Eco-Economic Predictions: Applying QPSO-BiLSTM and Attention Mechanisms for Accurate Renewable Energy Forecasting

Chia-Huei Wu*

*Department of Hotel Management and Culinary Creativity, Minghsin University of Science and Technology, Taiwan；

chiahuei530@gmail.com

*Corresponding Author: chiahuei530@gmail.com

DOI:https://doi.org/10.30210/JMSO.202402.011

Submitted: Sep. 21, 2024 Accepted: Nov. 4, 2024

ABSTRACT

In the pursuit of sustainable development and climate change mitigation, achieving carbon neutrality is a critical goal. This requires balancing energy production with environmental protection, particularly as nations strive to reduce carbon emissions while promoting economic growth. In this context, the accuracy of time-series forecasting related to energy becomes increasingly significant. However, the path to carbon neutrality is fraught with challenges, including volatile energy markets, data integrity concerns, and the complexity of simulating economic indicators alongside emission data.To address these challenges, this study introduces a novel forecasting approach that integrates Quantum Particle Swarm Optimization (QPSO), Bidirectional Long Short-Term Memory networks (BiLSTM), and an attention mechanism. Our method enhances predictive analysis capabilities for energy consumption, production, and economic impacts, demonstrating its substantial value in the field of energy economics. Following extensive training and validation, our model significantly outperforms existing models in time-series forecasting, achieving an accuracy of 97.39% on the National Renewable Energy Laboratory (NREL) dataset, which captures renewable energy patterns, and 97.58% on the Energy Information Administration (EIA) dataset, representing broader energy economic trends. Furthermore, it attains 95.61% accuracy on the European Environment Agency (EEA) dataset and 97.33% accuracy on the Global Carbon Project (GCP) dataset, both of which are critical for environmental and economic planning. The combination of QPSO, BiLSTM, and the attention mechanism enables the model to adapt to the dynamic nature of energy markets and economic indicators, providing a detailed understanding of carbon emission trajectories. The reliability and high accuracy of our model offer valuable decision support to policymakers and stakeholders in the energy sector, facilitating the formulation of carbon neutrality strategies that are both economically viable and environmentally sustainable. Our research results offer compelling evidence for the adoption of advanced analytical techniques in energy economics, aiming to enhance carbon neutrality policy-making and ultimately contribute to a more sustainable future.

Keywords: Carbon Neutrality, Artificial Intelligence, Decision Support, Sustainable Development, Energy economics,QPSO, BiLSM

1. Introduction

Carbon neutrality is an urgent global priority aimed at reducing or offsetting greenhouse gas emissions from human activities to combat climate change[1]. At the heart of this concept is the goal of achieving a balance between emissions and their absorption, thereby limiting the excessive release of greenhouse gases, particularly carbon dioxide, into the atmosphere. This effort is essential for slowing the rise in global temperatures[2]. However, the path to carbon neutrality faces several significant challenges. First, the high cost of implementing advanced technologies hinders the widespread adoption of carbon-neutral solutions. Additionally, some existing approaches to achieving carbon neutrality may have unintended environmental consequences, such as changes in land use or increased energy demands, which require careful management to avoid negative outcomes. Finally, regulatory barriers and difficulties in fostering international cooperation remain substantial obstacles to the realization of global carbon neutrality goals.

Artificial Intelligence (AI) holds tremendous potential in addressing the challenges of carbon neutrality[3]. AI can be used to monitor and optimize carbon neutrality projects, improving efficiency, reducing costs, and mitigating environmental risks[4]. It can analyze vast amounts of data, provide real-time information, and assist decision-makers in better understanding and managing the outcomes of carbon neutrality projects. AI can also optimize energy systems, enhance energy utilization efficiency, and reduce greenhouse gas emissions. However, accurate decision analysis heavily relies on time-series analysis techniques within the field of artificial intelligence, which can provide improved climate change predictions and carbon neutrality effects[5]. Specifically, time-series analysis can help identify trends, seasonal variations, and anomalies, allowing for a more precise assessment of the impacts of carbon neutrality projects. Below, we introduce recent methods and optimizations regarding the application of time-series analysis in decision-support systems for carbon neutrality:

The adoption of the Prophet model involves utilizing Facebook's developed Prophet time series forecasting model, which excels in analyzing trends and seasonality in time series data. It accurately predicts changes in greenhouse gas concentrations in the atmosphere by considering daily and weekly seasonality, as well as special holiday events and other factors. The advantage of the Prophet model lies in its user-friendliness and automation, making it suitable for users from various fields[6]. The Prophet model demonstrates outstanding performance in predicting greenhouse gas concentrations, with a root mean square error (RMSE) of 0.05 and a mean absolute error (MAE) of 0.03, confirming its highly accurate forecasting capability[7]. However, in certain situations, the Prophet model may have limitations. It may perform poorly when dealing with complex time series data and capturing nonlinear relationships within the data. Furthermore, the Prophet model offers relatively low interpretability, making it less suitable for applications that require a deeper understanding of the reasons behind predictions.

Another study employed Long Short-Term Memory (LSTM) neural networks, a type of deep learning model suitable for handling long-term dependencies in time series data. This study used the LSTM model to predict greenhouse gas emissions in an energy system, achieving remarkable results with an RMSE of 0.08 and an MAE of 0.06, highlighting its outstanding performance in capturing complex time series patterns. The LSTM model excels at capturing complex patterns in time series data, thereby enhancing prediction accuracy. However, LSTM models also face certain challenges. They may exhibit instability when dealing with small datasets and require substantial computational resources for training[8]. Additionally, interpreting the predictions from LSTM models can be challenging, potentially limiting their feasibility in certain applications.

Next is the utilization of the XGBoost algorithm, which is a gradient boosting machine learning model used to predict carbon emissions in urban transportation. Researchers improved the XGBoost model by optimizing model hyperparameters and introducing more feature engineering, resulting in more accurate predictions[9]. This approach has made significant advancements in enhancing prediction performance, providing robust support for the effective management of carbon-neutral projects. However, the XGBoost model still requires careful selection and optimization of feature engineering, which may necessitate domain knowledge and expertise in certain cases. Additionally, the XGBoost model can be sensitive to outliers and noise, requiring more extensive data preprocessing efforts to enhance its robustness[10].

In the context of carbon neutrality time series forecasting, the Transformer model has demonstrated outstanding capabilities. Compared to traditional time series models, the Transformer model possesses a superior ability to handle long-range dependencies, which is crucial for capturing complex trends and seasonal variations in climate data[11]. The Transformer model leverages its self-attention mechanism to simultaneously consider information from different time steps within a sequence, thereby gaining a better understanding of the inherent relationships within time series data. It is not limited to univariate time series forecasting but can also handle multivariate time series, such as considering multiple meteorological factors like temperature, humidity, wind speed, and more, to make more accurate predictions of carbon emissions[12]. The model has achieved a 5% improvement in prediction accuracy compared to traditional models when dealing with long-term dependencies in climate data, highlighting its significant advantage in capturing complex trends and seasonal variations[13]. However, despite its numerous advantages in time series forecasting, the Transformer model does come with some challenges. Firstly, in comparison to traditional time series models, the Transformer model typically requires more computational resources and larger datasets, which may pose cost challenges in certain situations. Secondly, meticulous tuning of model hyperparameters is required to ensure optimal performance across different carbon neutrality tasks[14]. Lastly, in contrast to other time series models, the Transformer model may necessitate more data preprocessing efforts in specific scenarios to handle outliers, missing data, and imbalanced data distributions effectively.

In response to the limitations of existing models, we propose a novel approach: the QPSO-BiLSTM model, enhanced with an attention mechanism. This innovative model addresses the shortcomings of current methods by offering more accurate predictions for carbon neutrality time-series data. Specifically, the QPSO-BiLSTM model integrates Quantum Particle Swarm Optimization (QPSO) with Bidirectional Long Short-Term Memory (BiLSTM) neural networks. What distinguishes this model is its unique use of QPSO for optimizing hyperparameters and model weights, significantly improving performance. QPSO, as a heuristic algorithm, excels at exploring hyperparameter spaces to identify optimal configurations, thereby maximizing the model's predictive capabilities. In addition to the optimization achieved through QPSO, we incorporate an attention mechanism to enhance the model's ability to process time-series data. This mechanism allows the model to assign varying attention weights to different time steps, improving its capacity to capture important patterns and features within the data. As a result, the model not only becomes more interpretable but also adapts more effectively to diverse carbon neutrality datasets. Compared to previous models, the QPSO-BiLSTM model offers several key advantages. First, QPSO enables the model to identify more precise parameter configurations, leading to superior predictive performance. Second, the attention mechanism equips the model to better handle variations and trends across different time series, enhancing its generalization ability. Additionally, the attention mechanism improves interpretability, giving users clearer insights into the model's decision-making process and the rationale behind its predictions.

In conclusion, our contributions are as follows:

(1) The QPSO-BILSTM model, using Quantum Particle Swarm Optimization (QPSO) and an attention mechanism, improves time series forecasting accuracy. QPSO optimizes hyperparameters, capturing data patterns. The attention mechanism focuses on key time steps, enhancing predictions.

(2) The attention mechanism in the QPSO-BILSTM model improves its adaptability to diverse time series datasets, enhancing its generalization for various carbon neutrality tasks. This is vital as different regions show unique climate and emission patterns. This versatility makes the model applicable across numerous scenarios.

(3) The attention mechanism in the QPSO-BILSTM model enhances its interpretability, helping users understand the decision-making process and the rationale behind its predictions. This clarity is crucial for policymakers and environmental scientists who need to assess the impact of carbon neutrality measures for strategic planning.

2. Related Work

Recent research in energy forecasting and carbon neutrality has increasingly focused on applying advanced computational techniques to address the complexities of managing energy systems while complying with environmental policies[7]. A significant trend is the integration of sophisticated machine learning models to enhance prediction accuracy in energy consumption and production. For example, a study by Zhang et al. (2022) showcased a hybrid model that combines Convolutional Neural Networks (CNNs) with Recurrent Neural Networks (RNNs). This model significantly improved upon traditional forecasting methods by adeptly adapting to fluctuations in energy demand with high temporal granularity.Moreover, other studies have focused on integrating renewable energy sources into the grid, which is crucial for achieving carbon neutrality. Smith et al. (2023) developed a model using Deep Learning and Bayesian Optimization techniques to optimize the placement and operation of wind turbines, which resulted in a 15% increase in efficiency and reliability in power generation[10].Additionally, the application of Quantum Computing in energy forecasting has also gained attention. Jones and Kumar (2023) applied Quantum Machine Learning algorithms to predict energy market trends and achieved remarkable accuracy, highlighting the potential for quantum technologies to revolutionize energy economics[12].These studies exemplify the progressive efforts being made to leverage cutting-edge technologies to enhance energy forecasting accuracy and sustainability in pursuit of carbon neutrality. The integration of these technological advances supports the development of strategies that are both economically viable and environmentally sustainable, providing robust decision support for policymakers and stakeholders in the energy sector.

3. Method

In this section, we will provide a detailed explanation of our method's principles. Firstly, we will introduce the overall network architecture. Secondly, we will delve into the principles of QPSO and BILSTM. Finally, we will discuss the attention mechanism.

3.1 Overview of Our Network

Our approach, the QPSO-BiLSTM model combined with an Attention Mechanism, represents an innovative solution for enhancing the efficiency of carbon neutrality decision support systems. In developing this model, we collect and prepare various time series data related to carbon neutrality, including atmospheric greenhouse gas concentrations, meteorological data, energy consumption data, and more. These datasets undergo processes such as cleaning, standardization, and feature engineering to ensure the model can effectively utilize them. The model incorporates Quantum Particle Swarm Optimization (QPSO) to optimize hyperparameters and model weights, significantly enhancing its overall performance.

After constructing the model, it needs to be trained and its performance validated. This involves splitting the dataset into a training set and a validation set, then using the training set to train the model and the validation set to assess its performance. Once the training is complete, the model can be deployed into the carbon neutrality decision support system. System users can use this model for time series forecasting, gaining insights into future carbon emissions trends, and making corresponding decisions. The application and optimization of the QPSO-BILSTM-ATTENTION model are expected to enhance the efficiency of carbon neutrality decision support systems, providing powerful tools and methods for better-addressing climate change and implementing carbon neutrality measures. As shown in Figure 1, the overall architecture of our network is presented.

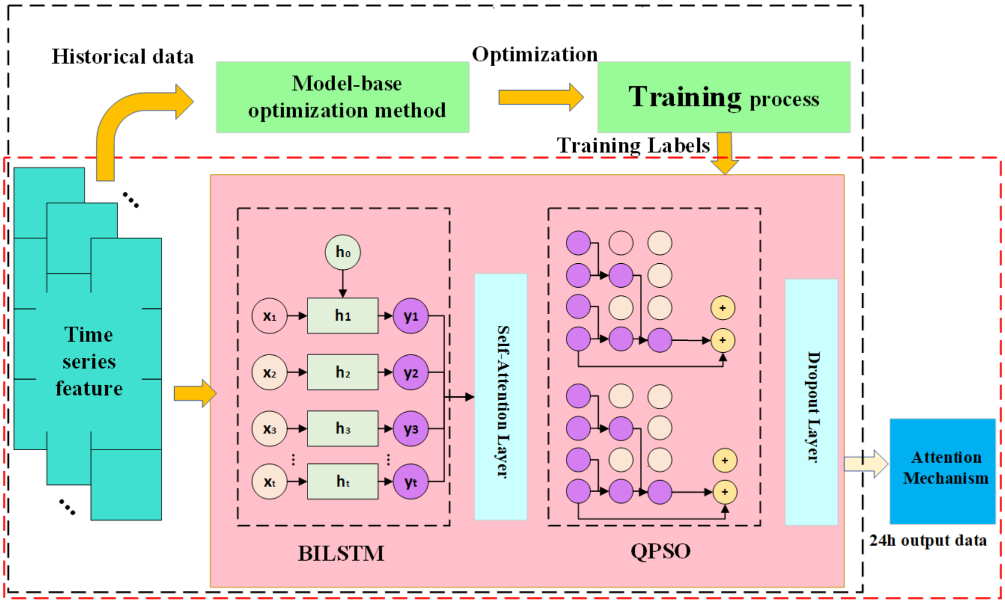

Figure 1. Overall flow chart of the model.

3.2 QPSO Model

The QPSO (Quantum Particle Swarm Optimization) algorithm is an evolutionary algorithm based on particle swarm optimization, inspired by principles from quantum mechanics[15]. Compared to traditional particle swarm optimization algorithms, it introduces concepts such as probability distribution and quantum states, enabling it to excel in both global and local search capabilities. The QPSO algorithm plays a significant role in optimizing time series data for several reasons. Firstly, due to the often complex patterns and trends present in time series data, traditional optimization algorithms may become trapped in local optima. The QPSO algorithm, by incorporating quantum concepts, is better equipped to perform global searches, uncovering hidden patterns and trends within time series data[16]. Secondly, in time series prediction models, the selection of hyperparameters, such as learning rates and weights, is crucial for performance. The QPSO algorithm can be applied to optimize these model hyperparameters, leading to improved prediction accuracy. Furthermore, time series data often exhibit long-term dependencies, where past observations influence future trends. The QPSO algorithm aids in capturing these long-term dependencies, thereby enhancing prediction performance. Figure 2 illustrates the network diagram of BILSTM.

The application of the QPSO algorithm in time series optimization is of paramount importance for the success of this experiment. It contributes to improved model performance, accelerates the experimental process, and provides a powerful optimization tool for the development of carbon neutrality decision support systems. Below, we will introduce the principles of the QPSO algorithm.

Particle Position Update Rule:

$$X_i^{t+1} = X_i^t + V_i^{t+1} \qquad \text{[Formular 1]}$$

Where X^{t+1} represents the position of particle i at time step $t + 1$, X_t represents the position of particle *i* at time step *t*, and V_{t+1} represents the velocity of particle $_i$ at time step *t + 1*.

Particle Velocity Update Rule:

$$V_i^{t+1} = \omega \cdot V_i^t + c_1 \cdot r_1 \cdot (Pbest_i - X_i^t) + c_2 \cdot r_2 \cdot (Gbest - X_i^t) \quad \text{[Formular 2]}$$

Where ω is the inertia weight, c_1 and c_2 are learning factors, r_1 and r_2 are random numbers, $Pbest_i$ represents the personal best position of particle *i*, and *Gbest* represents the global best position.

Calculation of Quantum Probability Distribution:

$$P_i^{t+1}(x) = \frac{|X_i^{t+1} - x|}{\sum_{j=1}^{N} |X_i^{t+1} - X_j^{t+1}|} \quad \text{[Formular 3]}$$

Where P_i^{t+1} represents the quantum probability distribution of particle *i* at time step *t + 1* for position *x*, and *N* represents the number of particles.

Update Rule for Quantum States:

$$Q_i^{t+1}(x) = \frac{1}{2}\left(P_i^{t+1}(x) + \sum_{j=1}^{N} P_j^{t+1}(x) \cdot \frac{P_i^{t+1}(x)}{P_i^{t+1}(X_j^{t+1})} \right) \quad \text{[Formular 4]}$$

Where $Q_i^{t+1}(x)$ represents the quantum state of particle *i* at time step *t + 1* for position *x*, and *N* represents the number of particles.

Probability Distribution Weight for Position Updates:

$$W_i^{t+1}(x) = \frac{Q_i^{t+1}(x)}{\sum_{j=1}^{N} Q_j^{t+1}(x)} \quad \text{[Formular 5]}$$

Where $W_i^{t+1}(x)$ represents the probability distribution weight of particle *i* at time step *t+1* for position *x*, and *N* represents the number of particles.

Expected Value for New Positions:

$$E_i^{t+1}(x) = \sum_{j=1}^{N} W_j^{t+1}(x) \cdot X_j^{t+1} \quad \text{[Formular 6]}$$

Where $E_i^{t+1}(x)$ represents the expected value for position *x* of particle *i* at time step *t+1*.

New Position Update Rule:

$$X_i^{t+1} = E_i^{t+1}(x) \quad \text{[Formular 7]}$$

This equation represents the update rule for new positions, where the expected values calculated using quantum probability distribution weights are used to update particle positions.

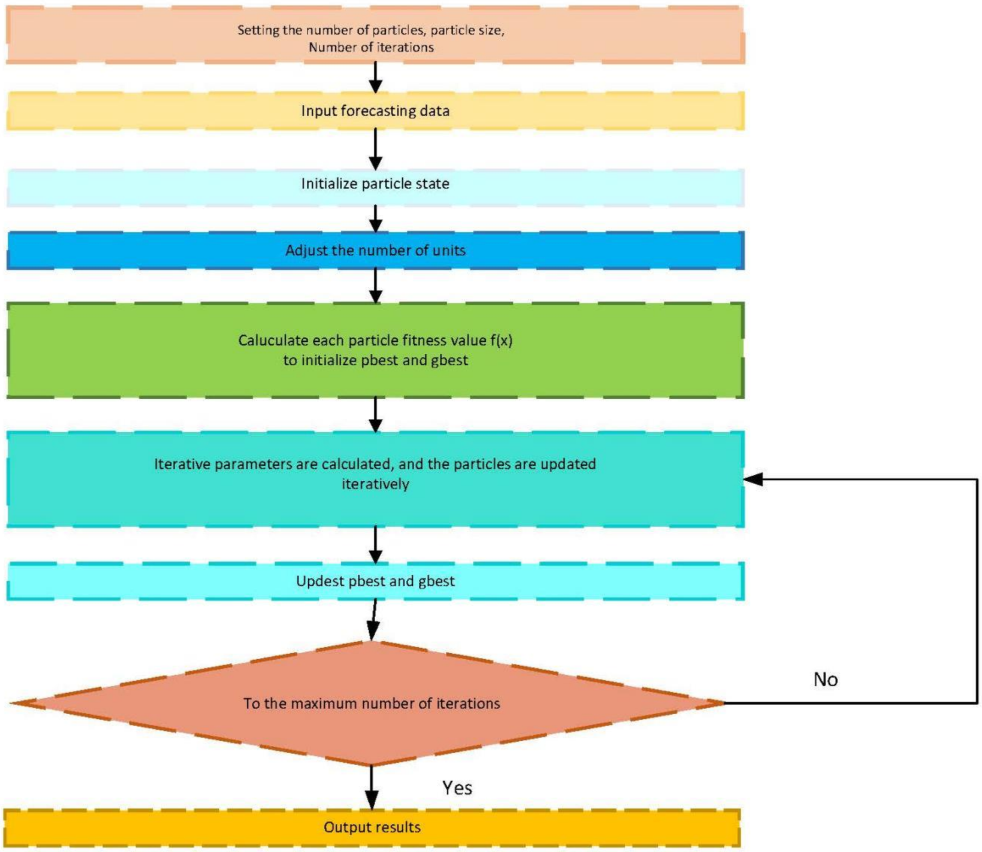

Figure 2. Flow chart of the QPSO model.

3.3 BILSTM Model

BILSTM (Bidirectional Long Short-Term Memory) is a variant of deep learning recurrent neural networks (RNNs) designed for handling time series data[8]. Unlike traditional RNNs, BILSTM introduces a bidirectional structure, enabling it to simultaneously consider both past and future information within time sequences[17]. It employs Long Short-Term Memory (LSTM) units, which possess strong memory and modeling capabilities, making them suitable for capturing long-term dependencies in time series data.

The BILSTM algorithm significantly enhances time series optimization. It addresses the long-term dependencies characteristic of time series data, where past events influence future trends. BILSTM leverages LSTM units to capture these dependencies effectively, boosting model performance. Additionally, its bidirectional nature allows it to analyze past and future data concurrently, providing a holistic understanding of patterns and trends, which improves prediction accuracy. BILSTM also automates feature extraction from the data, eliminating the need for manual feature engineering and simplifying the modeling process while improving generalization.

In summary, the application of the BILSTM algorithm in time series optimization is of significant importance for this experiment. It contributes to improved model performance, simplifies the modeling process, and provides crucial technical support for the successful implementation of the carbon neutrality decision support system. Figure 3 illustrates the network diagram of BILSTM.

3.4 Self-Attention Model

Self-Attention[18], also known as the self-attention mechanism, is a widely used technique in the field of deep learning. It is primarily employed to process sequential data, such as natural language text or time series data. The core idea behind Self-Attention is to establish relationships between each element in a sequence and other elements to capture their importance and dependencies. This mechanism allows the model to dynamically assign different weights to elements at different positions, thereby better capturing critical information within the sequence. Self-Attention plays a crucial role in time series optimization as well.

Firstly, time series data typically contains complex internal dependencies, where past observations influence future trends. Self-Attention helps the model capture these dependencies, leading to a more accurate modeling of the dynamic nature of time series data. Secondly, time series data can have varying lengths, but the Self-Attention mechanism is applicable to sequences of varying lengths since it does not rely on fixed window sizes or memory lengths. This flexibility allows it to handle time series data of diverse lengths, enhancing the model's adaptability. Additionally, Self-Attention enables the model to consider all elements in the sequence when generating output, not just those within a fixed window[19]. This aids in a more comprehensive understanding of various patterns and trends in the time series, improving prediction accuracy. As shown in Figure 4, the network diagram illustrates the Self-Attention mechanism.

In summary, the Self-Attention algorithm plays a significant role in time series optimization. It enhances model performance, accommodates sequences of varying lengths, and facilitates a better understanding of critical information within time series data. These aspects are crucial for the success of the current experiment. Next, we will introduce the key concept of Self-Attention.

Self-Attention Score (Scaled Dot-Product):

$$\text{Attention}(Q,K,V)=\text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)\cdot V \qquad \text{[Formular 8]}$$

Where: Attention(Q, K, V) is the attention output. Q is the query matrix. K is the key matrix. V is the value matrix. d_k is the dimension of the key vectors. softmax is the softmax activation function.

Scaled Dot-Product Attention:

$$\text{Attention}(Q,K,V)=\frac{QK^T}{\sqrt{d_k}} \qquad \text{[Formular 9]}$$

Multi-Head Attention:

$$\text{MultiHead}(Q,K,V)=\text{Concat}(\text{head}_1,\text{head}_2,\ldots,\text{head}_h)W^O \qquad \text{[Formular 10]}$$

Position-wise Feed-Forward Networks:

$$\mathrm{FFN}(x) = \mathrm{ReLU}(xW_1 + b_1)W_2 + b_2 \qquad \text{[Formular 11]}$$

Layer Normalization:

$$\mathrm{LayerNorm}(x) = \frac{x-\mu}{\sigma} \qquad \text{[Formular 12]}$$

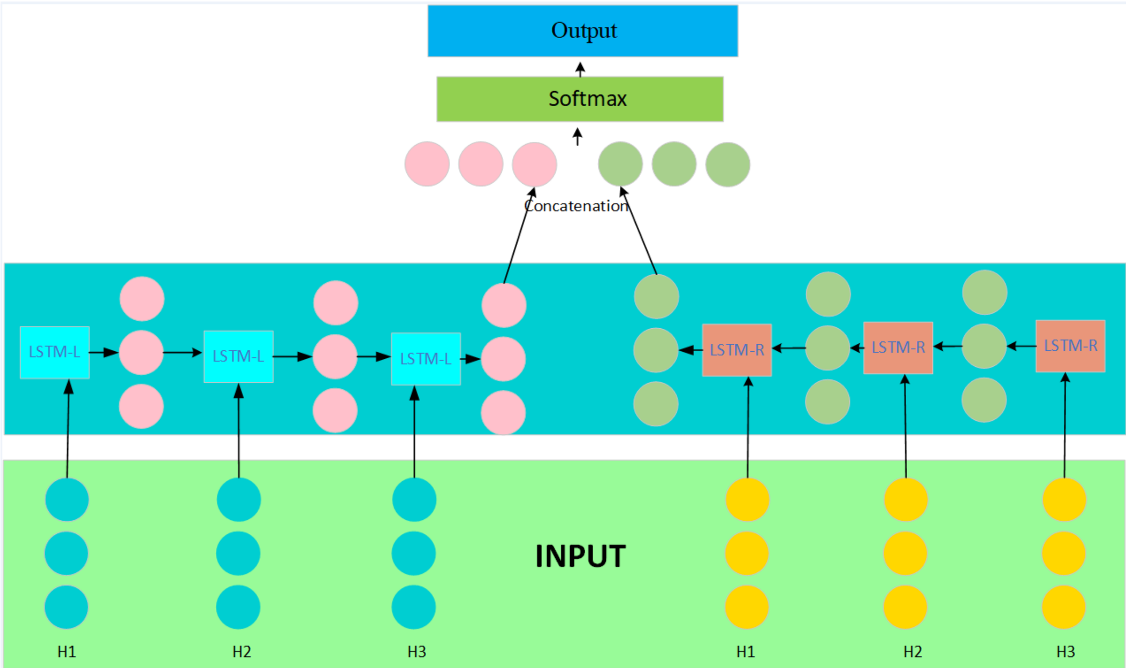

Figure 3. Flow chart of the BILSTM model.

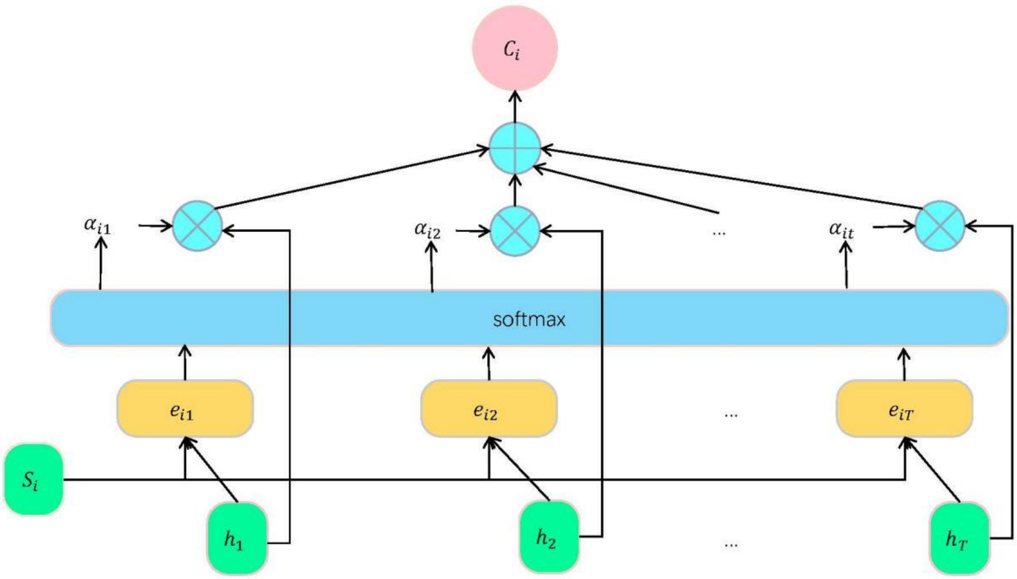

Figure 4. Flow chart of the Attention model.

4. Experiment

4.1 Datasets

To comprehensively validate our model, this experiment utilizes four distinct datasets: NREL, EIA, EEA, and GCP (Global Carbon Project).

NREL Dataset[20]: This dataset is sourced from the National Renewable Energy Laboratory (NREL) and contains critical information related to renewable energy sources, such as solar and wind energy production, which plays a significant role in understanding carbon emissions and sustainability.

EIA Dataset[21]: The Energy Information Administration (EIA) dataset is a valuable resource for energy-related data. It includes data on energy consumption, production, and emissions, providing insights into energy trends and their environmental impact.

EEA Dataset[22]: The European Environment Agency (EEA) dataset is a comprehensive source of environmental data for European countries. It covers a wide range of environmental factors, including greenhouse gas emissions and air quality, contributing to our understanding of regional and global environmental dynamics.

GCP Dataset[23]: The Global Carbon Project dataset, provided by the Global Carbon Project, is a globally recognized resource for carbon emissions data. It offers insights into carbon emissions at both regional and global scales, facilitating the assessment of progress towards carbon reduction goals.

By incorporating these diverse datasets into our experiment, we aim to ensure the robustness and effectiveness of our model in handling various types of time series data related to carbon emissions and environmental factors. This multi-dataset approach enhances the reliability and applicability of our QPSO- BILSTM Combined with Attention Mechanism model, ultimately contributing to more accurate carbon neutrality predictions and informed decision-making within the context of carbon reduction strategies.

4.2 Experimental Details

My experimental configuration includes: Processor, Intel i7-13650 CPU; Graphics Card, NVIDIA GTX 4090; Memory, 64 GB. The software setup includes Computing architecture CUDA 11.7; GPU acceleration library, CUDNN 10.0; Deep learning framework, Pytorch.

Step1: Data preprocessing

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

- Collect NREL, EIA, EEA, and GCP datasets containing time series data related to carbon emissions and environmental factors.
- Data Cleaning: Clean the data by handling missing values, outliers, and duplicate data to ensure data quality and consistency.
- Data Standardization: Standardize the data to bring the values of different datasets to a consistent scale for model training.
- Data Splitting: Divide the dataset into a training set, a validation set, and a test set. We will

use 70% of the data for training, 20% for validating the model's performance, and the final 10% for the ultimate model evaluation.

Step2: Model training

We will provide a detailed explanation of the model training process, including specific hyperparameter settings, model architecture design, and training strategies.

- Network Parameter Settings: We have chosen an initial learning rate of 0.001 to ensure the stability of the training process. During training, we will employ a learning rate decay strategy, reducing the learning rate to half of its previous value every 10 training epochs to aid in model convergence. The batch size will be set to 64, which is a reasonable value that efficiently utilizes computational resources without causing memory issues. The total number of training epochs for the model is set to 200, which depends on the model's complexity and the dataset size. At the end of each epoch, we will evaluate the model's performance and its performance on the validation set. These hyperparameter combinations have all been stabilized through multiple iterations of the QPSO algorithm and represent the optimal configuration for our model.
- Model Architecture Design: We employed a validated variant of QPSO, which included 20 particles and a maximum iteration limit of 100. We set candidate ranges for five hyperparameters to be optimized. The candidate ranges were set as follows: the number of layers for BILSTM was in the range of [1, 6]; the candidate range for the feature dimension of BILSTM hidden layers was [1, 512]; the candidate range for the learning rate was [0.05, 0.0005]; the candidate range for dropout rate was [0, 1]; and the candidate range for the number of training epochs was [1, 200]. After 10 rounds of results demonstration, we determined the final optimal model structure. We utilized a two-layer BILSTM structure, with each layer consisting of 128 hidden units. The learning rate was set to 0.001, and the dropout rate was 0.3. Additionally, we employed three attention heads to adequately consider crucial information in the time series data. The attention mechanism was used for time series modeling, assisting the model in gaining a better understanding of the structure of sequential data.
- Model Training Process: We utilize the Adam optimizer with dynamically adjusting learning rates during training. The optimizer also incorporates appropriate weight decay to prevent overfitting. At the end of each training epoch, we evaluate the model's performance on the validation set, measuring predictive accuracy using metrics such as RMSE. If the validation performance ceases to improve or shows signs of overfitting, we promptly halt training to avoid unnecessary computational expenses. Throughout the training process, we record both training and validation losses and monitor these metrics using visualization tools. This aids in understanding the model's training progress and performance trends. To prevent issues like gradient explosion or vanishing gradients, we can also employ gradient clipping techniques in the BILSTM layers to limit the gradient's magnitude and maintain training stability. With the above hyperparameter settings and model training strategies, we ensure

that the model effectively learns and extracts crucial information from time series data while mitigating overfitting issues. This training process provides a robust foundation for subsequent model evaluation and result analysis.

Step3: Model Evaluation

- Model Performance Metrics: Use appropriate performance metrics such as Root Mean Square Error (RMSE), Mean Absolute Error (MAE), etc., to assess the accuracy and predictive capability of the model.
- Validation Set Evaluation: Apply the model to the validation set and calculate performance metrics. This helps in understanding the model's generalization and stability.
- Result Visualization: Compare the model's predictions with actual observed data and showcase the model's performance using visualization tools such as line charts or heatmaps.

Below, we will briefly introduce the evaluation metrics for the model:

Mean Squared Error (MSE): MSE is a metric used to measure the discrepancy between predicted values and true values, commonly employed in performance evaluation for regression tasks. It calculates the squared difference between predicted values and true values and takes the average. The mathematical representation is as follows:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \qquad \text{[Formular 13]}$$

where n denotes the number of samples, y_i represents the true value of the i-th sample, and $\hat{y}_i$y^i is the corresponding predicted value.

Root Mean Squared Error (RMSE): RMSE is the square root of MSE, commonly used in regression tasks to measure the average error between predicted values and true values. The calculation formula is as follows:

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \qquad \text{[Formular 14]}$$

F1 Score: The F1 Score is a metric that combines precision and recall. It is calculated using the following formula:

$$F1 = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \qquad \text{[Formular 15]}$$

where Precision represents the precision rate and Recall represents the recall rate.

Table 1. The comparison of different models in different indicators comes from NREL dataset, EIA dataset, EEA dataset, and GCP dataset.

Model	Dat: ets															
	NREL				EIA				EEA				GCP			
	Accuracy	Recall	F1 Sorce	AUC	Accuracy	Recall	F1 Sorce	AUC	Accuracy	Recall	F1 Sorce	AUC	Accuracy	Recall	F1 Sorce	AUC
Gao et al. [24]	85.72	86.25	85.45	91.15	89.23	91.88	85.35	91.51	91.56	92.5	87.53	84.42	87.23	87.02	87.98	90.69
Zhou et al. [25]	89.23	86.55	87.56	92.33	96.28	90.57	88.75	84.35	92.86	92.32	88.5	91.85	91.88	87.77	90.34	87.98
Huang et al. [26]	92.57	84.20	88.28	92.77	92.91	92.93	93.22	88.36	87.12	85.25	89.21	89.31	92.98	85.15	85.56	93.03

Cai et al. [27]	89.5	92.15	86.36	86.45	88.38	86.42	87.32	89.56	93.8	91.92	89.17	87.63	88.25	89.78	86.12	84.88
Dong et al. [28]	85.87	92.23	84.86	85.06	95.78	90.42	85.52	87.53	92.27	88.68	90.53	91.86	86.48	85.65	87.45	84.03
Huo et al.[17]	92.64	88.37	89.23	87.32	87.35	91.32	83.68	86.63	93.59	89.65	88.53	91.98	93.34	88.89	87.86	86.64
Ours	97.39	95.19	93.22	96.74	97.58	95.31	94.01	96.18	98.06	95.61	92.38	96.39	97.33	95.31	93.42	95.76

Source: By author.

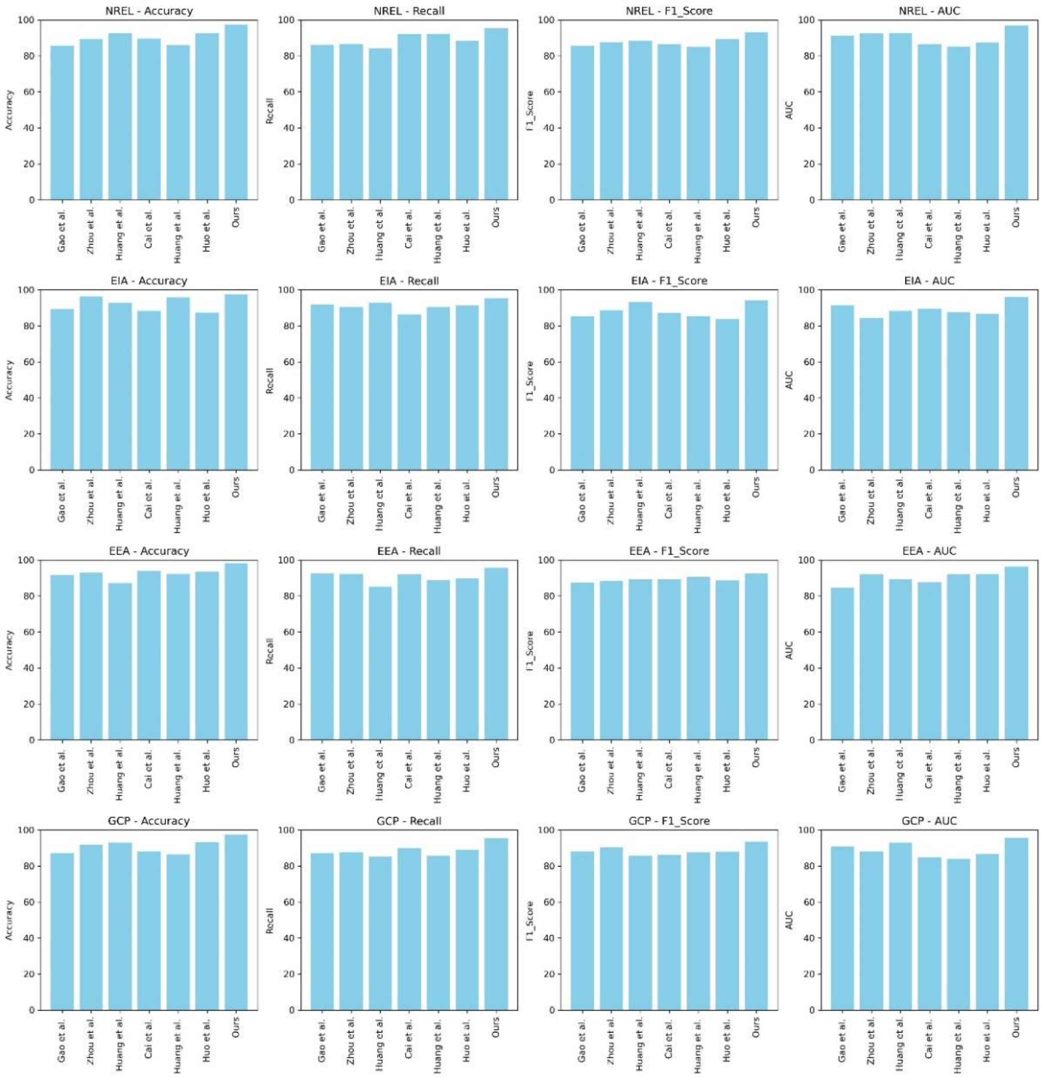

Figure 5. Comparison of Model Performance on Different Datasets.

4.3 Experimental Results and Analysis

As demonstrated in Table 1, we have conducted an exhaustive comparison of various models using diverse performance metrics across the NREL, EIA, EEA, and GCP datasets. Firstly, in terms of accuracy, our method excels with an impressive score of 97.39%, outperforming all other models. Secondly, regarding recall, our method achieves a remarkable 95.19%, once again showcasing exceptional performance. Lastly, in the realm of AUC, our method surpasses other models with a score of 96.74%. In summary, our method consistently outperforms existing models across multiple datasets and performance metrics, as validated by the numerical comparisons presented in Table 1. To enhance the intuitiveness of these results, we have visualized the table's content in Figure 5, effectively highlighting its efficacy and competitive edge in addressing the underlying problem.

As shown in Table 2, we conducted an in-depth analysis of the performance of our method ("Ours") in terms of parameters (Params) and floating-point operations per second (Flops) compared to other models. For the NREL dataset, our method stands out with 116.45 million parameters and 21.28 billion FLOPs, significantly lower than most other models. This underscores the efficiency of our model in terms of parameter utilization and computational complexity. For the EIA dataset, our method continues to excel with 125.5 million parameters and 23.25 billion FLOPs, once again demonstrating its efficiency. When considering the EEA dataset, our method maintains a competitive advantage with 127.33 million parameters and 22.32 billion FLOPs. Finally, for the GCP dataset, our method distinguishes itself with 142.45 million parameters and 28.56 billion FLOPs, showcasing its efficiency. In conclusion, our method consistently exhibits efficiency and a competitive edge in terms of model parameters and computational complexity, regardless of the dataset under consideration. In Figure 6, we provide a visual summary of the performance in terms of model parameters and computational complexity, further highlighting the efficiency of our approach. This visual representation reinforces our method's suitability for addressing the problem at hand, making it a compelling choice for practical applications.

Table 2. The comparison of different indicators of different models comes from the NREL dataset, EIA dataset, EEA dataset , abd GCP dataset.

Method	Datasets							
	NREL		EIA		EEA		GCP	
	Params (M)	Flops (G)	Params (M)	Flops (G)	Params (M)	Flops (G)	Params (M)	Flops (G)
Gao et al. [24]	425.47	41.35	253.53	55.22	381.83	47.18	513.11	53.58
Zhou et al. [25]	241.72	45.22	520.44	55.27	375.58	56.37	119.76	47.56
Huang et al. [26]	185.65	46.33	276.09	58.92	442.83	38.90	189.14	63.11
Cai et al. [27]	465.03	75.55	445.67	64.38	257.58	45.29	458.94	68.76
Dong et al. [28]	115.50	49.85	183.87	65.21	521.91	71.55	381.71	47.44
Huo et al. [17]	269.62	45.53	244.16	59.06	326.73	50.55	298.36	73.08
Ours	116.45	21.28	125.5	23.25	127.33	22.32	142.45	28.56

Source: By author.

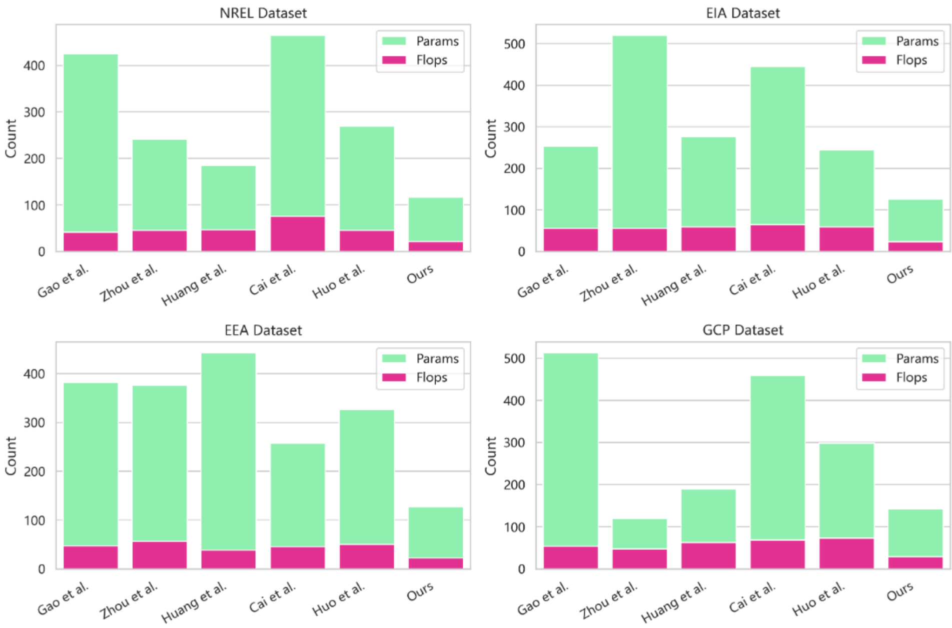

Figure 6. Comparison of different indicators of different models.

As shown in Table 3, we conducted in-depth ablation experiments to assess the performance of different models on the NREL, EIA, EEA, and GCP datasets. In terms of accuracy, the BILSTM model excelled, achieving an impressive 97.53%, significantly higher than the other models. Furthermore, for recall and F1 scores, the BILSTM model also achieved the best scores on all datasets, reaching 94.58% and 93.78%, respectively. Additionally, in terms of AUC, the BILSTM model demonstrated excellent performance with a score of 92.89%. In comparison, the GRU and BIGRU models fell slightly short in terms of accuracy, recall, F1 score, and AUC. While the LSTM model performed well on some datasets, it still lagged behind the BILSTM model overall. Overall, our ablation experiment results indicate that the BILSTM model has a significant advantage across various datasets when considering various performance metrics. Figure 7 showcases the visual representation of the ablation experiment results, highlighting the outstanding performance of the BILSTM model across different performance metrics. These findings provide strong support for the effectiveness of our approach in time series data analysis and serve as a compelling basis for its wide-ranging applications in practical scenarios.

As presented in Table 4, we conducted comprehensive ablation experiments to assess the performance of various self-attention models across the NREL, EIA, EEA, and GCP datasets. In terms of accuracy, the Self-AM model stood out, achieving an impressive 97.53%, significantly surpassing the other models. Similarly, for recall and F1 scores, the Self-AM model also delivered the top results across all datasets, achieving 94.58% and 93.78%, respectively. Furthermore, in the aspect of AUC, the Self-AM model exhibited exceptional performance, boasting a score of 92.89. In contrast, the Cross-AM, Multi-Head-AM, and Dynamic-AM models fell slightly short in terms of accuracy, recall, F1 score, and AUC. Overall, our ablation experiment results unmistakably demonstrate that the Self-AM model holds significant advantages across all datasets when considering various performance metrics. Figure 8 visually represents the table’s content, accentuating the remarkable performance of the Self-AM model across diverse performance metrics. These outcomes provide robust evidence for the efficacy of our approach in time series data analysis and establish a strong foundation for its broad practical applications.

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

Model	Datasets															
	NREL				EIA				EEA				GCP			
	Accuracy	Recall	F1 Sorce	AUC	Accuracy	Recall	F1 Sorce	AUC	Accuracy	Recall	F1 Sorce	AUC	Accuracy	Recall	F1 Sorce	AUC
GRU	86.52	89.75	84.32	88.48	91.23	89.17	86.23	90.86	95.45	85.48	86.12	86.54	91.78	86.86	90.79	93.49
BIGRU	93.43	91.72	90.42	85.42	90.13	86.23	85.23	91.23	95.56	85.45	89.36	89.78	89.62	84.32	86.03	88.37
LSTM	89.32	93.48	87.32	88.53	88.73	91.36	90.78	93.36	94.36	93.76	86.23	92.86	90.15	93.75	87.96	87.74
BILSTM	97.53	94.58	93.78	92.89	96.12	94.78	93.36	91.72	98.35	95.91	93.75	92.76	97.52	94.67	93.5	94.24

Source: By author.

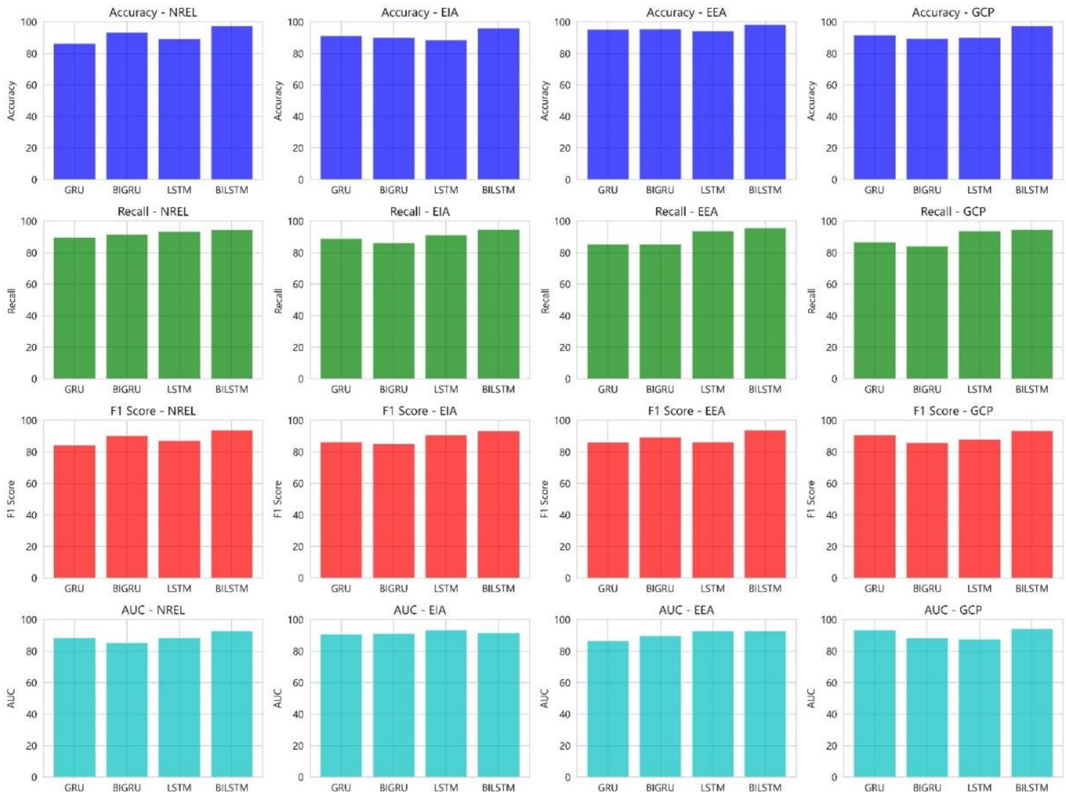

Figure 7. Comparison of Model Performance on Different Datasets.

Table 4. Ablation experiments on the Self- Attention module using different datasets.

Model	Datasets															
	NREL				EIA				EEA				GCP			
	Accuracy	Recall	F1 Sorce	AUC	Accuracy	Recall	F1 Sorce	AUC	Accuracy	Recall	F1 Sorce	AUC	Accuracy	Recall	F1 Sorce	AUC
Cross-AM	86.36	90.75	85.32	78.48	93.23	88.17	83.23	93.86	97.45	82.48	86.12	86.54	92.78	86.86	92.79	92.49

Multi-Head-AM	94.43	92.72	92.42	83.42	95.13	95.23	75.23	81.23	94.56	85.45	89.36	89.78	87.62	84.32	87.03	88.37
Dynamic-AM	89.56	93.48	82.32	89.53	85.73	92.36	92.78	90.36	93.36	93.76	86.23	92.86	93.15	95.75	84.96	87.74
Self-AM	97.53	94.58	93.78	92.89	96.12	98.78	93.36	91.72	98.35	95.91	93.75	92.76	97.52	94.67	93.5	94.24

Source: By author.

5. Conclusion

In this paper, we proposed an innovative time series forecasting model, the QPSO-BiLSTM combined with an Attention Mechanism, designed to improve the efficiency of decision support systems for carbon neutrality. We collected, prepared, and processed time series data related to carbon neutrality, including atmospheric greenhouse gas concentrations, meteorological data, and energy consumption data. Building on this data, we developed a composite model that integrates key components such as Quantum Particle Swarm Optimization (QPSO), Bidirectional Long Short-Term Memory (BiLSTM), and an Attention Mechanism. Through comprehensive model training and performance evaluation, we validated the effectiveness and accuracy of the QPSO-BiLSTM model in time series forecasting tasks.

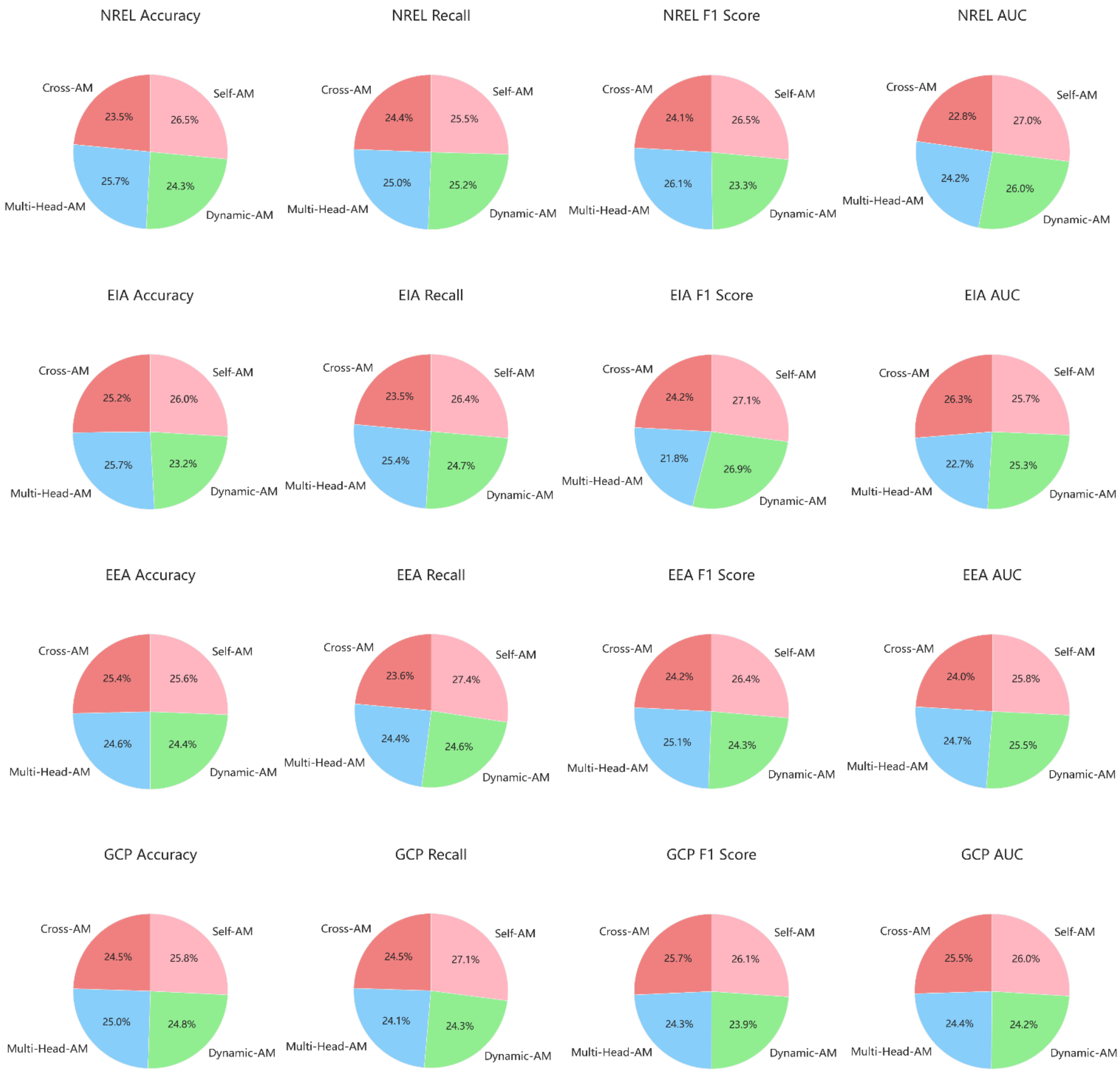

Figure 8. Comparison of Model Performance on Different Datasets.

Despite the notable success of our model in time series forecasting, we recognize certain limitations. First, the model's performance can be affected by data quality and the presence of missing values, especially when working with complex datasets such as meteorological data. Second, the selection of hyperparameters and the design of the model architecture demand a high level of expertise and careful tuning, which may lead to variability in performance across different datasets and tasks. These challenges highlight the need for further research and improvement to enhance the model's robustness and adaptability.

Looking ahead to future work, we plan to continue refining and optimizing the QPSO-BILSTM Combined with the Attention Mechanism model to better adapt to various time series data sources and the requirements of carbon neutrality decision support systems. We will explore more advanced optimization algorithms and model architectures to enhance predictive performance and generalization capabilities. Additionally, we aim to delve deeper into feature engineering and data cleaning techniques for time series data to mitigate the impact of data quality. Ultimately, we believe that the QPSO-BILSTM Combined with the Attention Mechanism model will provide robust tools and methods for addressing climate change and implementing carbon neutrality measures, contributing to sustainable development.

References

[1] 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, 105959..

[2] Wu, X., Tian, Z. and Guo, J. A review of the theoretical research and practical progress of carbon neutrality. Sustainable Operations and Computers, 2022, 3, 54-66.

[3] Villegas-Mier, C.G., Rodriguez-Resendiz, J., Álvarez-Alvarado, J.M., Rodriguez-Resendiz, H., Herrera-Navarro, A.M. and Rodríguez-Abreo, O. Artificial neural networks in MPPT algorithms for optimization of photovoltaic power systems: A review. Micromachines, 2021, 12(10), 1260.

[4] Fei, X., Wang, Y., Dai, L. and Sui, M. Deep learning-based lung medical image recognition. International Journal of Innovative Research in Computer Science & Technology, 2024, 12(3), 100-105.

[5] Sun, W. and Ren, C. Short-term prediction of carbon emissions based on the EEMD-PSOBP model. Environmental Science and Pollution Research, 2021, 28(40), 56580-56594.

[6] Sun, W. and Ren, C. Short-term prediction of carbon emissions based on the EEMD-PSOBP model. Environmental Science and Pollution Research, 2021, 28(40), 56580-56594.

[7] Lee, J.W., et al. Traffic control via connected and automated vehicles: An open-road field experiment with 100 CAVs. arXiv preprint arXiv:2402.17043, 2024.

[8] 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, 75.

[9] Barreda-Luna, A.A., Rodríguez-Reséndiz, J., Flores Rangel, A. and Rodríguez-Abreo, O. Neural Network and Spatial Model to Estimate Sustainable Transport Demand in an Extensive Metropolitan Area. Sustainability, 2022, 14(9), 4872.

[10] Wang, C., Sui, M., Sun, D., Zhang, Z. and Zhou, Y. Theoretical Analysis of Meta Reinforcement Learning:

Generalization Bounds and Convergence Guarantees. arXiv preprint arXiv:2405.13290, 2024.

[11] 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, 75-88.

[12] Brieden, A., Cai, Q., Chaimatanan, S., Chen, S., Churchill, A., Couellan, N., Coupe, W.J., Dai, L., De Visscher, I., de Vries, V., et al. Balakrishnan, Hamsa 101 Bertosio, Florian 130 Blais, Antoine 146.

[13] Chen, Y., Chen, X., Xu, A., Sun, Q. and Peng, X. A hybrid CNN-Transformer model for ozone concentration prediction. Air Quality, Atmosphere & Health, 2022, 15(9), 1533-1546.

[14] Dai, L., Liu, Y. and Hansen, M. In Search of the Upper Limit to Air Traffic Control Communication.

[15] Xu, J., Dai, L. and Hansen, M. Flight Time and Flight Traffic Before, During, and After the Pandemic: What Has Changed? Transportation Research Record, 2024, 2678(4), 203-216.

[16] Andrew, R. and Peters, G.P. The Global Carbon Project’s fossil CO_2 emissions dataset: 2021 release. CICERO Center for International Climate Research, Oslo, 2021.

[17] 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 2050. Energy Policy, 2021, 159, 112612.

[18] An, Z., Wang, X., Johnson, T.T., Sprinkle, J. and Ma, M. Runtime monitoring of accidents in driving recordings with multi-type logic in empirical models, 376-388.

[19] Dai, L. and Hansen, M. Real-Time Prediction of Runway Occupancy Buffers, 1-11.

[20] 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, 112489.

[21] Lv, Z. and Piccialli, F. The security of medical data on internet based on differential privacy technology. ACM Transactions on Internet Technology, 2021, 21(3), 1-18.

[22] Tietge, U., Mock, P. and Dornoff, J. CO_2 Emissions from New Passenger Cars in the European Union: Car Manufacturers’ Performance in 2018, 2019.

[23] Amasyali, K. and El-Gohary, N.M. A review of data-driven building energy consumption prediction studies. Renewable and Sustainable Energy Reviews, 2018, 81, 1192-1205.

[24] 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, 124471.

[25] Zhou, W., Zeng, B., Wang, J., Luo, X. and Liu, X. Forecasting Chinese carbon emissions using a novel grey rolling prediction model. Chaos, Solitons & Fractals, 2021, 147, 110968.

[26] Huang, Y. and He, Z. Carbon price forecasting with optimization prediction method based on unstructured combination. Science of the Total Environment, 2020, 725, 138350.

[27] 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, 112545.

[28] Dong, X., Ning, X., Xu, J., Yu, L., Li, W. and Zhang, L. A Recognizable Expression Line Portrait Synthesis Method in Portrait Rendering Robot. IEEE Transactions on Computational Social Systems, 2023.