Temperature Time Series Prediction Using Convolutional Bidirectional LSTM Optimized by FPA

Shuo Lv[1] , Bixian Cheng[1], Ji Li[1], Qing Shen[1*]

1. Department of Information Engineering, Wuhan Business University, Hubei, China

Corresponding Author: Qing Shen   Email: shenqing0611@126.com

***Abstract*—As global climate change intensifies, accurate temperature predictions have become increasingly important across various fields, including agriculture, meteorology, and environmental protection. This paper proposes a temperature time series prediction model that integrates a convolutional neural network (CNN) with a bidirectional long short-term memory network (BiLSTM), utilizing the flower pollination algorithm (FPA) for hyperparameter optimization. The model effectively captures bidirectional dependencies within sequence data and enhances prediction accuracy through the global optimization capabilities of the FPA algorithm. In a case study focused on temperature prediction in Beijing, the model exhibited superior prediction accuracy and faster convergence compared to traditional deep learning models. The findings offer a novel perspective on regional temperature prediction and highlight the model's potential for practical applications.**

Keywords—Temperature Forecasting;Convolutional Neural Network;Long Short-Term Memory Network;Flower Pollination Algorithm

I. INTRODUCTION

As global climate change intensifies, accurate temperature forecasting becomes increasingly important in sectors like agriculture, meteorology, and environmental protection.

Sepp Hochreiter and Jürgen Schmidhuber [1] initially introduced Long Short-Term Memory (LSTM). LSTM is widely utilized in forecasting applications. Juan M. Esparza-Gómez et al. [2] employed LSTM alongside extreme gradient boosting to forecast greenhouse temperatures; Hua Fan et al. [3] leveraged LSTM for modeling and predicting temperature time series. Compared to traditional machine learning algorithms, LSTM, while adept at capturing temporal correlations, faces limitations in processing spatially dependent and high-dimensional data in tasks like temperature forecasting that necessitate the simultaneous consideration of time and space characteristics, leading to inadequate prediction accuracy and stability. The CNN-LSTM model, integrating the strengths of both CNN and LSTM, demonstrates superior prediction accuracy and stability. Consequently, Jonathan Cahyadi et al. [4] developed a Bitcoin price prediction model utilizing CNN and LSTM. I Wayan Krisna Gita Santika et al. [5] employed a CNN-LSTM approach to predict gold prices. C.J. Chen et al. [6] combined convolutional CNN with LSTM for aviation visibility forecasting. İhsan Uluocak et al. [7] applied LSTM-CNN and GRU-CNN models in daily temperature forecasting. The LSTM within the CNN-LSTM model is unidirectional, capturing only forward dependencies in sequence data, unlike BiLSTM, which is capable of capturing both forward and backward dependencies simultaneously. BiLSTM's bidirectional nature allows for a more comprehensive capture of information in time series data, an advantage for temperature forecasting tasks requiring consideration of historical and future trends. This also endows it with stronger processing capabilities for long sequences. Ahmed Bahaa Farid et al. [8] employed a hybrid CNN-BiLSTM model for software defect prediction. Shuxin Liu et al. [9] forecasted the residual electrical life of railway relays utilizing a CNN-BiLSTM approach. CNN-BiLSTM hybrid models have been extensively applied across various forecasting domains, yet temperature forecasting remains an underexplored area.

In this study, we integrate CNN and LSTM to construct a bidirectional long short-term memory (BiLSTM) network, which captures the bidirectional dependencies within sequence data. Additionally, we employ the FPA algorithm to fine-tune the model's hyperparameters, aiming to achieve optimal predictive performance. The model is then applied to predict temperatures in Beijing. Ultimately, we compare the model's performance, along with its advantages and disadvantages, with other models to demonstrate its superiority and offer insights for regional temperature forecasting.

II. RESEARCH METHODS

A. *Deep Learning Related Technologies*

1) Convolutional Neural Network (CNN)

CNN has been widely used in weather and climate research and performs well[10]. CNN can extract local features from input data by utilizing local receptive fields, weight sharing, and pooling operations, thereby eliminating redundant information through pooling to streamline input for subsequent models[11]. The CNN model is composed of convolutional layers, maximum pooling layers (MaxPooling1D), and fully connected layers (Dense). The convolutional layer uses multiple convolution kernels to linearly transform the input data and add biases, thus extracting local features; the pooling layer reduces the data dimension through downsampling, ensuring that important feature information is retained; the fully connected layer integrates the features extracted by the previous layers and ultimately outputs a continuous value, which is the predicted maximum temperature.

2) Long Short-Term Memory Model(LSTM)

The LSTM is a neural network model that incorporates three gating mechanisms: the forget gate, the input gate, and the output gate. It boasts an excellent ability to capture long-term dependencies, along with flexibility, adaptability, and stability[12]. It can adjust predictions by learning from historical patterns, thereby modeling non-stationary relationships between variables. The LSTM can identify complex patterns and relationships within data, particularly those involving nonlinear relationships with long-term dependencies[13]. The bidirectional long short-term memory network (BiLSTM), which is based on the LSTM, captures bidirectional dependencies of sequences by integrating information from two LSTM layers(one processing future sequence values and the other processing past sequence values)[14].

B. C-BiLSTM Model Structure Introduction

The C-BiLSTM model offers significant advantages for processing one-dimensional time series data [15]. This approach, a deep learning model, amalgamates the strengths of convolutional neural network (CNN) and bidirectional long short-term memory networks(BiLSTM). Within this model, the CNN captures local features of the input data and compresses the feature count through convolutional and pooling layers, subsequently employing the BiLSTM network to glean more comprehensive time series information from the CNN's output, factoring in air temperature. Ultimately, the BiLSTM output is fed into the fully connected layer for temperature prediction.

C. Flower Pollination Algorithm(FPA)

The FPA algorithm, an intelligent bionic optimization algorithm, simulates the natural phenomena of flower pollination and reproduction. It emulates both self-pollination and cross-pollination processes found in nature [16]. Within the flower pollination algorithm, the search process is bifurcated into two phases:Local search and global search. Local pollination facilitates effective information exchange among neighboring pollen grains due to the proximity. This local pollination is characterized by:

$$X_i^{t+1} = X_i^t + \varepsilon(X_j^t - X_k^t) \quad (1)$$

Among them, X_j^t and X_k^t are the coordinates of different pollens from similar flowering plants, effectively two solution vectors randomly selected from the entire solution space; ε is a random variable within the interval [0,1] that follows a uniform distribution. The global search stage emulates the long-distance pollination, akin to the cross-pollination facilitated by insects, birds, and other organisms. Global pollination can be characterized by:

$$\mathrm{X}_\mathrm{i}^{\mathrm{t+1}} = \mathrm{X}_\mathrm{i}^\mathrm{t} + \theta \mathrm{L}(\mathrm{X}_*^{\mathrm{best}} - \mathrm{X}_\mathrm{i}^\mathrm{t}) \quad (2)$$

θ represents the step size scaling factor; X_*^{best} represents the global optimal solution of the current iteration; L represents the global pollination step size, which follows the Lévy distribution.

D. FPA-C-BiLSTM Deep Learning Model Design

The PFA-C-BiLSTM model not only leverages the strengths of convolutional neural network (CNN) and bidirectional long short-term memory network (BiLSTM), but also employs the global optimization capabilities of the flower pollination algorithm (FPA) to fine-tune the parameters of the CNN-BiLSTM, thereby enhancing the accuracy of the prediction outcomes. The overall process is depicted in Figure 1. The CNN-BiLSTM network model primarily consists of one-dimensional convolutional layer, one-dimensional pooling layer, BiLSTM layer, and fully connected layer. The training process of the entire prediction model can be broken down into the following steps:

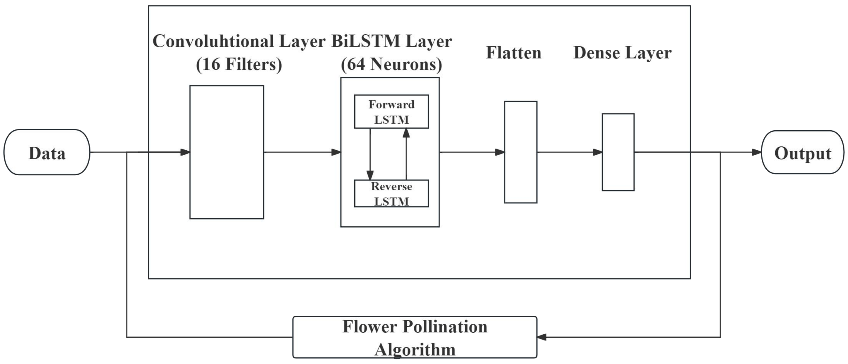

Fig. 1. FPA-C-BiLSTM model structure diagram

(1) Arrange the n-dimensional feature data into a one-dimensional matrix with n rows and columns, and then partition it into a training set and a test set;

(2) Utilize the convolutional and pooling layers to extract features from the training data, followed by feeding the resulting sequence data into the BiLSTM layer;

(3) Calculate the error and output after each iteration, and continuously refine the model parameters to ensure convergence of the error;

(4) Employ the test set to validate the model's predictive outcomes and utilize appropriate evaluation metrics to assess the model's performance.

III. RESULTS AND ANALYSIS

A. Data Collection And Preprocessing

The original dataset is sourced from the National Climatic Data Center of the United States, encompassing surface temperature data for Beijing from January 1, 2017, to December 31, 2021. It includes statistical parameters such as date, temperature, atmospheric pressure, wind direction, wind speed, and cloud coverage, sampled every three hours.

The dataset may contain missing or invalid data, potentially impacting the precise estimation of temperature accuracy. For temperatures outside the -50 to 50 degrees range, null and missing values are replaced with the preceding statistical data, resulting in a total of 14,432 processed data points. Time series features are constructed using the sliding window method, where data from the past 10 time steps are extracted as feature X for each window, and the corresponding target value y is generated. The dataset is split into an 80% training set and a 20% test set, with the processed features X and target y allocated accordingly. The input data for both sets are formatted as (number of samples, time step, number of features) to align with the model's input specifications. This processing method enables the model to effectively capture patterns within the time series data.

B. Model Evaluation Indicators

This study employs three metrics—root mean square error (RMSE), mean absolute error (MAE), and mean square error (MSE)—to assess the performance of the proposed FPA-C-BiLSTM model and other neural network models. RMSE quantifies the accuracy of prediction results, with lower values indicating more precise outcomes. MAE gauges the consistency of the prediction results, where lower values signify greater consistency. MSE serves as another measure of accuracy, with lower values indicating a better model prediction effect. These metrics will be used in conjunction to evaluate the model's predictive performance. The calculation formulas are as follows:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y_i})^2} \tag{3}$$

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y_i}| \tag{4}$$

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y_i})^2 \tag{5}$$

Among them, y_i is the true value, $\hat{y_i}$ is the predicted value, and n is the number of samples.

C. Prediction Results And Analysis

This study employs the FPA-C-BiLSTM model to forecast temperature datasets and compares its performance with classic models, including CNN and LSTM, as well as the pre-optimization C-BiLSTM model. The model's input shape is configured as (10, 1). The one-dimensional convolutional layer utilizes 16 filters and 2 convolution kernels, applying the ReLU activation function to extract features. The bidirectional long short-term memory (BiLSTM) layer is set with 32 units, also using the ReLU activation function. The flattened multidimensional data is then fed into a fully connected layer comprising 8 neurons, activated by ReLU. The output layer utilizes a linear activation function to produce the prediction outcomes. The model is compiled with the Adam optimizer and utilizes mean square error as the loss function. Training is executed over 20 epochs, with model performance tracked via the test set, ensuring the best results are retained.

Simultaneously, the pollen propagation algorithm (FPA) was utilized to optimize key model parameters, specifically the learning rate and batch size of the Adam optimizer, in pursuit of the optimal solution. The learning rate was varied between 10-5 and 10-2, and the batch size was explored within the range of 32 to 128. The FPA was configured with a population size of 10 and an iteration limit of 20. Through the iterative optimization process, the FPA identified the optimal parameters as a learning rate of 0.006238 and a batch size of 62, with the corresponding minimum Loss value of 2.92831. These optimal parameters were then applied to the model for prediction, with the results depicted in Figure 2. The alignment of the red predicted values with the red actual values in the figure suggests the effectiveness of the predictions. The model's average performance metrics, expressed as Mean Squared Error (MSE), Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE), were 3.26475, 1.29161, and 1.80686, respectively.

Furthermore, to validate the effectiveness of the global temperature prediction model founded on the FPA-C-BiLSTM neural network, this study trained separate models using CNN, LSTM, and C-BiLSTM architectures, and their convergence behaviors are presented in Figure 4.

All four models exhibited stability after the 10th iteration, with the FPA-C-BiLSTM model demonstrating a further reduction in loss value compared to the C-BiLSTM model, achieving convergence more rapidly. This observation indicates that the FPA-C-BiLSTM model not only converges more swiftly but also achieves higher predictive accuracy than the other three models.

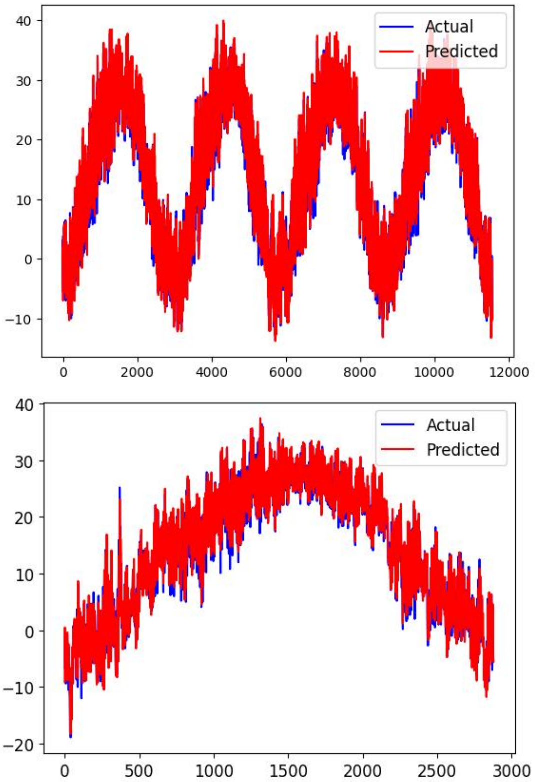

Fig. 2. FPA-C-BiLSTM model training result

The average training times of the four models are presented in Table I. It was observed that the CNN model had the shortest average iteration time, at 9.56 seconds. In contrast, the C-BiLSTM model exhibited the longest training time, amounting to 76.70 seconds. The FPA-optimized C-BiLSTM model reduced its training time to 51.63 seconds, which is a 32.69% reduction from its pre-optimization duration.

After multiple training sessions, the mean square errors (MSE) of the four models were calculated and box plot 3 was generated. As depicted in box plot 3, the LSTM model exhibits the highest MSE value distribution, signifying its largest prediction error among the models. In comparison to the classic deep learning models, the C-BiLSTM model demonstrates superior accuracy. The FPA-C-BiLSTM model, however, presents the lowest MSE value distribution, with its predictions being more tightly clustered than those of the pre-optimization C-BiLSTM model. This suggests that the FPA-C-BiLSTM model possesses enhanced stability and predictability.

TABLE I. COMPARISON OF THE TIME SPENT ON TRAINING EACH MODEL

	CNN	LSTM	C-BILSTM	FPA-C-BILSTM
Training Time(s)	9.56	49.01	76.70	51.63

The evaluation metrics for the four models are presented in Table II. The FPA-C-BiLSTM model demonstrates a reduction in MSE, MAE, and RMSE of 7.560%, 6.048%, and 3.854%, respectively, compared to the CNN model. When compared to the LSTM model, these reductions are 10.170%, 5.701%, and 5.221%, respectively. Additionally, the FPA-C-BiLSTM model shows a reduction of 4.730%, 4.000%, and 2.394% in MSE, MAE, and RMSE, respectively, when contrasted with the C-BiLSTM model. Consequently, the model presented in this paper outperforms the aforementioned models in overall performance.

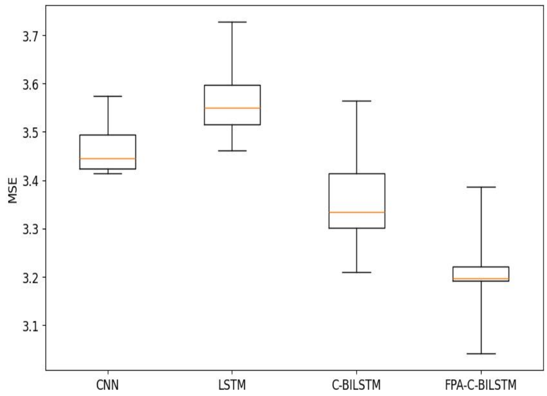

Fig. 3. Comparison of errors among models

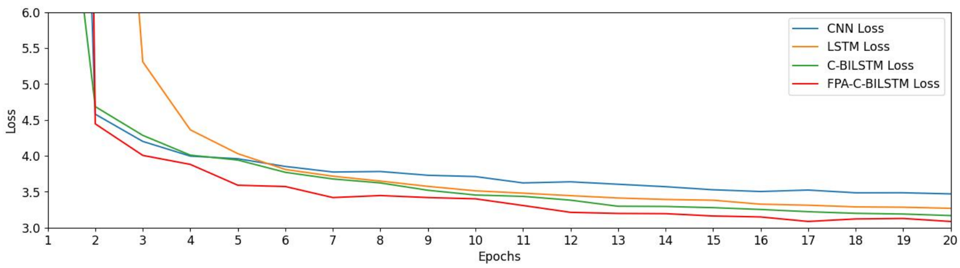

Fig. 4. Convergence efficiency and convergence degree comparison of different models

TABLE II. AVERAGE RESULTS OF MODEL EVALUATION INDICATORS

	MSE	MAE	RMSE
CNN	3.46280	1.37475	1.86086
LSTM	3.56341	1.36969	1.88769
C-BiLSTM	3.35996	1.34543	1.83302
FPA-C-BiLSTM	3.20102	1.29161	1.78914

IV. CONCLUSIONS

In this research, the FPA-C-BiLSTM model was utilized to forecast temperatures in Beijing. The experimental outcomes indicate that this model surpasses conventional deep learning models across several indicators. By integrating convolutional neural networks and bidirectional long short-term memory networks, and employing the pollen propagation algorithm for parameter optimization, the model effectively enhances prediction accuracy. Overall, the FPA-C-BiLSTM model excels in temperature prediction tasks and demonstrates a significant advantage in accuracy over other structural prediction models. Moreover, it reduces computation time while increasing prediction accuracy compared to the pre-optimization C-BiLSTM model, indicating high potential for practical application. Temperature predictions may also be influenced by variables such as weather conditions, cloud cover, and atmospheric pressure. In future endeavors, we intend to augment the model's generalization capabilities by incorporating additional meteorological data, thereby offering a more holistic approach to temperature forecasting.

ACKNOWLEDGMENT

This work was supported by the Research Project on the Protection, Inheritance and Promotion of Yangtze River Culture in Hubei Province in 2024 (HCYK2024Y57) and the Student Innovation and Entrepreneurship Program of Wuhan Business University (202211654009).

REFERENCES

[1] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long Short-Term Memory. Neural Comput. 9, 8 (November 15, 1997), 1735-1780.

[2] Juan M. Esparza-Gómez, Luis F. Luque-Vega, Héctor A. Guerrero-Osuna, et al. Long Short-Term Memory Recurrent Neural Network and Extreme Gradient Boosting Algorithms Applied in a Greenhouse's Internal Temperature Prediction. Applied sciences. 2023;13 (22):12341-12341.

[3] Hua Fan, Li Li, Cai Xinnan, et al. Application of Long Short-Term Memory Network in Temperature Prediction. Intelligent Computing and Applications, 2022, 12(11):92-95+102.

[4] Jonathan Cahyadi, Amalia Zahra. Bitcoin Price Prediction Model Development Using Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM). Journal La Multiapp. 2024;5 (2):52-62.

[5] I Wayan Krisna Gita Santika, Siti Saadah, Prasti Eko Yunanto. Gold price prediction using Convolutional Neural Network-Long Short-Term Memory (CNN-LSTM). Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control. 2021;0 (0):0-0.

[6] C.J. Chen, Chien-Hua Huang, Shih-Ming Yang. Aviation visibility forecasting by integrating Convolutional Neural Network and long short-term memory network. Journal of Intelligent and Fuzzy Systems. 2023;45 (3):5007-5020.

[7] İhsan Uluocak, Mehmet Bilgili. Daily air temperature forecasting using LSTM-CNN and GRU-CNN models. Acta Geophysica. 2023;0 (0):0-0.

[8] Ahmed Bahaa Farid, E. Fathy, Ahmed Sharaf Eldin, et al. Software defect prediction using hybrid model (CBIL) of convolutional neural network (CNN) and bidirectional long short-term memory (Bi-LSTM). PeerJ. 2021;7 (0):e739-e739.

[9] Shuxin Liu, Yankai Li, Shuyu Gao, et al. Prediction of Residual Electrical Life in Railway Relays Based on Convolutional Neural Network Bidirectional Long Short-Term Memory. Energies. 2023;16 (17):6357-6357.

[10] Ali J, Cheng L. Temperature forecasts for the continental United States: a deep learning approach using multidimensional features. Frontiers in Climate, 2024, 6.

[11] Wang Yi, Yang Jianbo, Li Rui, et al. CNN-GRU Temperature Forecasting Model Based on Spatiotemporal Sequence. Computer Applications, 2023, 43(S2):54-59.

[12] Yan Jinfu, He Qijun, Qu Yefeng, et al. Research on Damage TFM Localization and Detection of Plate Based on CNN-BiLSTM and ResNet Network. Journal of Nanjing University (Natural Sciences), 2024, 60(04):566-576.

[13] Abdullah M M A, Achira A, Jessica M, et al. Predicting the performance of green stormwater infrastructure using multivariate long short-term memory (LSTM) neural network. Journal of Hydrology, 2023, 625(PA).

[14] Heshan Wang, Yiping Zhang, Jian-Gang Liang, et al. DAFA-BiLSTM: Deep Autoregression Feature Augmented Bidirectional LSTM network for time series prediction. Neural Networks, 2023;157 (0):240-256.

[15] Yan Jinfu, He Qijun, Qu Yefeng, et al. Research on Damage TFM Localization and Detection of Plate Based on CNN-BiLSTM and ResNet Network. Journal of Nanjing University (Natural Sciences), 2024, 60(04):566-576.

[16] Shi Tao, Xiong Teng, Zhao Lingzhu. A Comprehensive Review of Pollen Germination Algorithm. Software Guide, 2023, 22(04):245-252.