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

Tehreem Murtaza[1*], Iqra Nawab[2]

[1*]Department of International Relations University of Okara, Pakistan; murtazatehreem25@gmail.com

[2]Department of Software Engineering Fast University, Pakistan; nawabiqra68@gmail.com

*Corresponding Author: murtazatehreem25@gmail.com

DOI: https://doi.org/10.30212/JITI.202604.001

Submitted: Sep. 22, 2025 Accepted: Dec. 14, 2025

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 methodologies 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 understanding 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 understanding and optimization of carbon neutrality strategies. Through extensive experimental validation, our research demonstrates the exceptional performance of the TCN-BILSTM-Attention model in the domain of carbon neutrality. Specifically, the proposed model outperforms existing approaches across four datasets (EPA, EIA, EEA, NREL). For instance, it achieved an accuracy of 97.53% on the EPA dataset and 96.12% on the EIA dataset. Overall, this study has significant implications not only for the practical application of carbon neutrality principles but also for providing novel perspectives and methodologies in global environmental sustainability and climate change mitigation. By offering innovative models and analytical tools for sustainable development, this work contributes valuable resources toward achieving carbon neutrality and advancing environmental conservation efforts.

Keywords: Sustainable development, Carbon neutrality, Time-Series data, TCN-BILSTM-Attention model, Decision support

1. Introduction

In today's world, carbon neutrality has emerged as one of the most critical challenges facing global society. With the escalating impacts 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 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 important position in carbon neutrality research. Time series data consist of sequences of data points ordered chronologically, including meteorological data, carbon emission records, and energy consumption data, among others. These data are essential for decision-making and planning in carbon neutrality initiatives, as they reflect temporal changes in environmental and industrial activities. Leveraging deep learning techniques for time series prediction enables decision-makers to better anticipate future trends and challenges, thereby facilitating the formulation of more effective carbon neutrality strategies. Consequently, time series forecasting holds substantial potential in carbon neutrality research, and this paper explores how deep learning methods can enhance forecasting accuracy to support carbon neutrality goals [3,4].

To date, time series analysis has made significant contributions in fields such as energy forecasting and sustainable development [5]. For example, its application in ping pong training camps has proven valuable. By applying time series analysis to track and analyze operational and performance data, critical temporal trends and patterns can be identified, offering essential insights for planning and decision-making within training camps [6,7]. Furthermore, time series analysis helps managers understand seasonal variations, such as fluctuations in student enrollment during summer and winter breaks, thereby enabling more efficient resource and personnel allocation. Recent developments in time series analysis within carbon neutrality research include the following approaches:

Some researchers have employed Convolutional Neural Networks (CNNs) for carbon emission prediction tasks [8]. CNNs excel at extracting features from spatiotemporal data, and effectively capturing patterns across spatial and temporal dimensions [8]. However, these models typically require large datasets for training, making them less effective in data-scarce scenarios, and they often lack flexibility in integrating heterogeneous data sources [9].

On the other hand, Long Short-Term Memory (LSTM) networks are widely used for energy consumption optimization [10]. LSTM models are well suited for time series data due to their memory mechanisms, which enable forecasting of future energy demand. 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 [14]. By simulating carbon markets and policy environments, DRL models can learn optimal emission strategies. However, training, and tuning DRL models require substantial computational resources, and concerns regarding robustness and interpretability remain in real-world policy applications [15].

Time Series Generative Adversarial Networks (TS-GANs) have been applied to generate realistic time series data, including carbon emission records. These models improve data completeness by generating synthetic samples to address missing data issues. However, TS-GANs face challenges in long-term forecasting and stability, particularly when external influencing factors are present [16,17,18].

Bidirectional Long Short-Term Memory (BILSTM), and an attention mechanism, specifically devised to forecast carbon neutrality. The TCN effectively captures long-term dependencies and spatiotemporal patterns, improving temporal dynamics and prediction accuracy of carbon emission. The BILSTM efficiently processes sequential data, capturing both long-term and short-term dependencies, while integrating information over varying temporal scales. The attention mechanism selectively emphasizes crucial time steps, enhancing both predictive accuracy and model interpretability. By leveraging the strengths of each component, the proposed approach generates accurate forecasts that support carbon neutrality strategies, sustainable development, and climate change mitigation.

2. Method

The TCN-BILSTM-Attention network integrates a Temporal Convolutional Network (TCN), Bidirectional Long Short-Term Memory (BILSTM), and an attention mechanism to improve carbon emission prediction accuracy and interpretability. Each component plays a distinct yet complementary role in addressing the complexities of carbon neutrality forecasting. The BILSTM processes sequential data by capturing both long- and short-term dependencies, while integrating information across multiple temporal scales to enhance prediction performance. The attention mechanism highlights critical time steps, improving accuracy and providing interpretability, which enables decision-makers to understand the basis of model predictions and formulate effective carbon offset strategies.

The architecture of the TCN-BILSTM-Attention network is structured as follows: First, the model receives historical time series data related to carbon emissions and associated influencing factors. The TCN module is then applied to capture long-term dependencies within the input data and extract temporal features. It utilizes dilated convolutions to expand the perceptual field and enable the identification of spatiotemporal patterns over extended time horizons. Next, the processed data is passed through the BILSTM layer, which refines the feature representation by capturing both forward and backward dependencies. Subsequently, the attention mechanism is applied to the BILSTM outputs. It dynamically assigns attention weights to different time steps, accentuating the most influential time points for prediction. Finally, the weighted BILSTM outputs are used to generate accurate predictions of future carbon emissions trends. The overall network architecture and data flow are illustrated in Figure 1.

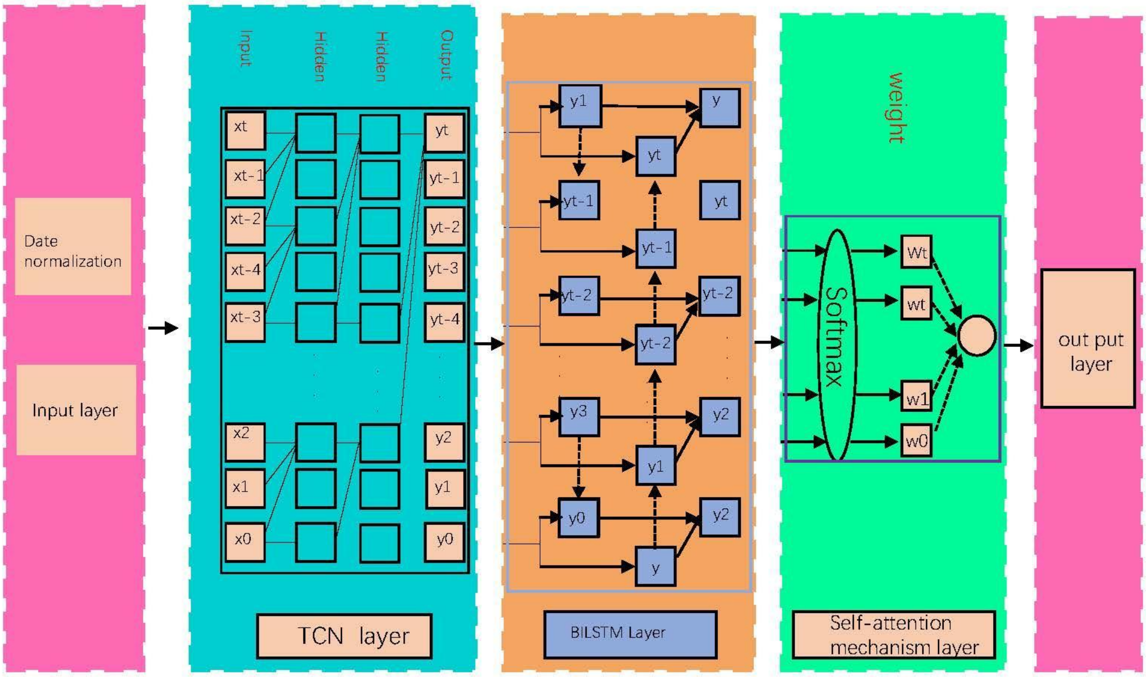

Figure 1. TCN-BILSTM-Attention Network Architecture Diagram.

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].

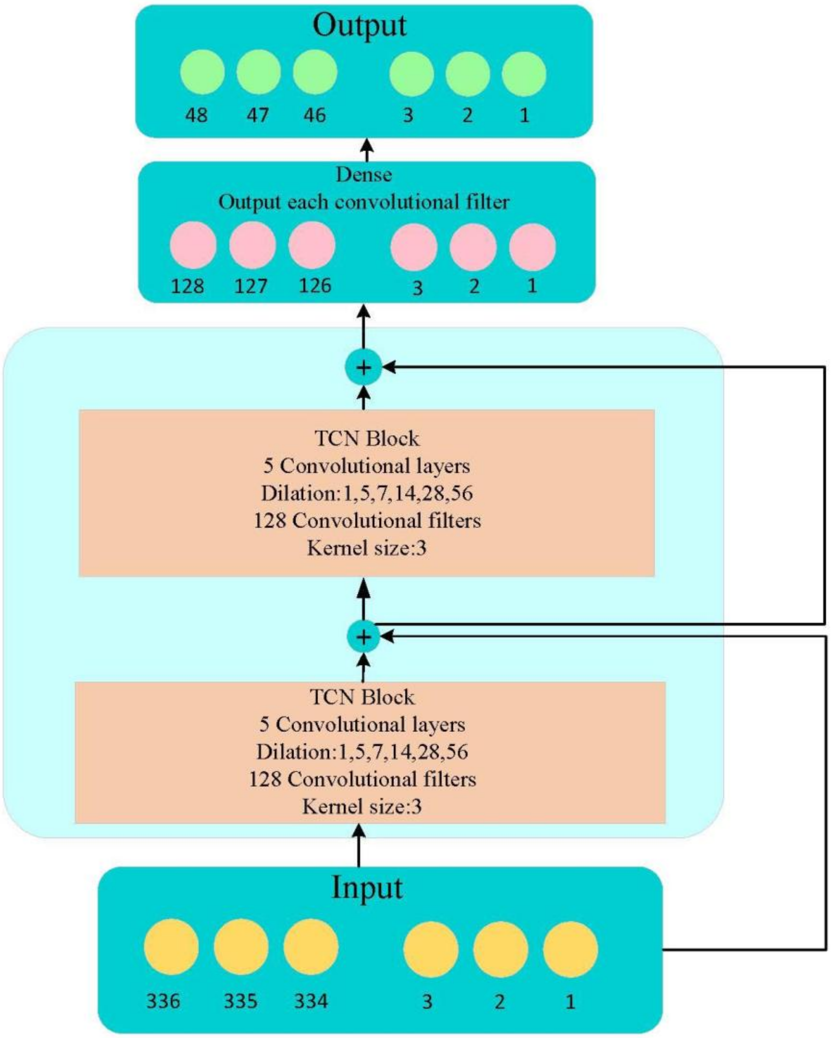

Figure 2. Flow chart of the TCN model.

Convolution Operation:

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

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

Dilated Convolution:

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

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:

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

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.

Temporal Convolutional Layer:

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

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:

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

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, a 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].

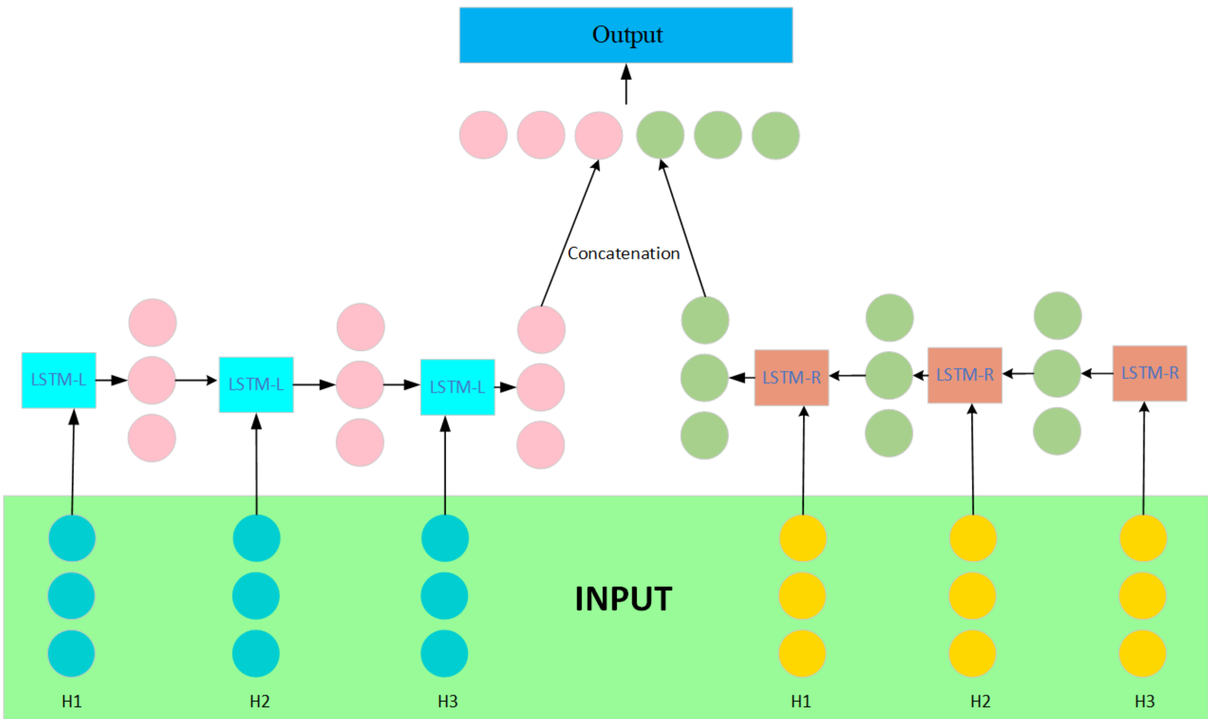

Figure 3. Flow chart of the BILSTM model.

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]. This is crucial for tasks like carbon offset prediction because it enables us to gain a better understanding of the temporal dependencies between carbon emissions and related factors. 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:

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

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

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:

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

where: $\mathcal{L}$ represents the loss function, T is the sequence length, and N indicates the number of classes. The true label for class i at time step t is denoted by $y_{t,i}$, while $\hat{y}_{t,i}$ denotes the predicted probability for class i at time step t.

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:

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

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

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

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

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

2.3 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.

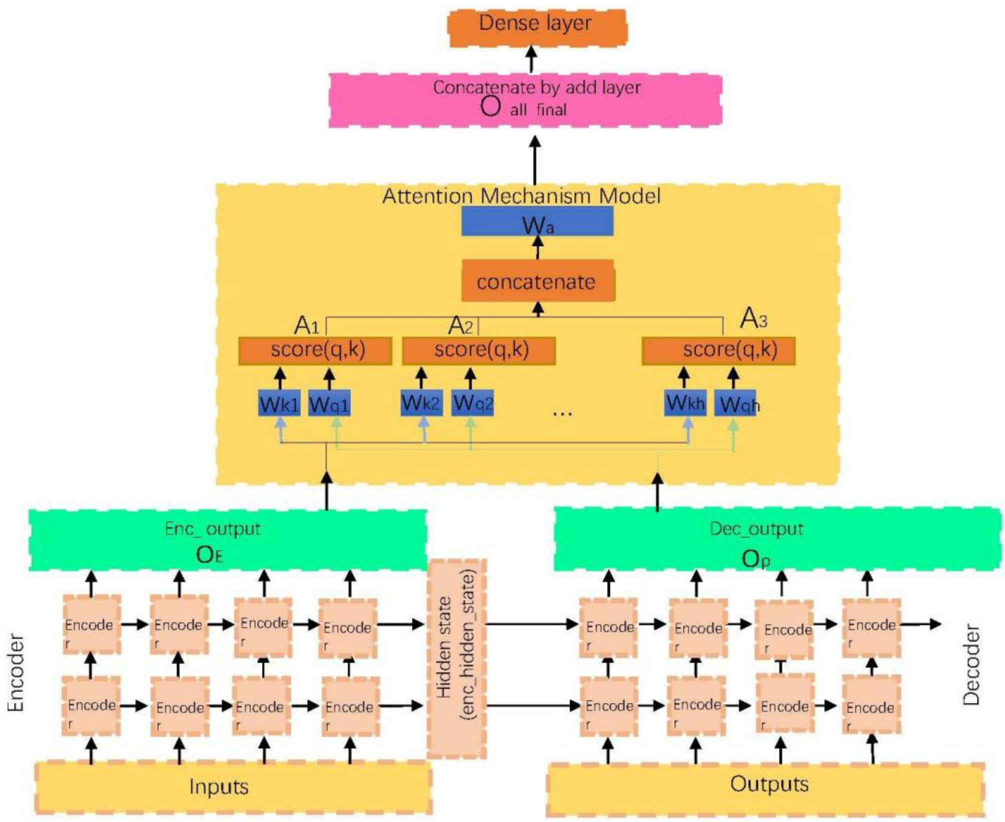

Figure 4. Flow chart of the attention model.

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.

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

where: Weighted_t is the weighted sum of values at time step t. N is the total number of time steps or elements. Value_i represents the value at time step i.

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

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

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

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.

3. Experiments

3.1 Dataset

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

EPA Dataset [27]: The Environmental Protection Agency (EPA) dataset contains extensive 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]: The Energy Information Administration dataset provides crucial data on energy production, consumption, and trends. Integrating this dataset improves our understanding of energy usage patterns and their relationship with carbon emissions, thereby refining predictive accuracy.

EEA Dataset [29]: The European Environment Agency (EEA) dataset offers insights into carbon emissions and environmental conditions specific to European regions. Including this dataset enhances the geographical diversity of our study and ensures the adaptability of the proposed model across different regulatory environments.

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

By employing multiple datasets, we conduct a comprehensive evaluation of the model's effectiveness across diverse regions, energy sources, and environmental conditions. This multi-faceted approach ensures that the proposed Dynamic Attention Mechanism can provide valuable insights and accurate predictions to support the development of effective carbon neutrality strategies on a global scale.

3.2 Experimental Details

Step1: Data preprocessing

- In the initial data preprocessing stage, raw datasets are cleaned to ensure data integrity and quality by addressing missing values, handling outliers, and resolving inconsistencies. This process establishes a solid foundation for subsequent analysis.
- To streamline the dataset and focus on the most relevant factors, feature selection is performed.

This step involves identifying and retaining key variables that have a significant impact on the research objectives, thereby reducing dimensionality and improving computational efficiency.

- 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.

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 selected to achieve an optimal balance between computational efficiency and gradient estimation accuracy. (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: A Temporal Convolutional Network (TCN) module is employed to capture long-term dependencies and extract temporal features. The module consists of four 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 learn spatiotemporal patterns over extended time intervals.
- BILSTM Module: Bidirectional Long Short-Term Memory (BILSTM) networks are utilized to refine feature representations. The BILSTM module comprises two LSTM layers, each containing 128 hidden units. This design effectively captures both forward and backward dependencies, providing a solid foundation for modeling temporal data.
- Attention Mechanism: An attention mechanism on top of the BILSTM outputs to dynamically assign weights to different time steps, highlighting the most influential temporal features. A multi-head self-attention mechanism with 8 attention heads is employed 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) Mean squared error (MSE) was used 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 to

preserve the best-performing models for future deployment.

3.3 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 dataset, EIA dataset, EEA dataset, and NREL dataset.

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

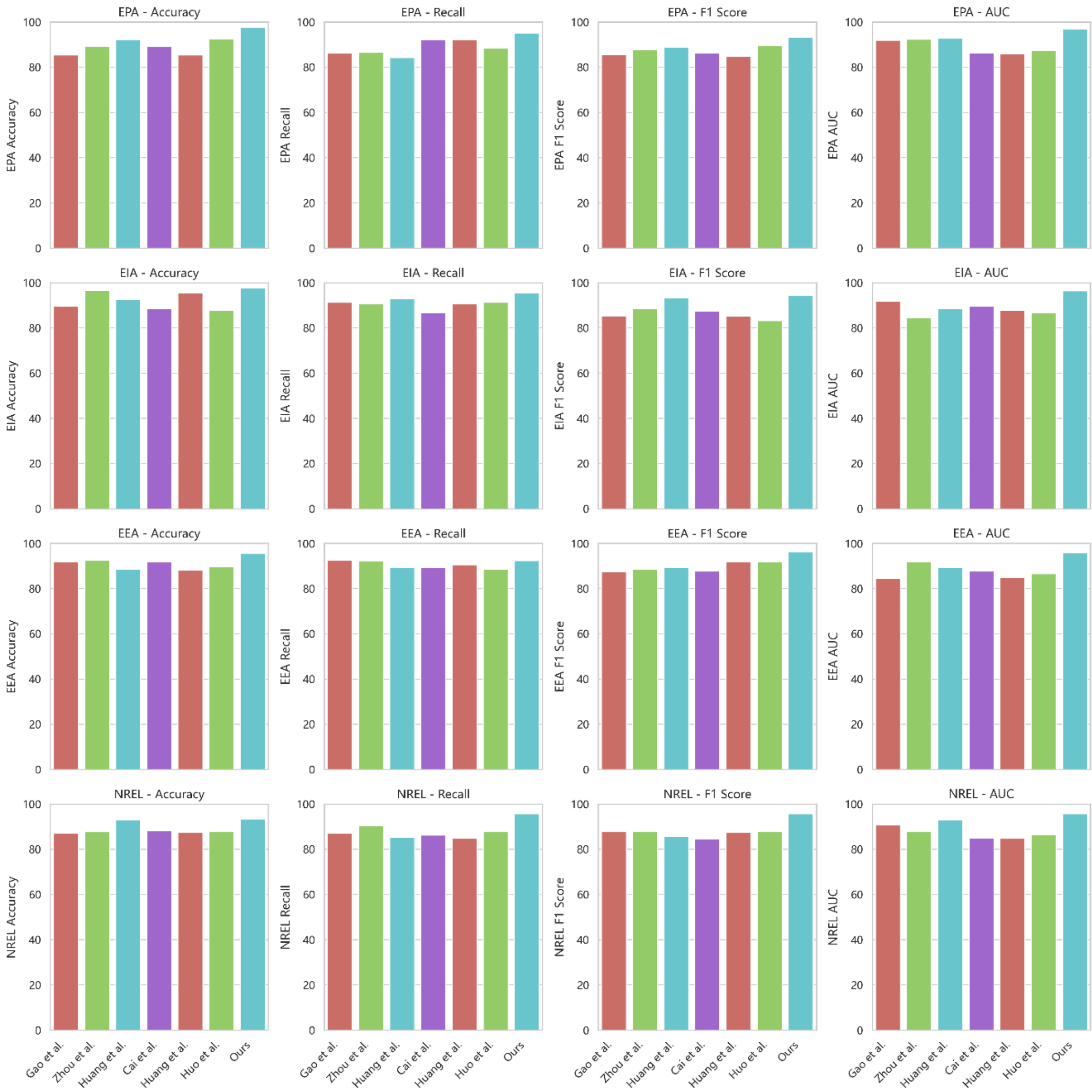

Figure 5. Comparison of model performance on different datasets.

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.

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

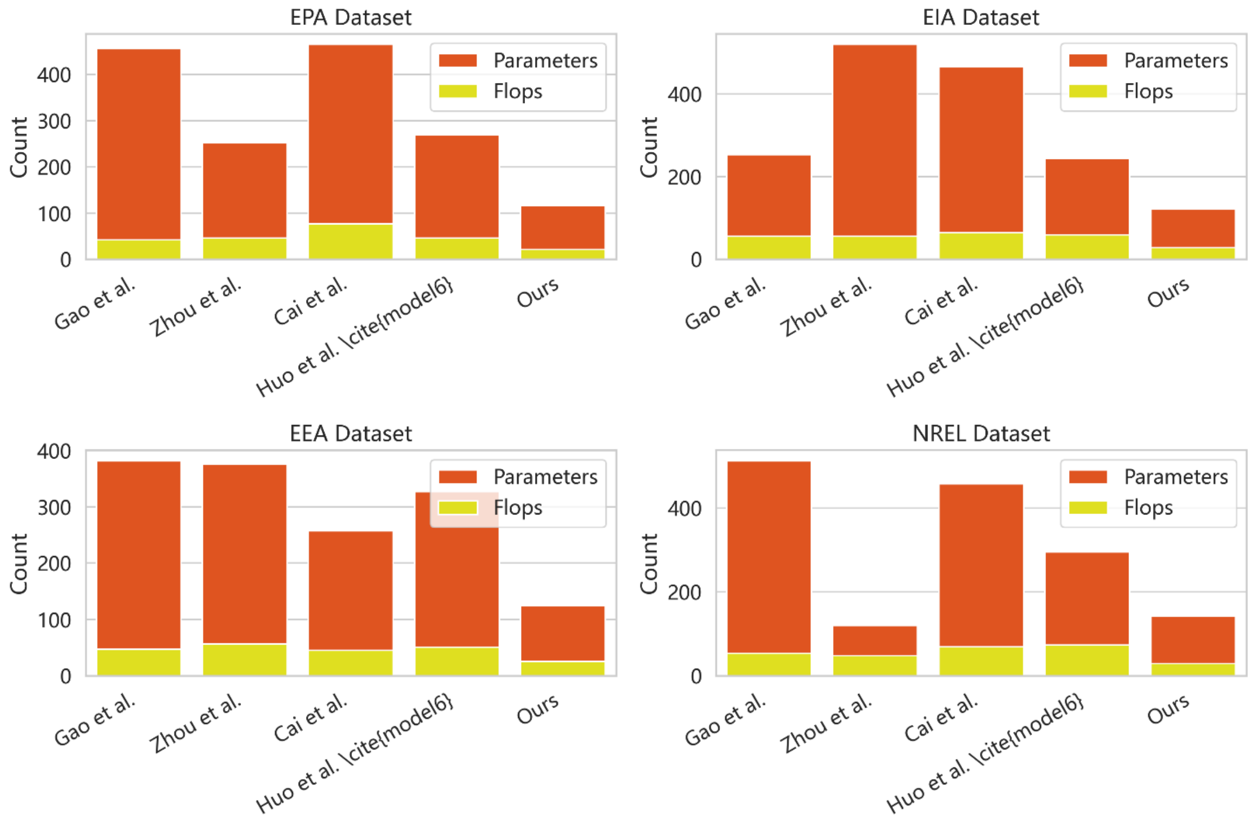

Figure 6. Comparison of different indicators of different models.

3.4 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
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

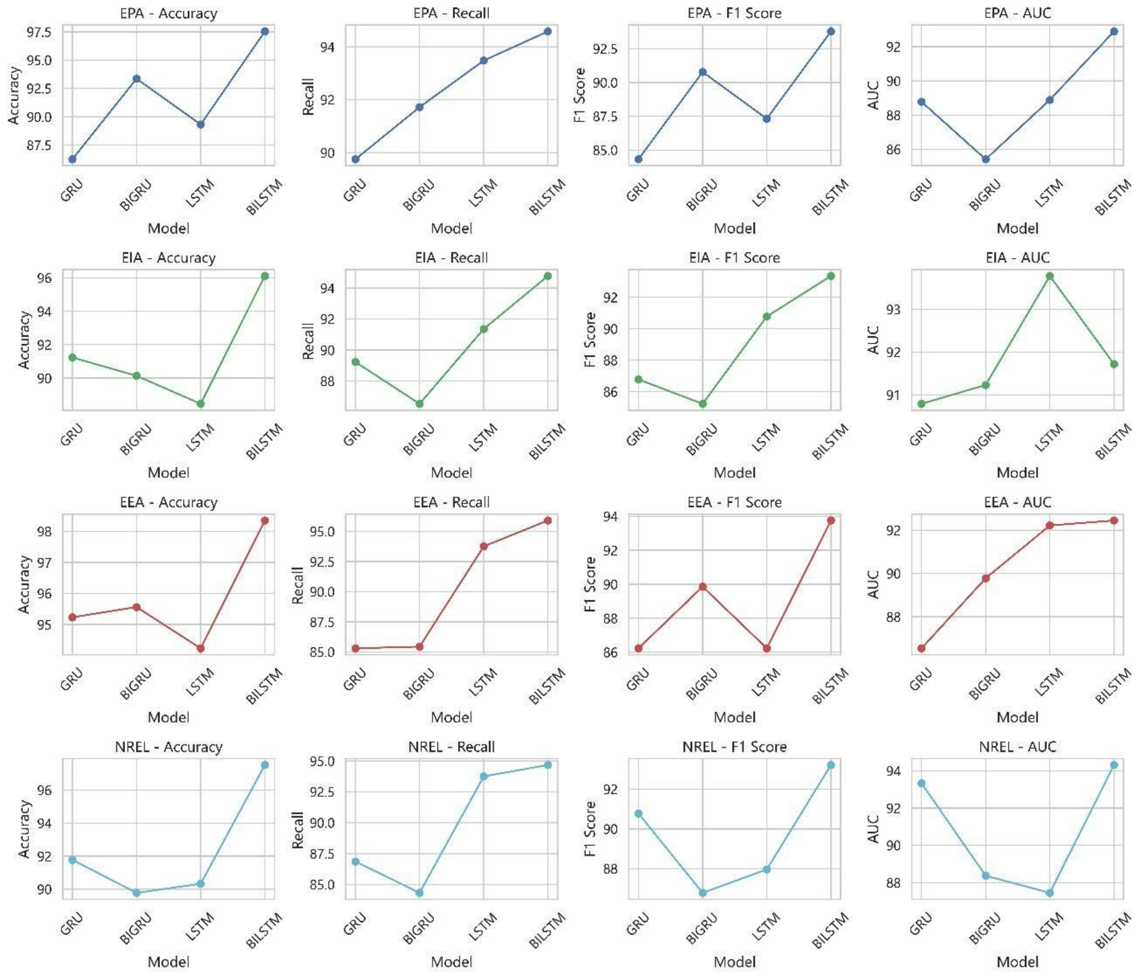

Figure 7. Comparison of model performance on different datasets.

4. 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 outcomes of the comparative analysis indicate that our model markedly outperforms existing baseline models, including those developed by Gao et al., Zhou et al., and Huang et al., across several performance metrics such as accuracy, recall, F1 score, and AUC. Additionally, ablation studies were performed to confirm the critical role of the dynamic attention mechanism module. The experimental results highlight the remarkable efficacy of our model in carbon offset prediction, thereby providing substantial support for both scholarly research and practical applications in the field of carbon neutrality.

Notwithstanding the considerable efficacy demonstrated by our model in predicting carbon offsets, we acknowledge certain limitations inherent to its current framework. Primarily, the model encounters challenges when handling 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 real-world scenarios may be constrained by variations in data quality and availability. Consequently, further investigation is warranted to refine the model’s implementation in practical settings to more adequately fulfill pragmatic criteria.

In subsequent research, we intend to further enhance the 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 application of the proposed model to practical challenges in 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.

Acknowledgements

This article received no financial or funding support.

Conflicts of Interest

The authors confirm that there are no conflicts of interest.

References

[1] Moshari, A., Aslani, A., Zolfaghari, Z., Malekli, M. and 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. and Farahani, O.N. Estimation of NOx pollutants in a spark engine fueled by mixed methane and hydrogen using neural networks and genetic algorithm. Heliyon, 2023, 9.

[3] Zahedi, R., Aslani, A., Seraji, M.A.N. and 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, 2022, 39.

[4] Zahedi, R., Ayazi, M. and Aslani, A. Comparison of amine adsorbents and strong hydroxides soluble for direct air CO2 capture by life cycle assessment method. Environmental Technology and Innovation, 2022, 28.

[5] Wang, J., Sun, X., Cheng, Q. and 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, 2021, 762.

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

[7] Chen, Z., Yu, Y., Zhang, S., Yang, S., Mamat, A. and Wang, Y. 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. and El-Gohary, N.M. A review of data-driven building energy consumption prediction studies. Renewable and Sustainable Energy Reviews, 2018, 81.

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

[10] Anthony, L.F.W., Kanding, B. and Selvan, R. Carbontracker: tracking and predicting the carbon footprint of training deep learning model. sarXiv preprint arXiv:2007.03051, 2020.

[11] Chen, Y., Yang, J., Luo, L., Zhang, H., Qian, J., Tai, Y. and Ding, X. Adaptive noise dictionary construction via IRRPCA for face recognition. Pattern Recognition, 2016, 59.

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

[13] Pacheco, K.A., Reis, A.C., Bresciani, A.E., Nascimento, C.A. and Alves, R.M. Assessment of the Brazilian market for products by carbon dioxide conversion. Frontiers in Energy Research, 2019, 7.

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

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

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

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

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

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

[20] Shen, Z., Wu, Q., Qian, J., Gu, C., Sun, F. and Tan, J. Federated learning for long-term forecasting of electricity consumption towards a carbon-neutral future. In: 2022 7th International Conference on Intelligent Computing and Signal Processing (ICSP), IEEE, 2022.

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

[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, 2023, 11.

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

[24] Oyando, H.C., Kanyolo, T.N. and Chang, C.K.RNN-based main transformer OLTC control for SMR integration into a high renewable energy penetrated grid. Journal of Electrical Engineering and Technology, 2023.

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

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

[27] Kow, P.Y., Chang, L.C., Lin, C.Y., Chou, C.C.K. and Chang, F.J. Deep neural networks for spatiotemporal PM2.5 forecasts based on atmospheric chemical transport model output and monitoring data. Environmental Pollution, 2022, 306.

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

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

[30] Papi, F. and 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, 2022, 162.

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

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

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

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

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