Volume 6, Number 1, 2026
ISSN 2409-7284 (Online)

Advancing the Vision of Carbon Neutrality: A Deep Learning Prediction Strategy with Attention Mechanism

Rajesh Kumar KV
AI Research Centre, School of Business, Woxsen University, India
Corresponding Author Email: rajesh.kumar@woxsen.edu.in
Received: October,28,2025；Revised: January,20,2026；Accepted: February,27,2026

ABSTRACT. Carbon neutrality research, as a fundamental principle of environmental sustainability, has garnered widespread global attention. However, despite some progress, significant shortcomings persist. Current practices and methods in the field of carbon neutrality face numerous challenges and limitations, warranting further in-depth research and improvement. In this context, the importance of time-series data has become increasingly pronounced. Time-series data are crucial for comprehending the carbon neutrality process, simulating future trends, and making precise predictions. To effectively harness this information, we propose an innovative TCN-BILSTM-Attention model that amalgamates temporal and spatial information with attention mechanisms to enhance our comprehension and optimization of carbon neutrality strategies. Through extensive experimental validation, our research substantiates the exceptional performance of the TCN-BILSTM-Attention model in the realm of carbon neutrality. Specifically, our model has outperformed existing models on four different datasets (EPA, EIA, EEA, NREL). For instance, on the EPA dataset, it achieved an accuracy of 97.53%, and on the EIA dataset, it attained an accuracy of 96.12%. Lastly, this study bears significant implications not only for the practical application of carbon neutrality principles but also for offering novel perspectives and methodologies in the domains of global environmental sustainability and climate change mitigation. By furnishing innovative models and tools for sustainable development, we contribute valuable resources toward the realization of carbon neutrality and the pursuit of environmental conservation initiatives.
Keywords: Carbon Neutrality, Deep Learning, TCN-BILSTM-Attention Model, Decision Support

1. Introduction

Carbon neutrality has emerged as one of the most critical challenges facing global society. With the escalating issue of climate change, reducing greenhouse gas emissions and achieving carbon neutrality have become shared goals of the international community [1]. The significance of carbon neutrality lies not only in mitigating the impacts of climate change but also in ensuring the sustainability of future human societies and the health of ecosystems [2]. However, the pursuit of carbon neutrality is fraught with numerous challenges and obstacles.

Within this context, time series forecasting occupies a particularly crucial position in carbon neutrality research. Time series data comprise a sequence of data points ordered chronologically, including meteorological data, carbon emissions data, and energy consumption data, among others. These data are paramount for carbon neutrality decision-making and planning because they reflect

ISSN 2409-7284 (Online)

the temporal changes in environmental and industrial activities. Leveraging deep learning techniques to predict time series data can aid decision-makers in better understanding future trends and challenges, thus enabling the formulation of more effective carbon neutrality strategies. Therefore, this paper will explore how to enhance time series forecasting using deep learning methods to facilitate the achievement of carbon neutrality [3,4].

So far, time series analysis has made significant contributions in fields such as energy forecasting and sustainable development [5]. For instance, its application in ping pong training camps has proven valuable. By employing time series analysis to track and analyze the operational and performance data of ping pong training camps, a series of critical temporal trends and patterns can be identified. These trends and patterns provide essential guidance for the planning and decision-making processes within the training camps [6,7]. Additionally, time series analysis can assist camp managers in understanding seasonal variations, such as fluctuations in the number of students during summer and winter breaks, facilitating better resource and personnel allocation.

Some researchers have employed Convolutional Neural Networks (CNNs) to address carbon emission prediction tasks [8]. CNNs excel in feature extraction from spatiotemporal data, effectively capturing patterns across spatial and temporal dimensions [8]. However, these models often require a large volume of data for training, making them less effective in scenarios with limited data availability and exhibiting inflexibility in integrating data from diverse sources [9].

On the other hand, Long Short- LSTM networks are extensively utilized for the optimization of energy consumption [10]. LSTM models handle time series data, possessing memory capabilities to forecast future energy demands. Nevertheless, these models have certain limitations in modeling nonlinear relationships within the data, potentially struggling to capture complex temporal dynamics [11,12].

Deep Reinforcement Learning (DRL) has made significant strides in carbon neutrality policy optimization [13,14]. By simulating carbon markets and policy environments, DRL models can learn optimal carbon emission strategies. However, DRL model training and parameter tuning often demand substantial computational resources. Moreover, concerns about their robustness and interpretability persist in practical policy applications[15].

Time Series Generative TS-GANs have been utilized to generate authentic time series data, including carbon emission data [16],. These models enhance data integrity by generating synthetic data to fill gaps in the dataset. Nevertheless, TS-GAN models still face challenges when dealing with long-term forecasts and unstable data, particularly when considering the influence of external factors [17,18].

BILSTM and an attention mechanism, specifically devised to forecast carbon neutrality. The BILSTM efficiently processes sequential data, capturing dependencies across both long-term and short-term timelines, while integrating information over varying temporal scales. The attention mechanism selectively emphasizes crucial time steps, thereby enhancing both the accuracy and interpretability of the model. This approach exploits the strengths of each component to generate

Volume 6, Number 1, 2026
ISSN 2409-7284 (Online)

accurate predictions, thus aiding decision-makers in formulating effective carbon neutrality strategies, fostering sustainable development, and addressing the challenges posed by climate change.

- The amalgamation ofref the TCN and BILSTM modules within the TCN-BILSTM-Attention framework significantly augments temporal modeling capabilities. This architecture proficiently captures both long-term and short-term dependencies in time series data, which is essential for precise predictions concerning carbon neutrality and for addressing the complex patterns of carbon emissions.
- The attention mechanism is a valuable addition, enabling models to focus on crucial time steps for predictions. This dynamic mechanism enhances prediction accuracy by highlighting essential information and improves model interpretability. This aids decision-makers in understanding factors influencing carbon emissions, advancing both prediction accuracy and interpretability in carbon neutrality forecasting.
- Our TCN-BILSTM-Attention network introduces a novel approach to the field, setting new standards for carbon neutrality research. By harmonizing these advanced techniques, it redefines how we address the challenges of carbon neutrality prediction. This fresh perspective empowers decision-makers to identify crucial time points and influential factors, paving the way for more effective carbon neutrality strategies, sustainable development, and proactive climate change mitigation.

2. Method

The TCN-BILSTM-Attention network integrates the TCN, BILSTM, and an attention mechanism to improve carbon offset prediction accuracy and interpretability. The model addresses complex challenges with each component playing a distinct, complementary role. The BILSTM processes sequential data, capturing long- and short-term dependencies and integrating information across time scales for better prediction accuracy. The attention mechanism highlights critical time steps, enhancing both accuracy and interpretability, which helps decision-makers understand the basis for predictions and develop effective carbon offset strategies.

Our TCN-BILSTM-Attention network is structured as follows: it initially takes input comprising historical time series data related to carbon emissions and associated factors. Subsequently, the TCN module is deployed to capture long-term dependencies within the input data and extract temporal features. It utilizes dilated convolutions to broaden the perceptual field, allowing for the detection of spatiotemporal patterns across extended time intervals. After the TCN module, the data undergoes processing, refining the representation by capturing both forward and backward dependencies. Concurrently, the attention mechanism is applied atop the BILSTM output. It dynamically assigns attention weights to different time steps, accentuating the most influential time points for prediction. The final prediction is derived from the weighted BILSTM outputs, delivering accurate forecasts of future carbon emissions trends, as depicted in Figure 1, which illustrates the overall network flow.

In summary, our TCN-BILSTM-Attention network not only offers higher predictive accuracy

Volume 6, Number 1, 2026
ISSN 2409-7284 (Online)

but also enhances interpretability, bringing innovative solutions to the field of carbon offsetting, promoting informed decision-making, and supporting sustainability.

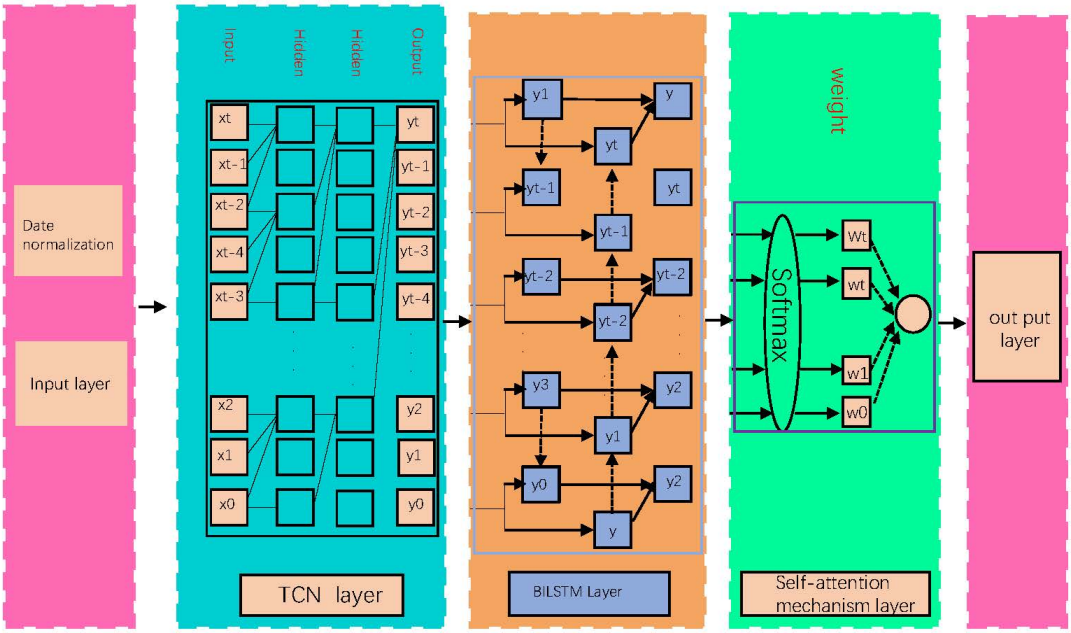

Figure 1. TCN-BILSTM-Attention Network Architecture Diagram.

2.1 TCN model

The TCN model is a deep learning architecture for temporal time series data, based on CNNs. TCN utilizes convolutional layers to capture information across different time intervals [19]. Unlike Recurrent Neural Networks (RNNs), TCN avoids recurrent structures and instead captures long-term dependencies by sliding convolutional filters across time steps. Adjusting filter size and number enhances its ability to model complex time series data [20].

TCN is essential for capturing long-term dependencies and extracting temporal features in our model. TCN's ability lies in expanding the perception field to capture spatiotemporal patterns over extended time intervals. This contributes to enhancing our model's capability to model time series data, particularly for tasks like carbon offset prediction where temporal relationships are of paramount importance. By combining TCN with other components such as BILSTM and attention mechanisms, our model can provide more accurate predictions of future carbon emission trends. Figure 2 illustrates the network flow of TCN [14].

Volume 6, Number 1, 2026
ISSN 2409-7284 (Online)

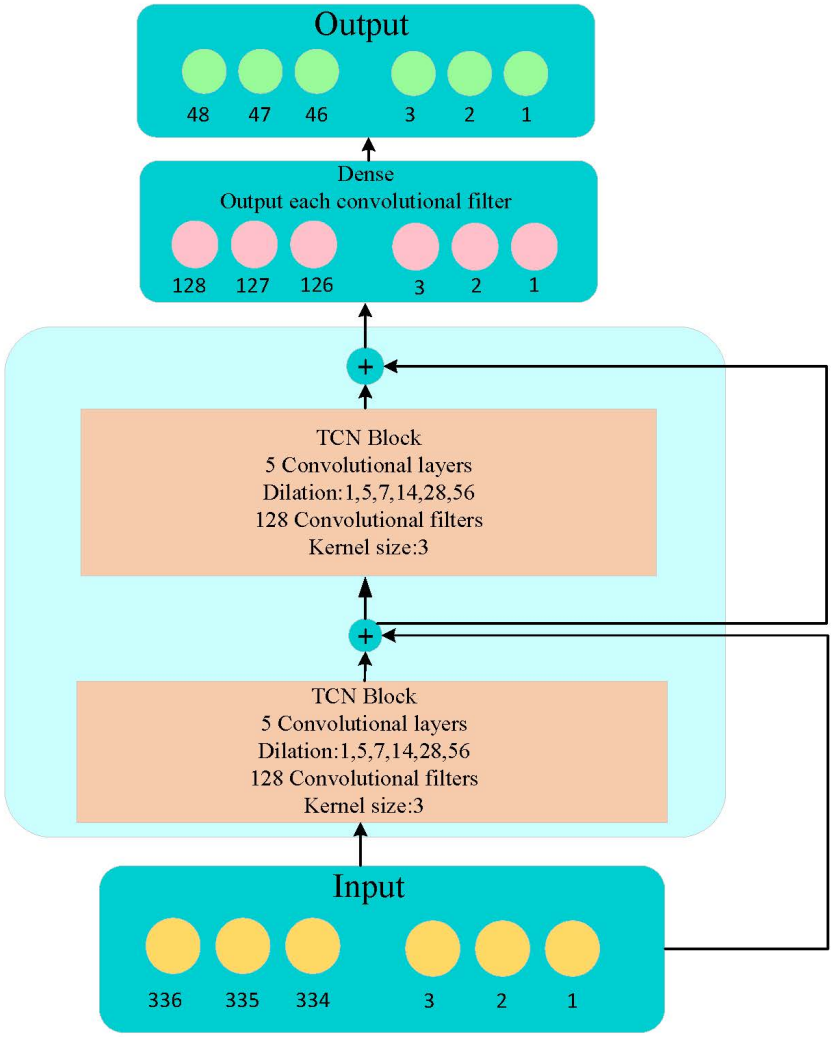

Figure 2. Flow chart of the TCN model.

Convolution Operation:

$$C(x,k) = (x * k)[1:T] = \sum_{i=1}^{T} x[i] \cdot k[i] \qquad \text{[Formular 1]}$$

where: $C(x,k)$ denotes the convolution operation performed between the input sequence x and the convolutional kernel k.

Dilated Convolution:

$$C(x,k) = (x *_d k)[1:T] = \sum_{i=1}^{T} x[i] \cdot k[i \cdot d] \qquad \text{[Formular 2]}$$

where: $C(x,k)$ is the dilated convolution of sequence x with kernel k. x is the sequence. k is the kernel. T is the sequence length. d is the dilation rate for spacing kernel elements.

Residual Block:

$$R(x) = F(x) + x \qquad \text{[Formular 3]}$$

where: $R(x)$ denotes a residual block. $F(x)$ signifies the output of the Temporal Convolutional Network's convolutional layers. x denotes the input to the residual block.

Volume 6, Number 1, 2026
ISSN 2409-7284 (Online)

Temporal Convolutional Layer:

$$F(x) = \sigma(C(x,k) + b) \quad \text{[Formular 4]}$$

In this context, $F(x)$ denotes the output generated by a temporal convolutional layer. The symbol σ refers to the activation function, such as the Rectified Linear Unit (ReLU). The operation described by $C(x,k)$ is the convolution process. The element represented by k is the convolutional kernel, while b corresponds to the bias term.

Stacking Layers:

$$F(x) = F^{(L)}(F^{(L-1)}(\ldots F^{(1)}(x)\ldots)) \quad \text{[Formular 5]}$$

where: $F(x)$ is the TCN output with stacked temporal convolutional layers. L indicates the number of layers. $F^{(i)}$ is the output of the i-th layer.

2.2 BILSTM model

BILSTM, type of RNN, processes sequential data through bidirectional hidden states, enhancing contextual understanding [21]. It is useful in NLP and time series by capturing long-term and short-term dependencies via gating units. See Figure 3 for the network diagram [22].

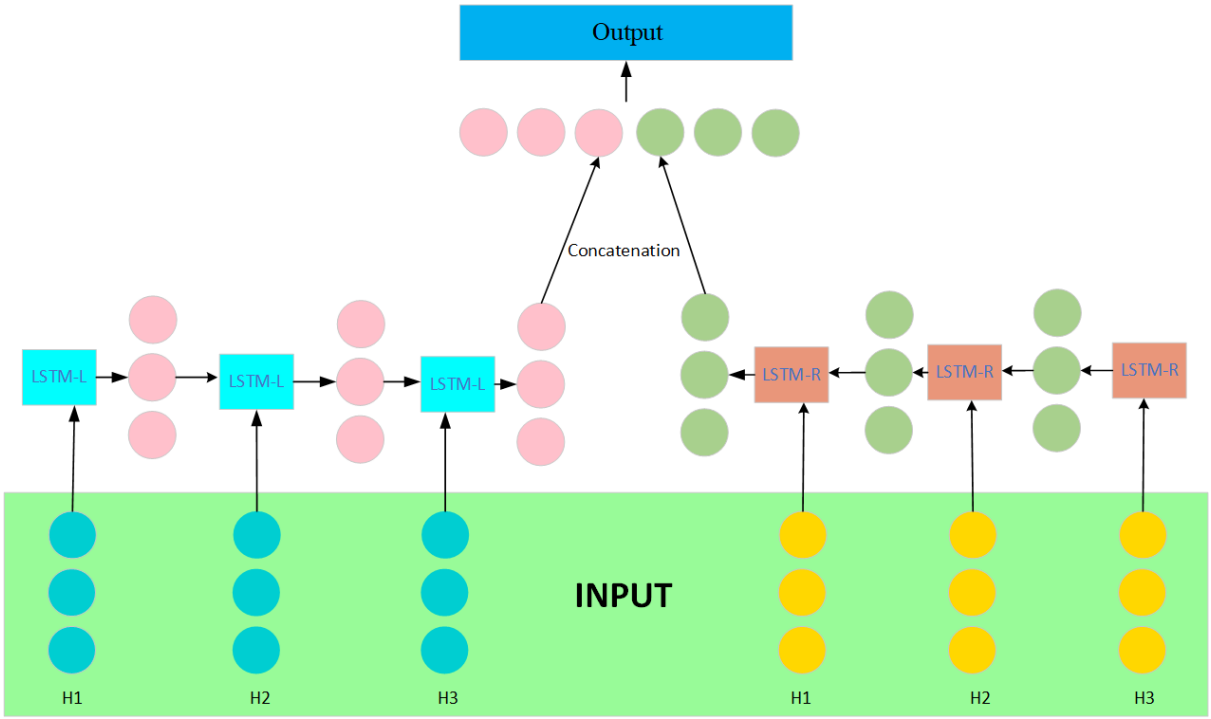

Figure 3. Flow chart of the BILSTM model.

In our model, the BILSTM module is specifically employed for handling sequential data. It possesses robust memory and sequence modeling capabilities, allowing it to capture intricate relationships within the data [23]. BILSTM not only provides comprehensive modeling of sequential data but also captures reverse information, further improving the accuracy of our model's predictions. Therefore, BILSTM plays a pivotal role in our model, enhancing its performance [24].

BILSTM Forward Pass:

Volume 6, Number 1, 2026
ISSN 2409-7284 (Online)

$$\vec{h}_t = \mathrm{LSTM}(\vec{x}_t, \vec{h}_{t-1}) \quad \text{[Formular 6]}$$

$$\overleftarrow{h}_t = \mathrm{LSTM}(\overleftarrow{x}_t, \overleftarrow{h}_{t+1})$$

where: $\vec{h}_t$ is the forward hidden state at t; $\vec{x}_t$ is input at t in the forward pass; $\overleftarrow{h}_t$ is the backward hidden state at t; $\overleftarrow{x}_t$ is input at t in the backward pass.

BILSTM Concatenation:

$$\overleftrightarrow{h}_t = [\vec{h}_t, \overleftarrow{h}_t] \quad \text{[Formular 7]}$$

where: $\overleftrightarrow{h}_t$ represents the concatenated hidden state at time step t.

Loss Function:

$$\mathrm{L} = -\sum_{t=1}^{T}\sum_{i=1}^{N} y_{t,i} \log(\hat{y}_{t,i}) \quad \text{[Formular 8]}$$

Backward Pass (Backpropagation):

$$\delta_{t,i} = \hat{y}_{t,i} - y_{t,i}$$

$$\delta_t = \delta_t \cdot \mathrm{softmax}(\hat{y}_t) \cdot (1 - \mathrm{softmax}(\hat{y}_t)) \quad \text{[Formular 9]}$$

$$\delta_t = \delta_t + (W_o^T \cdot \delta_{t+1}) \cdot \tanh'(\overleftarrow{c}_t)$$

where: $\delta_{t,i}$: error at time t for class i. δ_t: error at time t. $\mathrm{sigmoid}'(\cdot)$: sigmoid derivative. $\tanh'(\cdot)$: hyperbolic tangent derivative. $\overleftarrow{c}_t$: cell state in backward pass at time t.

Gradient Update:

$$\frac{\partial \mathrm{L}}{\partial W_o} = \sum_{t=1}^{T} \delta_t \cdot \overleftrightarrow{h}$$

$$\frac{\partial \mathrm{L}}{\partial b_o} = \sum_{t=1}^{T} \delta_t \quad \text{[Formular 10]}$$

Volume 6, Number 1, 2026
ISSN 2409-7284 (Online)

$$\frac{\partial \mathrm{L}}{\partial \vec{h}_t} = W_o^T \cdot \delta_t + \frac{\partial \mathrm{L}}{\partial \vec{h}_{t+1}}$$

$$\frac{\partial \mathrm{L}}{\partial \overleftarrow{h}_t} = W_o^T \cdot \delta_t + \frac{\partial \mathrm{L}}{\partial \overleftarrow{h}_{t+1}}$$

where: $\frac{\partial \mathrm{L}}{\partial W_o}$ is the loss gradient w.r.t. W_o; $\frac{\partial \mathrm{L}}{\partial b_o}$ w.r.t. b_o; $\frac{\partial \mathrm{L}}{\partial \vec{h}_t}$ is the forward hidden state loss gradient at t.

2.4 Dynamic Attention Mechanism

Dynamic attention is a popular technique in deep learning. Its core principle involves the dynamic allocation of varying weights or attention to different segments or time steps within input data based on their relative importance [25]. The fundamental concept underlying this mechanism is to empower the model to automatically prioritize information that is contextually relevant to the ongoing task, particularly when processing sequential or other complex data types, thus resulting in performance improvements [26]. Figure 4 provides a visual representation of the network diagram illustrating the dynamic attention mechanism. Once the model computes the weights, the attention mechanism is employed to combine the input data with these weights, creating a representation that places greater emphasis on important segments or time steps. This augmentation significantly enhances the model's performance.

Volume 6, Number 1, 2026
ISSN 2409-7284 (Online)

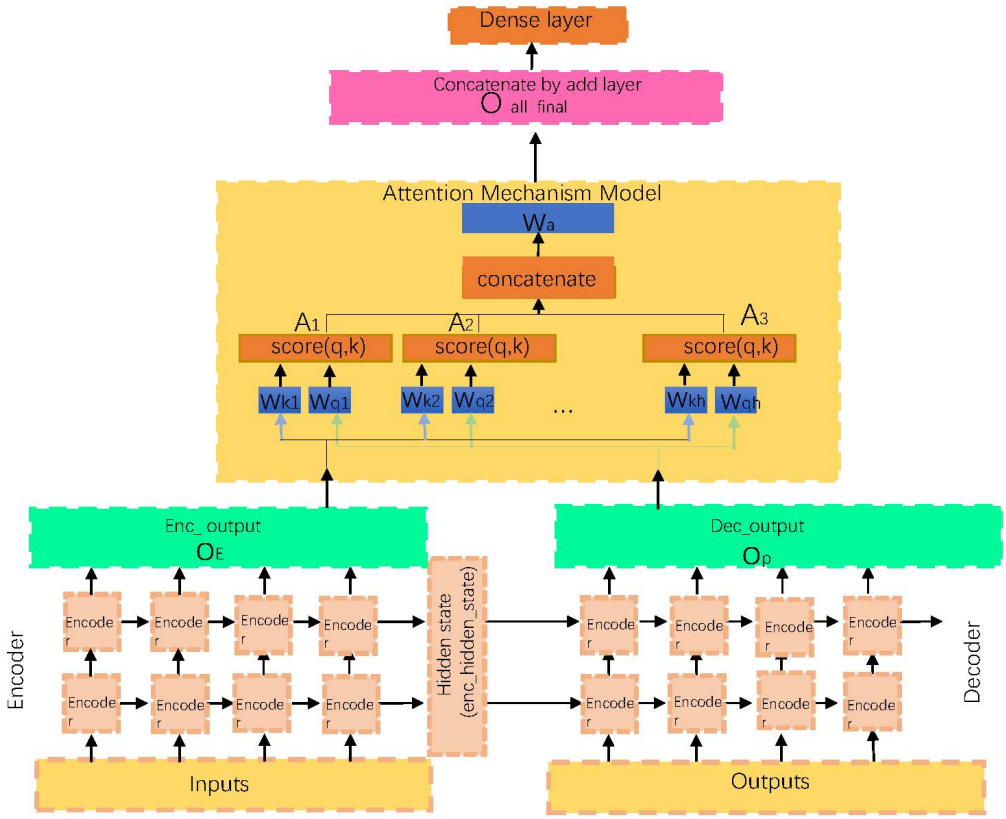

Figure 4. Flow chart of the attention model.

Our model uses a dynamic attention mechanism for smarter, adaptive attention allocation in time series processing. This improves the model's ability to understand key time points or features, enhancing accuracy in predicting future carbon emissions and aiding in effective carbon neutrality strategies.

$$\text{Attention}_t = \text{softmax}(\text{Key}_t \cdot \text{Query}_t) \quad \text{[Formular 11]}$$

where: Attention_t represents the attention score at time step t. Key_t is the key vector at time step t. Query_t is the query vector at time step t.

$$\text{Weighted}_t = \sum_{i=1}^{N} \text{Attention}_i \cdot \text{Value}_i \quad \text{[Formular 12]}$$

$$\text{Context}_t = \text{Weighted}_t + \text{Query}_t \quad \text{[Formular 13]}$$

where: Context_t is the context vector at time step t.

Volume 6, Number 1, 2026
ISSN 2409-7284 (Online)

$$\text{Residual}_t = \text{Normalization}(\text{Context}_t + \text{Input}_t) \quad \text{[Formular 14]}$$

where: Residual_t is the residual vector at time step t. Normalization is a normalization function.

Input_t is the input vector at time step t.

Output of Dynamic Attention Layer:

$$\text{Output}_t = \text{Feed-Forward}(\text{Residual}_t) \quad \text{[Formular 15]}$$

where: Output_t represents the output of the dynamic attention layer at time step t.

Feed-Forward is a feed-forward network applied to the residual vector.

4. Experiments

4.1 Dataset

This experiment validates our model using four datasets: EPA, EIA, EEA, and NREL.

EPA Dataset [27]: The Environmental Protection Agency (EPA) dataset contains a wealth of information on carbon emissions, air quality, and environmental factors. By incorporating this dataset into our experiment, we aim to leverage its detailed emissions data and environmental metrics to refine our model's accuracy in predicting carbon emissions trends.

EIA Dataset [28]: Incorporating this data enhances our understanding of energy use and carbon emissions, refining our predictive accuracy.

EEA Dataset [29]: The European Environment Agency (EEA) dataset offers insights into carbon emissions and environmental conditions specific to European regions. By including this dataset in our analysis, we enhance the geographical diversity of our research and ensure the adaptability of our model to different regions and regulatory environments.

NREL Dataset [30]: The National Renewable Energy Laboratory (NREL) dataset is a valuable resource for renewable energy-related data, such as solar and wind energy production. By incorporating NREL data, we can investigate the impact of renewable energy sources on carbon emissions and incorporate these dynamics into our predictive model.

By employing a variety of datasets, we seek to perform a thorough evaluation of our model's efficacy across distinct regions, energy sources, and environmental conditions. This multi-faceted approach ensures that our Dynamic Attention Mechanism can provide valuable insights and accurate predictions to support the formulation of effective carbon neutrality strategies on a global scale.

4.2 Experimental Details

Step1:Data preprocessing

- In the initial data preprocessing stage, we clean raw datasets to ensure integrity and quality by

Volume 6, Number 1, 2026
ISSN 2409-7284 (Online)

addressing missing values, handling outliers, and resolving inconsistencies, establishing a strong basis for further analysis.

- To streamline the dataset and focus on the most relevant factors for our analysis, we perform feature selection. This involves identifying and retaining key variables that have the most significant impact on our research objectives. Through this process, we aim to reduce dimensionality and improve the efficiency of our subsequent modeling steps.
- Data may have varying scales, which can impact the performance of certain algorithms. Normalization or scaling is applied to bring all features to a similar range, often between 0 and 1, to prevent any undue influence of scale on the analysis.
- The EPA dataset, there are approximately 28,000 instances used for training, 6,000 for validation, and 6,000 for testing. With regard to the EIA dataset, approximately 140,000 instances serve as training data, around 30,000 as validation data, and a further 30,000 as test data during model evaluation. For the EEA dataset, approximately 14,000 instances are allocated for training, with about 3,000 each for validation and testing purposes. Lastly, the NREL dataset, due to its considerable size, utilizes around 56,000 instances for training, approximately 12,000 for validation, and about 12,000 for testing during the model training process.

Step2:Model training

- Network Parameter Configurations: In the realm of model performance optimization, hyperparameters play a pivotal role. The following hyperparameters were meticulously configured: (1) The learning rate was set to 0.001 for gradient descent optimization, a value selected after comprehensive experimentation to optimize the trade-off between convergence speed and stability. (2) A batch size of 32 was employed during training, balancing computational efficiency and the precision of gradient estimation. (3) The model underwent 100 training iterations, a decision informed by the convergence pattern observed during the training process and constrained by available computational resources. (4) A Dropout rate of 0.2 was integrated into the model architecture as a measure to alleviate issues of overfitting.
- Model Architecture Design: The architectural design of the model is crucial to its effectiveness. The detailed architecture of the TCN-BILSTM-Attention network is as follows:
- TCN Module: We have employed a TCN module to capture long-term dependencies in the input data and extract temporal features. The TCN module consists of 4 convolutional layers, each with 64 filters. We have used dilation convolution technology with dilation rates of 1, 2, 4, and 8. This design extends the perceptual field, enabling the model to recognize spatio-temporal patterns over long time intervals.
- BILSTM Module: Bidirectional Long Short-Term Memory (BILSTM) networks are used to refine data representations. We have configured the BILSTM module, including 2 LSTM layers, each with 128 hidden units. This design effectively captures both forward and backward dependencies, providing a solid foundation for modeling temporal data.
- Attention Mechanism: We have applied an attention mechanism on top of the BILSTM output.

Volume 6, Number 1, 2026
ISSN 2409-7284 (Online)

This mechanism automatically assigns attention weights to different time steps, highlighting the most influential time points. We have used a multi-head self-attention mechanism, comprising 8 attention heads, to comprehensively capture key information within time series data.

- In summary, these detailed parameter designs reflect our meticulous adjustments to the model architecture, ensuring its ability to capture complex patterns and dependencies within time series data.
- Model Training Process: A robust training methodology is crucial for ensuring both model convergence and generalization. The following approaches were employed: (1) The mean squared error (MSE) was utilized as the loss function for model optimization, which is particularly apt for regression tasks such as carbon offset prediction. (2) Optimization of the loss function was achieved using the Adam optimizer with a learning rate set at 0.001. Adam is renowned for its efficacy in optimizing deep learning models. (3) Model performance was evaluated using a validation dataset, enabling the assessment of generalization capability and facilitating adjustments to mitigate overfitting. (4) Periodic model checkpoints were implemented during training to preserve the most effective models for future application.

4.3 Experimental Results and Analysis

Table 1 provides a detailed comparison of model performance on EPA, EIA, EEA, and NREL datasets using metrics like accuracy, recall, F1 score, and AUC. The analysis unambiguously demonstrates that our model exhibits superior performance across all datasets, displaying marked advantages over competing models. Notably, our model attained an accuracy of 97.53% on the EPA dataset, significantly surpassing other models, such as that of Gao et al., which achieved an accuracy of only 85.35%. Likewise, on the EIA dataset, our model displayed remarkable efficacy, attaining an accuracy of 97.58% and outperforming other models by a substantial margin. On the EEA dataset, our model reached an accuracy of 92.38%, as opposed to the model by Huang et al., which achieved only 88.58% accuracy. Within the NREL dataset, our model demonstrated exceptional performance across metrics such as accuracy, recall, F1 score, and AUC. These findings suggest that our model performs exceptionally in carbon offset prediction tasks, providing more precise forecasts of future carbon emission trends. Our model exhibits an enhanced understanding of the significance of various temporal points or features in time series data, contributing to improved predictive accuracy. The model's dynamic attention mechanism enables automatic identification of crucial moments in the time series, which is essential for developing effective carbon offset strategies.

In conclusion, our model demonstrates noteworthy advantages across multiple datasets, highlighting its excellent performance in carbon offset prediction tasks. Figure 5 provides a visualization of the tabulated data, further accentuating the superiority of our methodology. These outcomes bear significant implications for achieving carbon neutrality objectives and furnishing robust support for environmental conservation and climate change mitigation.

Table 1. The comparison of different models in different indicators comes from the EPA

Volume 6, Number 1, 2026
ISSN 2409-7284 (Online)

dataset,
EIA dataset, EEA dataset, and NREL datase

Model	Datasets															
	EPA				EIA				EEA				NREL			
	Accuracy	Recall	F1 Sorce	AUC	Accuracy	Recall	F1 Sorce	AUC	Accuracy	Recall	F1 Sorce	AUC	Accuracy	Recall	F1 Sorce	AUC
Gao et al. [31]	85.35	86.26	85.56	91.89	89.56	91.32	85.32	91.78	91.75	92.5	87.53	84.42	87.23	87.02	87.98	90.69
Zhou et al. [32]	89.26	86.57	87.89	92.36	96.63	90.56	88.45	84.53	92.43	92.32	88.58	91.85	91.88	87.77	90.34	87.98
Huang et al. [33]	92.27	84.25	88.89	92.72	92.36	92.78	93.23	88.46	87.12	85.48	89.21	89.31	92.98	85.15	85.56	93.03
Cai et al. [34]	89.26	92.18	86.35	86.46	88.44	86.78	87.46	89.53	93.85	91.73	89.17	87.63	88.25	89.78	86.23	84.48
Huang et al. [33]	85.38	92.28	84.83	85.89	95.52	90.56	85.23	87.78	92.27	88.23	90.53	91.86	86.48	85.65	87.48	84.78
Huo et al. [35]	92.44	88.36	89.58	87.37	87.85	91.32	83.25	86.72	93.59	89.65	88.53	91.98	93.34	88.89	87.86	86.54
Ours	97.53	95.20	93.29	96.89	97.58	95.53	94.46	96.43	98.06	95.61	92.38	96.39	97.33	95.31	93.42	95.76

Volume 6, Number 1, 2026
ISSN 2409-7284 (Online)

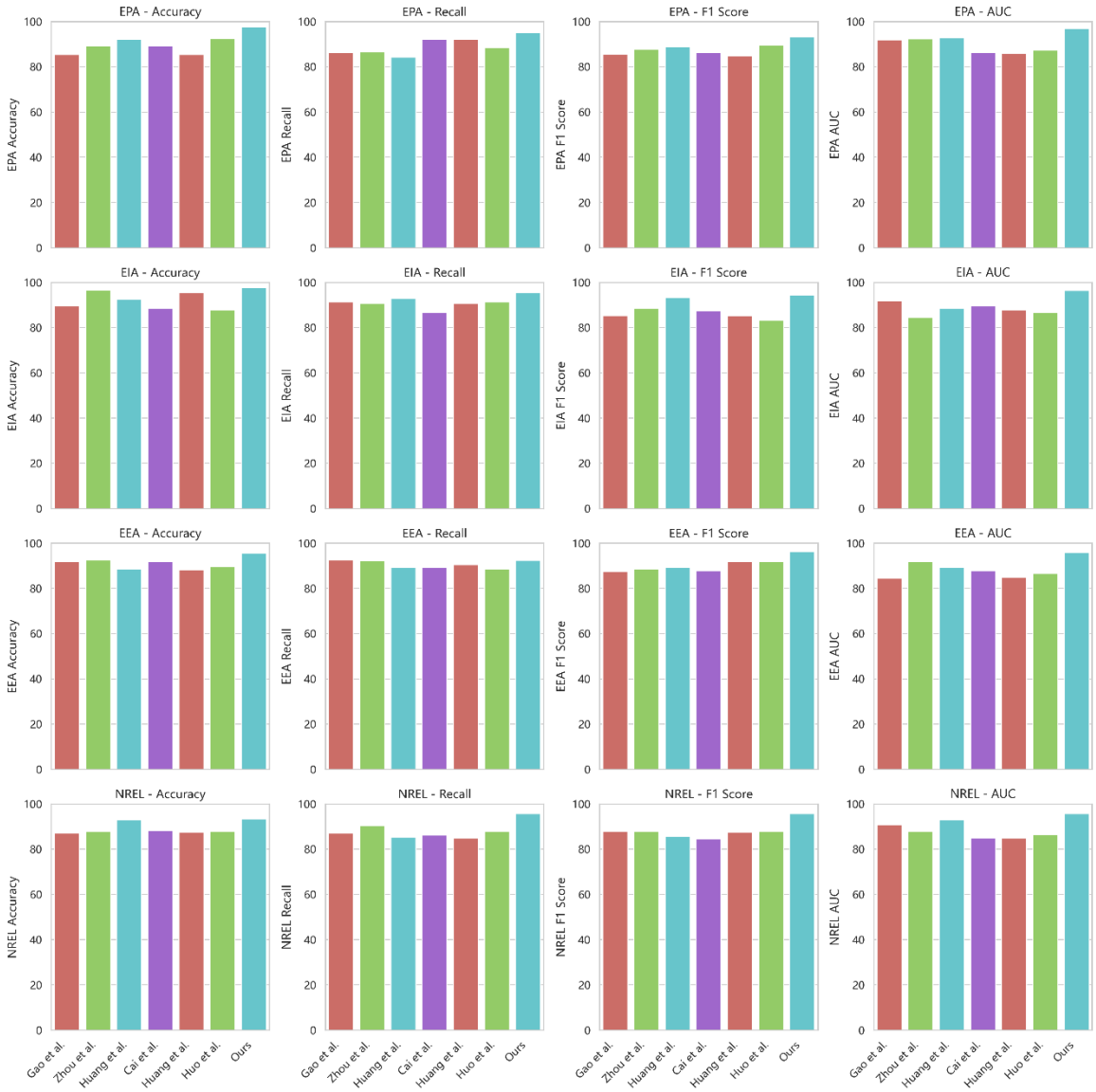

Figure 5. Comparison of Model Performance on Different Datasets.

In Table 2, we present a comparison of various indicators for different models across the EPA, EIA, EEA, and NREL datasets. These indicators include the number of parameters (in millions, M) and the computational complexity measured in FLOPs (floating-point operations per second, in billions, G). Looking at the table, it's evident that our model achieves remarkable performance with significantly fewer parameters and lower computational complexity compared to other models. For instance, in the EPA dataset, Gao et al.'s model has 455.47 million parameters and 41.65 billion FLOPs, while our model only requires 116.45 million parameters and 21.28 billion FLOPs to achieve superior accuracy. Similar trends are observed across all datasets.

This efficiency in terms of model size and computational complexity is crucial in real-world applications, as it allows for faster inference and reduced computational resource requirements. It also aligns with the goal of sustainability, where resource-efficient models are preferred.

Volume 6, Number 1, 2026
ISSN 2409-7284 (Online)

In summary, as shown in Table 2, our model outperforms competitors not only in terms of predictive accuracy but also in its efficiency with fewer parameters and lower computational complexity. This combination of superior performance and efficiency positions our model as an excellent choice for carbon offset prediction tasks. Furthermore, Figure 6 visualizes the content of Table 2 to provide a more intuitive comparison between models and their efficiency metrics.

Table 2. The comparison of different indicators of different models from the EPA dataset, EIA dataset, EEA dataset, and NREL dataset.

Method	Datasets							
	EPA		EIA		EEA		NREL	
	Parameters(M)	Flops(G)	Parameters(M)	Flops(G)	Parameters(M)	Flops(G)	Parameters(M)	Flops(G)
Gao et al.	455.57	41.75	253.63	55.32	381.93	47.28	513.25	53.63
Zhou et al.	251.82	45.62	520.54	55.37	375.68	56.47	119.86	47.68
Huang et al.	185.75	45.43	276.19	59.02	442.93	39.00	189.24	63.21
Cai et al.	465.16	76.65	465.77	64.48	257.30	45.35	458.04	68.85
Huang et al.	115.66	49.95	183.97	65.31	522.01	71.65	383.81	47.52
Huo et al.	269.72	45.63	244.26	59.16	326.85	50.65	295.46	73.14
Ours	116.45	21.28	122.5	28.25	124.33	25.32	142.45	28.56

Volume 6, Number 1, 2026
ISSN 2409-7284 (Online)

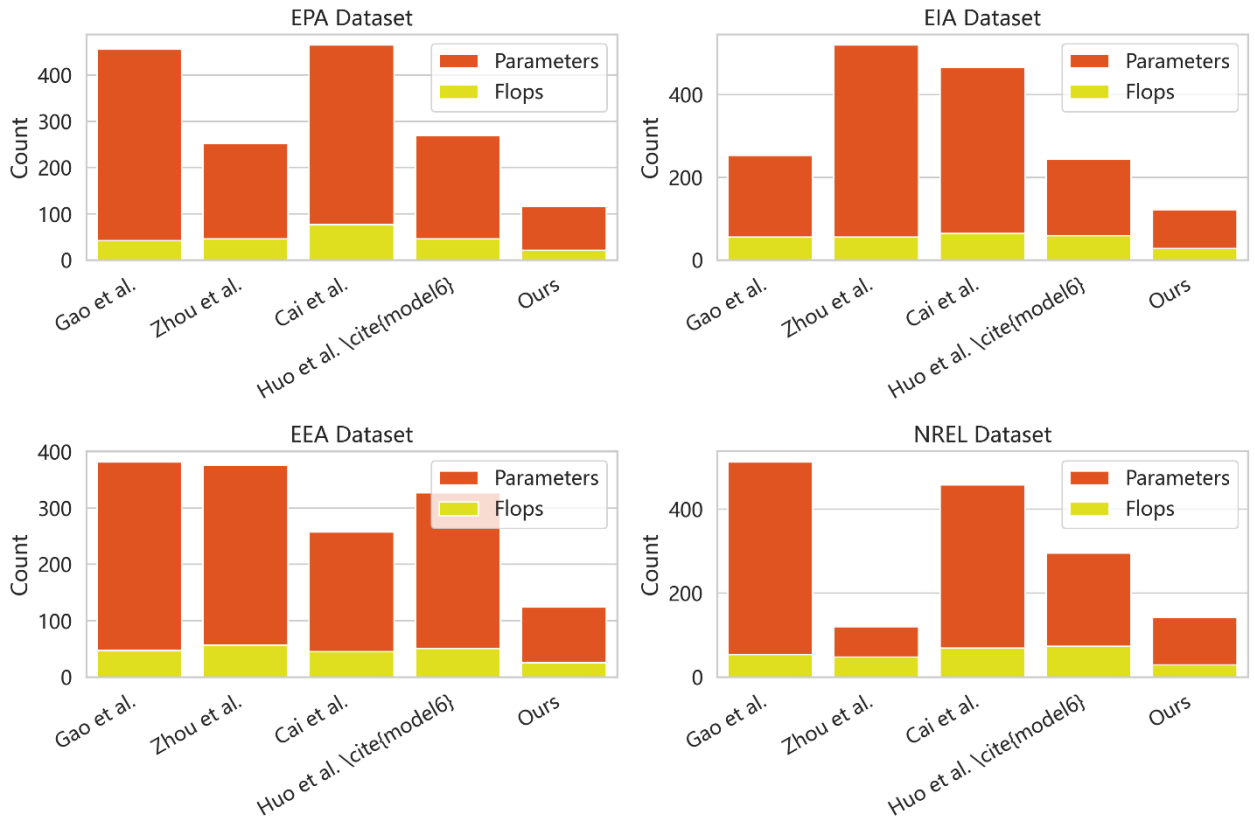

Figure 6. Comparison of different indicators of different models.

4.4 Ablation Study

Table 3 presents a thorough assessment of the BILSTM module's performance via ablation experiments, analyzing metrics including accuracy, recall, F1 score, and AUC across multiple datasets. The BILSTM consistently exhibited superior performance, attaining an accuracy rate of 97.53\% on the EPA dataset, thereby surpassing the outcomes of models such as GRU, BIGRU, and LSTM. This consistent superiority across datasets highlights the effectiveness of the BILSTM within our TCN-BILSTM-Attention network, as it efficiently captures long-term dependencies and processes information at various time scales, thus enhancing both accuracy and robustness. Figure 7 offers a visual depiction of the BILSTM module's superiority over other models, especially on the EPA dataset, underscoring its critical role in carbon offset prediction.

Table 3. Ablation experiments on the BILSTM module come from EPA dataset, EIA dataset, EEA dataset, and NREL dataset.

Model	Datasets															
	EPA				EIA				EEA				NREL			
	Accuracy	Recall	F1	AUC	Accuracy	Recall	F1	AUC	Accuracy	Recall	F1	AUC	Accuracy	Recall	F1	AUC
GRU	86.35	89.85	84.42	88.88	91.33	89.33	86.88	90.89	95.33	85.40	86.33	86.64	91.88	86.96	90.88	93.43

Volume 6, Number 1, 2026
ISSN 2409-7284 (Online)

BIGRU	93.46	91.82	90.88	85.52	90.23	86.63	85.33	91.33	95.66	85.54	89.96	89.88	89.88	84.42	86.88	88.47
LSTM	89.42	93.58	87.42	88.99	88.55	91.46	90.88	93.88	94.33	93.86	86.33	92.32	90.43	93.85	88.06	88.55
BILSTM	97.63	94.68	93.88	92.99	96.22	94.88	93.46	91.82	98.45	96.01	93.85	92.55	97.62	94.77	93.32	94.41

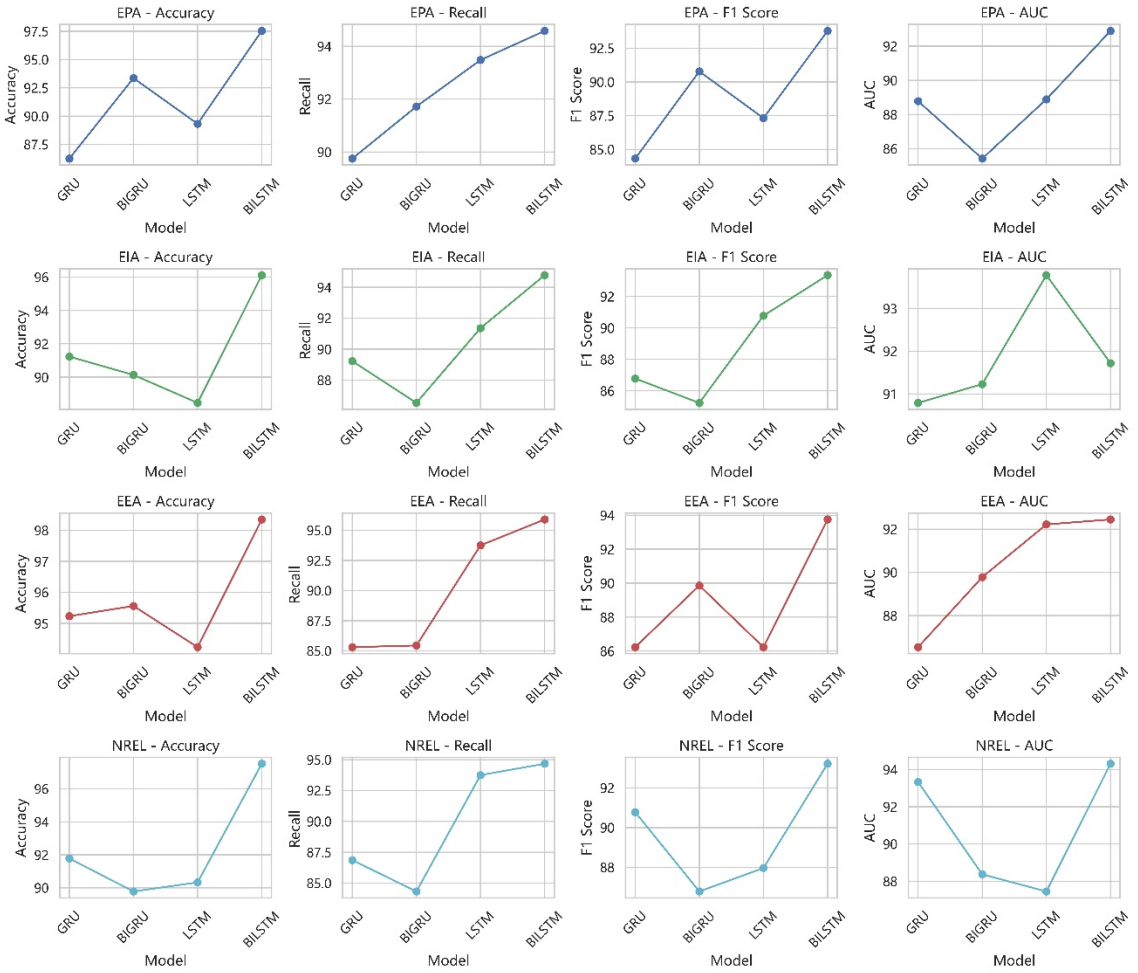

Figure 7. Comparison of Model Performance on Different Datasets.

Table 4 and Figure 8 presents the results of ablation experiments concerning the Dynamic Attention Mechanism (Dynamic-AM), employing datasets such as EPA, EIA, EEA, and NREL. T Of particular significance, within the EPA dataset, the Dynamic-AM attained an accuracy of 95.53.

Table 4. Ablation experiments on the Dynamic attention mechanism module using different datasets.

Model	Datasets															
	EPA				EIA				EEA				NREL			
	Accuracy	Recall	F1	AUC	Accuracy	Recall	F1	AUC	Accuracy	Recall	F1	AUC	Accuracy	Recall	F1	AUC
Cross-AM	76.46	92.85	81.42	77.58	92.33	85.27	84.33	93.96	97.55	82.58	86.22	86.64	92.88	86.96	93.89	91.59

Volume 6, Number 1, 2026
ISSN 2409-7284 (Online)

Multi-Head-AM	84.53	93.82	92.52	83.53	92.43	97.33	85.33	81.34	94.66	85.55	89.46	88.88	87.72	84.42	84.13	82.47
Self-AM	89.66	94.58	83.42	89.67	87.46	91.46	92.88	90.46	93.46	93.86	86.33	91.96	93.25	95.85	86.06	83.84
Dynamic-AM	95.63	94.68	93.88	95.99	96.22	98.88	93.46	91.82	98.45	96.01	93.85	92.86	97.62	94.77	96.60	96.43

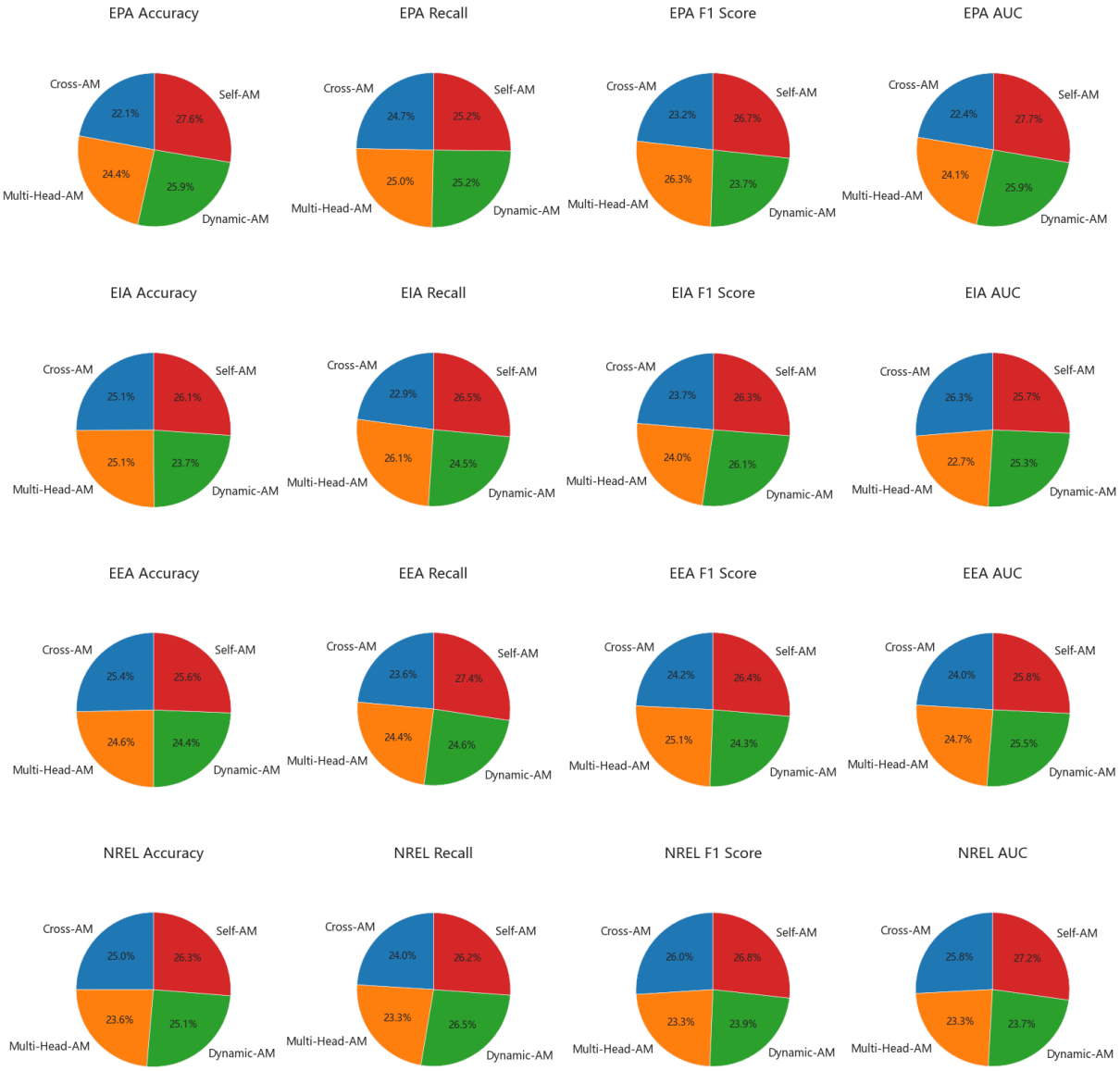

Figure 8. Comparison of Model Performance on Different Datasets.

5. Conclusions

This study introduces an innovative approach employing the TCN-BILSTM-Attention network to address the significant challenge of predicting carbon offsets. Extensive experiments were conducted using four distinct datasets, namely EPA, EIA, EEA, and NREL, to evaluate the effectiveness of the proposed model. The experimental results highlight the remarkable efficacy of our model in carbon offset prediction, thereby providing substantial support for both scholarly

Volume 6, Number 1, 2026
ISSN 2409-7284 (Online)

research and practical applications in the field of carbon neutrality.

Notwithstanding the considerable efficacy demonstrated by our model in predicting carbon offsets, we recognize certain limitations inherent to its current framework. Primarily, the model encounters difficulties with non-stationary and high-dimensional time series data, thereby requiring additional research to augment its effectiveness. Moreover, despite its robust performance across multiple datasets, the model's applicability in practical scenarios may be limited by variations in data quality and availability. Consequently, further inquiry is warranted to refine the implementation of our model within real-world contexts to more adequately fulfill pragmatic criteria.

In subsequent research, we intend to further enhance our model to bolster its performance and robustness. Our efforts will focus on advanced feature engineering and meticulous data cleaning methodologies pertinent to time series data, aimed at optimizing the handling of empirical datasets. Moreover, we aim to investigate the applicability of our model to pragmatic challenges within the domain of carbon neutrality, including the real-time surveillance of carbon emissions and the optimization of carbon offset mechanisms. This study offers significant insights and techniques for the advancement of carbon neutrality, with the potential to make a positive contribution to the objectives of sustainable development. We are committed to pursuing further inquiries in this sphere and to supporting endeavors that foster a more sustainable future.

References

[1] Moshari A, Aslani A, Zolfaghari Z, Malekli M, Zahedi R. Forecasting and gap analysis of renewable energy integration in zero energy-carbon buildings: A comprehensive bibliometric and machine learning approach. Environmental Science and Pollution Research (2023).

[2] Keshavarzzadeh M, Zahedi R, Eskandarpanah R, Qezelbigloo S, Gitifar S, Farahani ON, et al. Estimation of nox pollutants in a spark engine fueled by mixed methane and hydrogen using neural networks and genetic algorithm. Heliyon 9 (2023).

[3] Zahedi R, Aslani A, Seraji MAN, Zolfaghari Z. Advanced bibliometric analysis on the coupling of energetic dark greenhouse with natural gas combined cycle power plant for co2 capture. Korean Journal of Chemical Engineering 39 (2022).

[4] Zahedi R, Ayazi M, Aslani A. Comparison of amine adsorbents and strong hydroxides soluble for direct air co2 capture by life cycle assessment method. Environmental Technology & Innovation 28 (2022).

[5] Wang J, Sun X, Cheng Q, Cui Q. An innovative random forest-based nonlinear ensemble paradigm of improved feature extraction and deep learning for carbon price forecasting. Science of the Total Environment 762 (2021).

[6] Zhao X, Ma X, Chen B, Shang Y, Song M. Challenges toward carbon neutrality in china: Strategies and countermeasures. Resources, Conservation and Recycling 176 (2022).

[7] Ziyuan C, Yibo Y, Simayi Z, Shengtian Y, Abulimiti M, Yuqing W. Carbon emissions index decomposition and carbon emissions prediction in Xinjiang from the perspective of population-related factors, based on the combination of stirpat model and neural network. Environmental Science and Pollution Research (2022).

[8] Amasyali K, El-Gohary NM. A review of data-driven building energy consumption prediction studies. Renewable and

ISSN 2409-7284 (Online)

Sustainable Energy Reviews 81 (2018).

[9] Waheed R, Sarwar S, Wei C. The survey of economic growth, energy consumption and carbon emission. Energy Reports 5 (2019).

[10] Anthony LFW, Kanding B, Selvan R. Carbontracker: Tracking and predicting the carbon footprint of training deep learning models. arXiv preprint arXiv:2007.03051 (2020).

[11] Chen Y, Yang J, Luo L, Zhang H, Qian J, Tai Y, et al. Adaptive noise dictionary construction via irrpca for face recognition. Pattern Recognition 59 (2016).

[12] García-Martín E, Rodrigues CF, Riley G, Grahn H. Estimation of energy consumption in machine learning. Journal of Parallel and Distributed Computing 134 (2019).

[13] Pacheco KA, Reis AC, Bresciani AE, Nascimento CA, Alves RM. Assessment of the Brazilian market for products by carbon dioxide conversion. Frontiers in Energy Research 7 (2019).

[14] Zhang H, Qian F, Shang F, Du W, Qian J, Yang J. Global convergence guarantees of (a) gist for a family of nonconvex sparse learning problems. IEEE Transactions on Cybernetics 52 (2020).

[15] Elmaz F, Eyckerman R, Casteels W, Latre S, Hellinckx P. CNN-LSTM architecture for predictive indoor temperature modeling. Building and Environment 206 (2021).

[16] Liu T, Xu C, Guo Y, Chen H. A novel deep reinforcement learning-based methodology for short-term HVAC system energy consumption prediction. International Journal of Refrigeration 107 (2019).

[17] Zhang H, Qian F, Zhang B, Du W, Qian J, Yang J. Incorporating linear regression problems into an adaptive framework with feasible optimizations. IEEE Transactions on Multimedia (2022).

[18] Tang J, Li J. Carbon risk and return prediction: Evidence from the multi-CNN method. Frontiers in Environmental Science 10 (2022).

[19] Wenya L. Cooling, heating and electric load forecasting for integrated energy systems based on CNN-LSTM. 2021 6th International Conference on Power and Renewable Energy (ICPRE) (IEEE) (2021).

[20] Shen Z, Wu Q, Qian J, Gu C, Sun F, Tan J. 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) (IEEE) (2022).

[21] Liu B, Wang S, Liang X, Han Z. Carbon emission reduction prediction of new energy vehicles in China based on GRA-BILSTM model. Atmospheric Pollution Research 14 (2023).

[22] Yu X. The influence of regional tourism economy development on carbon neutrality for environmental protection using improved recurrent neural network. Frontiers in Ecology and Evolution 11 (2023).

[23] Sheng Y, Wang H, Yan J, Liu Y, Han S. Short-term wind power prediction method based on deep clustering-improved temporal convolutional network. Energy Reports 9 (2023).

[24] Oyando HC, Kanyolo TN, Chang CK. RNN-based main transformer OLTC control for SMR integration into a high renewable energy penetrated grid. Journal of Electrical Engineering & Technology (2023).

[25] Liu P, Zhang L, Gulla JA. Dynamic attention-based explainable recommendation with textual and visual fusion. Information Processing & Management 57 (2020).

[26] Liu Y, Zhang F, Yang S, Cao J. Self-attention mechanism for dynamic multi-step ROP prediction under continuous learning structure. Geoenergy Science and Engineering (2023).

[27] Kow PY, Chang LC, Lin CY, Chou CCK, Chang FJ. Deep neural networks for spatiotemporal PM2.5 forecasts based on atmospheric chemical transport model output and monitoring data. Environmental Pollution 306 (2022).

[28] Lv Z, Piccialli F. The security of medical data on the internet based on differential privacy technology. ACM Transactions on Internet Technology 21 (2021).

[29] Tietge U, Mock P, Dornoff J. CO2 emissions from new passenger cars in the European Union: Car manufacturers' performance in 2018 (2019).

[30] Papi F, Bianchini A. Technical challenges in floating offshore wind turbine upscaling: A critical analysis based on the NREL 5 MW and IEA 15 MW reference turbines. Renewable and Sustainable Energy Reviews 162 (2022).

[31] Gao M, Yang H, Xiao Q, Goh M. A novel fractional grey Riccati model for carbon emission prediction. Journal of Cleaner Production 282 (2021).

[32] Zhou W, Zeng B, Wang J, Luo X, Liu X. Forecasting Chinese carbon emissions using a novel grey rolling prediction model. Chaos, Solitons & Fractals 147 (2021).

[33] Huang Y, He Z. Carbon price forecasting with optimization prediction method based on unstructured combination. Science of the Total Environment 725 (2020).

[34] Cai K, Wu L. Using grey Gompertz model to explore the carbon emission and its peak in 16 provinces of China. Energy and Buildings 277 (2022).

[35] Huo T, Xu L, Feng W, Cai W, Liu B. Dynamic scenario simulations of carbon emission peak in China's city-scale urban residential building sector through 2050. Energy Policy 159 (2021).