32	83.57	84.66	94.36	85.78	85.36	89.36	93.76	92.23	82.86	91.15	90.75	86.96	86.74
TPA	94.53	93.32	92.11	91.11	93.12	92.78	91.36	92.72	91.35	95.91	94.75	93.76	92.52	91.67	92.50	91.33

Table 5. Ablation Experiments on the Resnet Module Using Different Data Sets

Model	Datasets															
	Air Quality Dataset				Energy Consumption Dataset				Environment Sensor Dataset				NAB Dataset			
	Accuracy	Recall	F1 Score	AUC	Accuracy	Recall	F1 Score	AUC	Accuracy	Recall	F1 Score	AUC	Accuracy	Recall	F1 Score	AUC
AlexNet	92.35	91.76	90.33	90.41	93.00	90.25	90.23	91.75	90.45	92.48	93.12	94.54	92.75	91.86	90.79	92.49
GoogLeNet	93.53	92.23	91.00	90.45	92.11	91.25	90.25	91.28	91.56	93.45	94.38	93.78	91.62	90.32	92.45	92.37
EfficientNet	93.55	93.30	90.33	91.10	91.65	91.36	90.78	90.38	90.36	9.76	94.23	92.86	92.15	91.75	90.96	90.74
ResNeXt	94.53	93.32	92.11	91.11	93.12	92.78	92.36	92.72	91.35	95.91	94.75	93.76	92.52	92.67	93.50	92.33

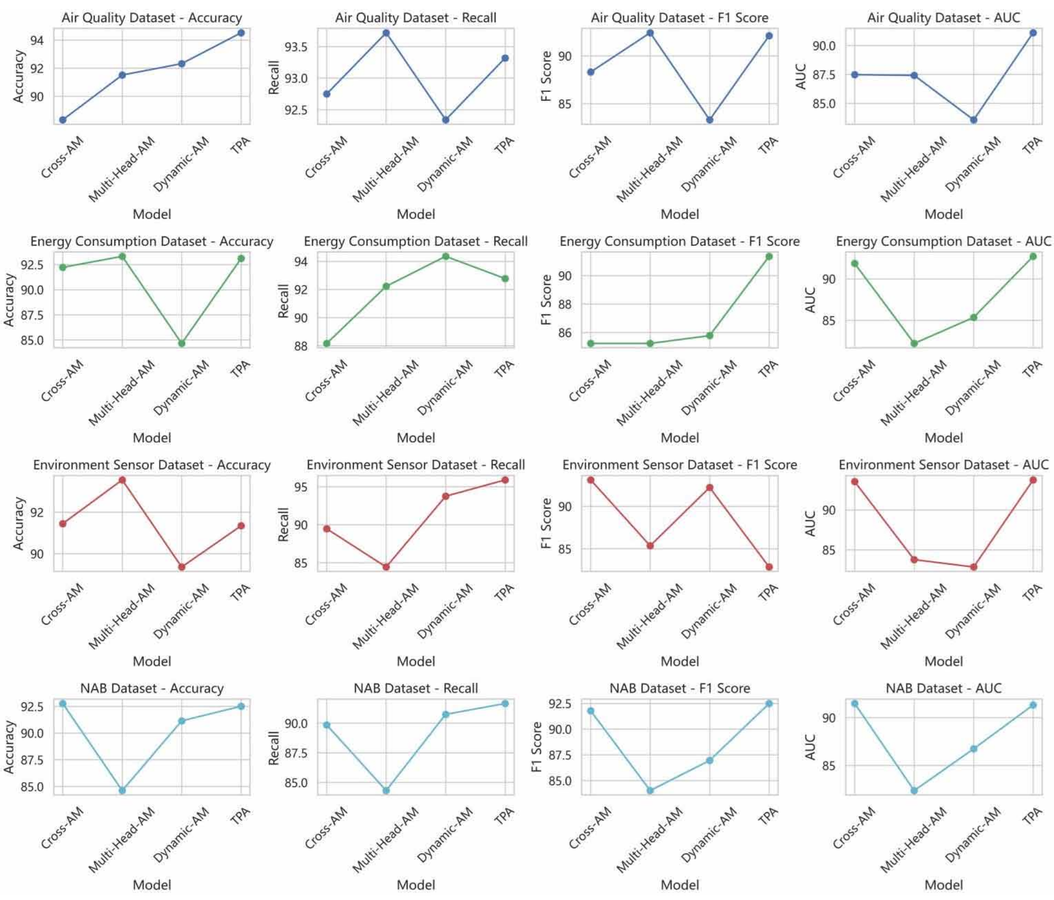
As shown in Table 5, we conducted ablation experiments on the ResNet module using different data sets to evaluate the performance of our method in the context of carbon offsetting anomaly detection. In these experiments, we applied different models (AlexNet, GoogLeNet, EfficientNet, and ResNeXt) to four distinct data sets (air quality data set, energy consumption data set, environment sensor data set, and NAB data set) and measured their performance in terms of accuracy, recall, F1 score, and AUC, among other performance metrics.

By comparing these performance metrics, we can clearly observe that our ResNeXt model excels in most cases. Particularly, on the air quality data set and energy consumption data set, ResNeXt achieves accuracy rates of 94.53% and 93.12%, respectively, which are significantly better than other models. Additionally, in terms of recall, F1 score, and AUC, ResNeXt also demonstrates a substantial advantage. It is worth noting that while other models such as GoogLeNet and EfficientNet perform well on some data sets, their overall performance across multiple data sets falls short of ResNeXt. This indicates the superior performance of our proposed ResNeXt model in handling carbon offsetting anomaly detection tasks.

CONCLUSION AND DISCUSSION

In this study, we have provided a detailed introduction to the ResNet-BIGRU-TPA network, an innovative model that integrates deep learning with time series attention mechanisms. Our research

Figure 8. Comparison of Model Performance on Different Data Sets



primarily focuses on addressing the challenges of anomaly detection in the carbon offsetting domain, particularly aiming to enhance the accuracy of detecting diverse and complex anomalous events. Regarding the model experiments, we conducted extensive testing and validation of the ResNet-BIGRU-TPA network using four different data sets. The experimental results demonstrate that our model performs exceptionally well in the task of anomaly detection. The ResNet component effectively extracts spatial distribution-related information associated with anomalous events, the BIGRU component successfully captures temporal dependencies within time series data, and the TPA mechanism enables our model to identify anomalies more precisely. These experimental findings highlight the significant potential of our model in addressing anomaly detection challenges.

However, our model still has certain limitations. First, the complexity of the model may require higher computational resources during both training and deployment, which may not be practical for certain application scenarios. Second, although our model performed well across multiple datasets, its generalization ability still needs improvement, especially when dealing with unknown domains or data distributions. Addressing these challenges will be a focal point of future research. Looking ahead to future work, we plan to enhance our model through various avenues. First, we will explore model lightweighting and optimization techniques to reduce computational resource demands, making it more suitable for real-world applications. Second, we will concentrate on improving the model's generalization capability, including adapting it to new domains and different data distributions. Finally, we will continue to refine the TPA mechanism to better model time series data. The significance of this study lies in providing a promising novel model for anomaly detection in the carbon offsetting domain and laying out directions for future improvements and research. Through ongoing efforts, we hope to extend the application of this research to a broader range of domains and contribute further to the reduction of greenhouse gas emissions, climate change mitigation, and the achievement of sustainable development goals.

FUNDINGS

This work was sponsored in part by Hebei Natural Science Foundation (G2023502005) and the Fundamental Research Funds for the Central Universities(2021MS103).

REFERENCES

Alzubi, J. (2018). Machine learning from theory to algorithms: An overview. In *Journal of Physics: Conference Series.* IOP Publishing.

Amasyali, K., & El-Gohary, N. M. (2018). A review of data-driven building energy consumption prediction studies. *Renewable & Sustainable Energy Reviews*, *81*, 1192–1205. doi:10.1016/j.rser.2017.04.095

Berriel, R. F. (2017). Monthly energy consumption forecast: A deep learning approach. *2017 International Joint Conference on Neural Networks (IJCNN)*. doi:10.1109/IJCNN.2017.7966398

Cai, K., & Wu, L. (2022). Using grey Gompertz model to explore the carbon emission and its peak in 16 provinces of China. *Energy and Building*, *277*, 112545. doi:10.1016/j.enbuild.2022.112545

Chen, M., Liu, Q., Huang, S., & Dang, C. (2022): Environmental cost control system of manufacturing enterprises using artificial intelligence based on value chain of circular economy. Enterprise Information Systems. 16(8-9), 1268–1287. doi:10.1080/17517575.2020.1856422

Chen, Y., Chen, X., Xu, A., Sun, Q., & Peng, X. (2022). A hybrid CNN-transformer model for ozone concentration prediction. *Air Quality, Atmosphere & Health*, *15*(9), 1533–1546. doi:10.1007/s11869-022-01197-w

Elmaz, F., Eyckerman, R., Casteels, W., Latré, S., & Hellinckx, P. (2021). CNN-LSTM architecture for predictive indoor temperature modeling. *Building and Environment*, *206*, 108327. doi:10.1016/j.buildenv.2021.108327

Feng, Z., & Chen, M. (2022). Platformance-based cross-border import retail e-commerce service quality evaluation using an artificial neural network analysis. *Journal of Global Information Management*, *30*(11), 1–17. Advance online publication. doi:10.4018/JGIM.306271

Gao, M., Yang, H., Xiao, Q., & Goh, M. (2021). A novel fractional grey Riccati model for carbon emission prediction. *Journal of Cleaner Production*, *282*, 124471. doi:10.1016/j.jclepro.2020.124471

Han, Y., Cao, L., Geng, Z., Ping, W., Zuo, X., Fan, J., Wan, J., & Lu, G. (2023). Novel economy and carbon emissions prediction model of different countries or regions in the world for energy optimization using improved residual neural network. *The Science of the Total Environment*, *860*, 160410. doi:10.1016/j.scitotenv.2022.160410 PMID:36427740

Huang, Y., & He, Z. (2020). Carbon price forecasting with optimization prediction method based on unstructured combination. *The Science of the Total Environment*, *725*, 138350. doi:10.1016/j.scitotenv.2020.138350 PMID:32304964

Huo, T., Xu, L., Feng, W., Cai, W., & Liu, B. (2021). Dynamic scenario simulations of carbon emission peak in China's city-scale urban residential building sector through 2050. *Energy Policy*, *159*, 112612. doi:10.1016/j.enpol.2021.112612

Lavin, A., & Ahmad, S. (2015). Evaluating real-time anomaly detection algorithms: The Numenta anomaly benchmark. *2015 IEEE 14th international conference on machine learning and applications (ICMLA).*

Li, R., Wang, Q., Liu, Y., & Jiang, R. (2021). Per-capita carbon emissions in 147 countries: The effect of economic, energy, social, and trade structural changes. *Sustainable Production and Consumption*, *27*, 1149–1164. doi:10.1016/j.spc.2021.02.031

Li, Y., Zuo, X., Zuo, W., Liang, S., Zhang, Y., & Zhu, Y. (2023). Causality extraction based on BERT-GCN. *Journal of Jilin University: Science Edition*, *61*, 325–330. doi:10.13413/j.cnki.jdxblxb.2022021

Liu, B., Wang, S., Liang, X., & Han, Z. (2023). Carbon emission reduction prediction of new energy vehicles in China based on GRA-BiLSTM model. *Atmospheric Pollution Research*, *14*(9), 101865. doi:10.1016/j.apr.2023.101865

Liu, H., Sun, C., He, H., & Hui, K. (2023). Metal surface defect detection method YOLOv3I. *Journal of Jilin University: Science Edition*, *61*, 612–622. doi:10.13413/j.cnki.jdxblxb.2022151

Lu, C., Chen, M., & Tian, G. (2022). Spatial-temporal evolution and influencing factors of urban green innovation efficiency in China. *Journal of Environmental and Public Health*. .10.1155/2022/4047572

Lv, Z., Wang, N., Lou, R., Tian, Y., & Guizani, M. (2023). Towards carbon neutrality: Prediction of wave energy based on improved GRU in maritime transportation. *Applied Energy*, *331*, 120394. doi:10.1016/j.apenergy.2022.120394

Monacchi, A. (2014). GREEND: An energy consumption dataset of households in Italy and Austria. *2014 IEEE International Conference on Smart Grid Communications (SmartGridComm)*. doi:10.1109/SmartGridComm.2014.7007698

Movassagh, A. A. (2021). Artificial neural networks training algorithm integrating invasive weed optimization with differential evolutionary model. *Journal of Ambient Intelligence and Humanized Computing*, 1–9.

Oyando, H. C., Kanyolo, T. N., & Chang, C. (2023). RNN-based main transformer OLTC control for SMR integration into a high renewable energy penetrated grid. *Journal of Electrical Engineering & Technology*, *18*(4), 1–13. doi:10.1007/s42835-022-01354-2

Pipattanasomporn, M., Chitalia, G., Songsiri, J., Aswakul, C., Pora, W., Suwankawin, S., Audomvongseree, K., & Hoonchareon, N. (2020). CU-BEMS, smart building electricity consumption and indoor environmental sensor datasets. *Scientific Data*, *7*(1), 241. doi:10.1038/s41597-020-00582-3 PMID:32686680

Sethi, J. K., & Mittal, M. (2019). A new feature selection method based on machine learning technique for air quality dataset. *Journal of Statistics and Management Systems*, *22*(4), 697–705. doi:10.1080/09720510.2019.1609726

Shen, Z. (2022). Federated learning for long-term forecasting of electricity consumption towards a carbon-neutral future. *2022 7th International Conference on Intelligent Computing and Signal Processing (ICSP).*

Shih, S. Y., Sun, F.-K., & Lee, H. (2019). Temporal pattern attention for multivariate time series forecasting. *Machine Learning*, *108*(8-9), 1421–1441. doi:10.1007/s10994-019-05815-0

Somu, N., Raman M R, G., & Ramamritham, K. (2021). A deep learning framework for building energy consumption forecast. *Renewable & Sustainable Energy Reviews*, *137*, 110591. doi:10.1016/j.rser.2020.110591

Sun, W., & Ren, C. (2021). Short-term prediction of carbon emissions based on the EEMD-PSOBP model. *Environmental Science and Pollution Research International*, *28*(40), 56580–56594. doi:10.1007/s11356-021-14591-1 PMID:34060019

Waheed, R., Sarwar, S., & Wei, C. (2019). The survey of economic growth, energy consumption and carbon emission. *Energy Reports*, *5*, 1103–1115. doi:10.1016/j.egyr.2019.07.006

Wang, J., Sun, X., Cheng, Q., & Cui, Q. (2021). An innovative random forest-based nonlinear ensemble paradigm of improved feature extraction and deep learning for carbon price forecasting. *The Science of the Total Environment*, *762*, 143099. doi:10.1016/j.scitotenv.2020.143099 PMID:33127140

Yang, W., Huang, B., Zhang, A., Li, Q., Li, J., & Xue, X. (2022). Condition prediction of submarine cable based on CNN-BiGRU integrating attention mechanism. *Frontiers in Energy Research*, *10*, 1023822. doi:10.3389/fenrg.2022.1023822

Zhang, K., Yang, X., Wang, T., Thé, J., Tan, Z., & Yu, H. (2023). Multi-step carbon price forecasting using a hybrid model based on multivariate decomposition strategy and deep learning algorithms. *Journal of Cleaner Production*, *405*, 136959. doi:10.1016/j.jclepro.2023.136959

Zhao, X., Ma, X., Chen, B., Shang, Y., & Song, M. (2022). Challenges toward carbon neutrality in China: Strategies and countermeasures. *Resources, Conservation and Recycling*, *176*, 105959. doi:10.1016/j.resconrec.2021.105959

Zheng, H., & Ge, L. (2022). Carbon emissions reduction effects of sustainable development policy in resource-based cities from the perspective of resource dependence: Theory and Chinese experience. *Resources Policy*, *78*, 102799. doi:10.1016/j.resourpol.2022.102799

Zhou, W., Zeng, B., Wang, J., Luo, X., & Liu, X. (2021). Forecasting Chinese carbon emissions using a novel grey rolling prediction model. *Chaos, Solitons, and Fractals*, *147*, 110968. doi:10.1016/j.chaos.2021.110968