Investigation of a transformer-based hybrid artificial neural networks for climate data prediction and analysis

Shangke Liu[1], Ke Liu[2]*, Zheng Wang[1], Yuanyuan Liu[1], Bin Bai[1] and Rui Zhao[1]

[1]State Grid Ningxia Electric Power Co. Ltd., Eco-tech Research Institute, Yinchuan, Ningxia, China,
[2]Shandong Chengxin Engineering Construction & Consulting Co. Ltd., Jinan, Shangdong, China

Introduction: Climate change isone of the major challenges facing the world today, causing frequent extreme weather events that significantly impact human production, life, and the ecological environment. Traditional climate prediction models largely rely on the simulation of physical processes. While they have achieved some success, these models still face issues such as complexity, high computational cost, and insufficient handling of multivariable nonlinear relationships.

Methods: In light of this, this paper proposes a hybrid deep learning model based on Transformer-Convolutional Neural Network (CNN)-Long Short-Term Memory (LSTM) to improve the accuracy of climate predictions. Firstly, the Transformer model is introduced to capture the complex patterns in cimate data time series through its powerful sequence modeling capabilities. Secondly, CNN is utilized to extract local features and capture short-term changes. Lastly, LSTM is adept at handling long-term dependencies, ensuring the model can remember and utilize information over extended time spans.

Results and Discussion: Experiments conducted on temperature data from Guangdong Province in China validate the performance of the proposed model. Compared to four different climate prediction decomposition methods, the proposed hybrid model with the Transformer method performs the best. The resuts also show that the Transformer-CNN-LSTM hybrid model outperforms other hybrid models on five evaluation metrics, indicating that the proposed model provides more accurate predictions and more stable fitting results.

KEYWORDS

climate prediction, sequentially, hybrid deep learning, CNN, LSTM, transformer

1 Introduction

Climate change, as one of the most important global issues of the 21st century, has attracted extensive attention from governments, scientists, and the public. The extreme weather events brought about by climate change, such as floods, droughts, heatwaves, and storms, have not only caused significant impacts on human society and the economy but also had profound

destructive effects on natural ecosystems Akpuokwe et al. (2024). Therefore, accurately analyzing and predicting climate change trends has become a focal issue for both academia and industry.Climate prediction methods are generally classified into five areas: (1) game theory models, (2) basic methods, (3) simplified models, (4) statistical models, and (5) artificial intelligence (AI) methods. Among these, statistical and AI approaches are most widely used due to their effectiveness. Statistical methods, among the earliest used in climate prediction, rely on historical time series data and external factors. Common techniques include Regression Models (RM) Sreehari and Ghantasala (2019), Autoregressive (AR) Broszkiewicz-Suwaj and Wyłomańska (2021), ARIMA Amjad et al. (2022), Transfer Function (TF) Ruigar and Golian (2016), GARCH Pandey et al. (2019), and ARMAX Sarhadi et al. (2014). However, climate time series are highly volatile and complex, complicating prediction tasks. Traditional statistical methods, usually based on linear assumptions, struggle to capture the nonlinear relationships in climate data. Climate time series often show significant fluctuations, noise, and time-varying properties, requiring more sophisticated models to adapt to these characteristics. Missing data further complicates predictions, as preprocessing is needed to address these gaps. External factors such as solar radiation, volcanic activity, and human activities add to the complexity of climate systems, making accurate predictions more challenging. AI methods, such as deep learning, offer enhanced modeling capabilities to address these issues. They are better suited for handling nonlinear patterns, high-frequency data, and changing climate conditions. However, these methods typically demand large datasets, significant computational resources, and may lack interpretability. Researchers often turn to more advanced algorithms to improve predictive accuracy despite these challenges. With the availability of data and the development of machine learning algorithms, the use of artificial intelligence methods for climate prediction has surged in recent years. Artificial intelligence methods include machine learning and deep learning. In the early 21st century, a number of papers attempted to use Artificial Neural Network (ANN) models to predict climate. Pande et al. used Support Vector Machines (SVM) as a basic model for climate prediction Pande et al. (2023). Weirich et al. achieved high accuracy using Random Forest models Weirich-Benet et al. (2023). Compared to machine learning methods, deep learning methods can analyze deep and complex nonlinear relationships through hierarchical and distributed feature representation Bauer et al. (2023). Several deep learning methods are widely used for climate prediction, including Deep Neural Networks (DNN), Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM), and Gated Recurrent Units (GRU). Among these deep learning methods, climate prediction often involves using LSTM and GRU models because they can address long-term dependencies and prevent gradient explosion issues.As climate science evolves and extreme weather events become more frequent due to climate change, the complexity of accurate climate prediction has intensified. Researchers have increasingly turned to hybrid models, which combine multiple machine learning approaches, to enhance prediction accuracy. These hybrid models consistently outperform traditional machine learning and deep learning models like SVM and LSTM Peng and Ni (2020). Among these, various feature selection methods play a crucial role in climate prediction. Convolutional Neural Networks (CNNs) have proven especially effective in feature extraction for hybrid models, as they excel at capturing both spatial and temporal

features from grid-like climate data, surpassing other machine learning architectures. Chen et al. (2019) introduced a CNN-LSTM hybrid model, which demonstrated superior accuracy, generalizability, and practicality compared to single models and other hybrid approaches. Similarly, a CNN-GRU hybrid model Han et al. (2023) also achieved strong predictive performance. While GRU offers competitive accuracy, LSTM has been more extensively researched and applied in a wider range of tasks, providing a wealth of resources and expertise, making it a favorable choice for many researchers. Thus, this paper selects CNN-LSTM as the core model. However, climate data contain mixed features, and selecting optimal model hyperparameters remains challenging, limiting the CNN-LSTM model's potential for further advancements in prediction accuracy. To address these limitations, feature decomposition can play a crucial role. Utilizing Transformers for data decomposition can help uncover hidden patterns and trends in climate sequences Ye et al. (2021). Additionally, optimizing hyperparameters can further refine predictions, leading to more accurate and reliable climate forecasting. This paper adopts the Transformer to process raw climate sequences and proposes a new hybrid method for climate prediction. By combining the strengths of each component, this method aims to improve accuracy and capture long-term dependencies in the data. The key innovation of our work is the application of the Transformer to data decomposition, enhancing the model's performance by effectively decomposing complex time series data into its underlying components. The combination of CNN and LSTM allows for the extraction of information features and the capture of local and long-term dependencies, respectively. The contributions of this paper can be summarized in the following three aspects:

- This paper introduces the Transformer model, which leverages its strong sequence modeling capabilities to decompose complex climate time series into trend, seasonality, and noise components, enhancing the model's accuracy and robustness in handling diverse patterns in the data.
- This paper employs Convolutional Neural Networks (CNNs) for local feature extraction from climate data, capturing local patterns and spatial features. This improves short-term predictions by focusing on short-term fluctuations and local variations in the climate data.
- This paper uses Long Short-Term Memory (LSTM) networks to capture long-term dependencies in climate data. LSTM effectively remembers and utilizes long-span historical information, improving the accuracy and stability of climate predictions by accounting for past data trends.

2 Related work

2.1 Importance of climate prediction

Climate prediction holds extremely important significance in modern society. Firstly, it can significantly enhance disaster prevention and mitigation capabilities. By providing early warnings of extreme weather events such as typhoons, floods, droughts, and snowstorms, relevant departments can take timely emergency measures, effectively reducing the damage caused by

natural disasters and safeguarding people's lives and property. Secondly, climate prediction plays a crucial role in agricultural production. Accurate climate forecasts can help farmers schedule planting and harvesting times reasonably, select suitable crop types and planting methods, improve agricultural production efficiency, and mitigate the adverse effects of climate change on agriculture, thereby ensuring food security Butt et al. (2024). Furthermore, climate prediction has significant impacts on water resource management and energy utilization. By forecasting precipitation and hydrological conditions, reservoir scheduling can be optimized, water resources can be allocated rationally, and water use efficiency can be improved. In the energy sector, especially the utilization of renewable energy such as wind and solar power, which highly depends on weather conditions, climate prediction can optimize energy dispatch, ensuring a stable energy supply. Moreover, climate prediction is indispensable for ecological protection and urban planning. Accurate climate forecasts can help ecological protection departments take measures in advance to maintain ecological balance, protect endangered species, and promote biodiversity. In terms of urban planning and infrastructure construction, understanding future climate trends aids in designing city layouts and infrastructures that are more adaptable to climate change, enhancing the sustainable development capacity of cities. Lastly, climate prediction is also of significant importance to public health. Climate change affects the spread of diseases and public health conditions. Through climate prediction, health departments can take preventive measures in advance to prevent and control the outbreak of climate-related diseases, thereby ensuring public health Sang et al. (2023). Therefore, climate prediction not only plays a critical role in responding to the direct impacts of climate change but also holds significant importance in promoting the coordinated and sustainable development of various social sectors.

2.2 Feature selection

Feature selection is a crucial step in climate prediction to avoid the curse of dimensionality. It involves selecting the most relevant and informative subset of features from the original dataset as inputs for machine learning or deep learning models Fahad et al. (2023). The aim of feature selection is to reduce the dimensionality of the data while retaining its important characteristics, thereby simplifying the task of pattern recognition and accurate prediction for machine learning or deep learning models. Research on feature selection algorithms is generally categorized into three types: filter, wrapper, and embedded methods Liu et al. (2023). Specifically, filter methods apply proxy measures to evaluate the relevance of features independently of the machine learning or deep learning model. Typical examples of filter methods include correlation-based feature selection, mutual information-based feature selection, and chi-square feature selection. However, filter methods cannot adapt to changes in data or machine learning models. Once features are selected, they are fixed and cannot be updated or modified Tao et al. (2022). Additionally, filter methods have limited capacity to handle nonlinear relationships between the selected features and the target variable. In machine learning or deep learning models, wrapper methods are often used instead of filter

methods for feature selection. Using wrapper methods can improve the accuracy and interpretability of machine learning models, especially in cases where the interactions between features and the target variable are complex and nonlinear. Finally, embedded methods integrate feature selection into the model training process, enabling the model to learn which features are most relevant to the task. Embedded methods can accelerate model training and have lower computational costs compared to wrapper methods Rahman et al. (2023). As climate prediction becomes increasingly complex, feature selection methods may vary depending on the prediction model used. Particularly, climate exhibits nonlinear dynamics. Traditional feature selection methods in filter, wrapper, and embedded methods may be limited in climate prediction Villia et al. (2022). To address this issue, nonlinear filter methods such as decision trees, neural networks, and SVMs have been proposed, which can enhance the performance of machine learning models. Recently, researchers have focused on using deep learning methods for feature selection in climate prediction, especially when dealing with time series data. The advantage of deep learning methods lies in their ability to automatically extract relevant and complex features from time series data, with the potential for high accuracy and robustness. In deep learning models, CNNs are consistently used as information feature selection methods because they can extract features at multiple scales or resolutions. CNNs are well-suited for handling grid-like data, such as time series data, and can capture various patterns and trends in climate data. By combining CNNs with machine learning or deep learning models, highly accurate and efficient climate prediction models can be constructed.

2.3 Recent models for climate prediction

Single models (such as a single machine learning or deep learning model) can predict climate based on input features. The advantage of single models is that they are relatively simple and straightforward, and may require fewer computational resources. However, single models might not meet some critical needs in climate prediction. Additionally, they may struggle to capture the complexity and variability of climate data. When using deep learning models to handle the randomness and distribution imbalances of climate, single models might encounter issues of overfitting. For the above reasons, developing hybrid models has become a popular topic in the field of climate prediction research. Hybrid models combine the strengths of different approaches. For instance, statistical methods can be used to preprocess data, machine learning methods to select relevant features, and deep learning methods to make predictions. Nikseresht and Amindavar (2023) established a hybrid method based on ARFIMA and enhanced fractional Brownian motion for short-term climate prediction. The results showed that the hybrid method outperformed other models or methods, such as statistical models or individual neural networks. Kou et al. (2023) proposed a new combined prediction model for predicting precipitable water vapor (PWV). This model combines Variational Mode Decomposition (VMD) and the Multi-Objective Harris Hawks Optimization (MOHHO) algorithm, integrating four nonlinear models and two linear models. Through effective data preprocessing and weight optimization,
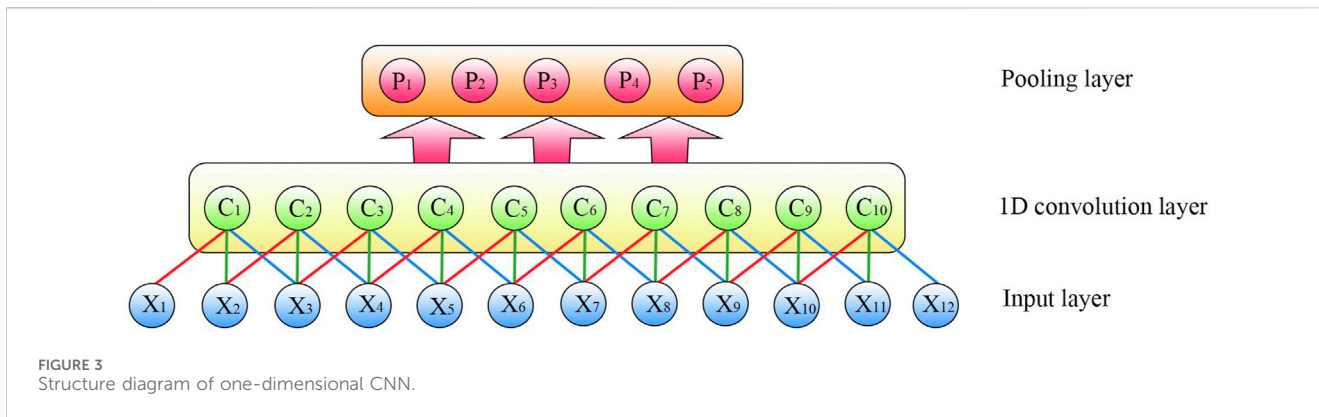
FIGURE 1
Overall algorithm flowchart.

FIGURE 2
Structure diagram of Transformer.
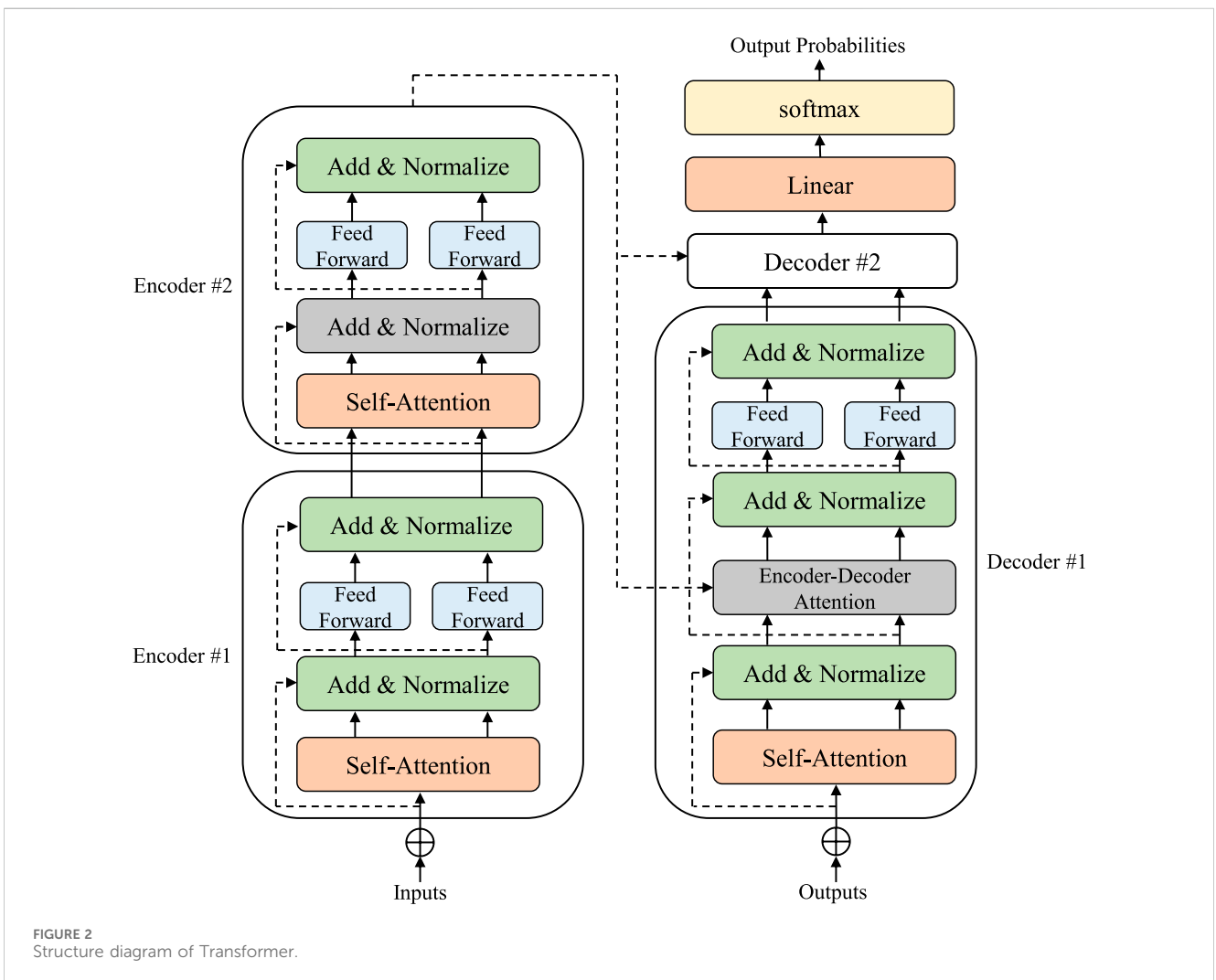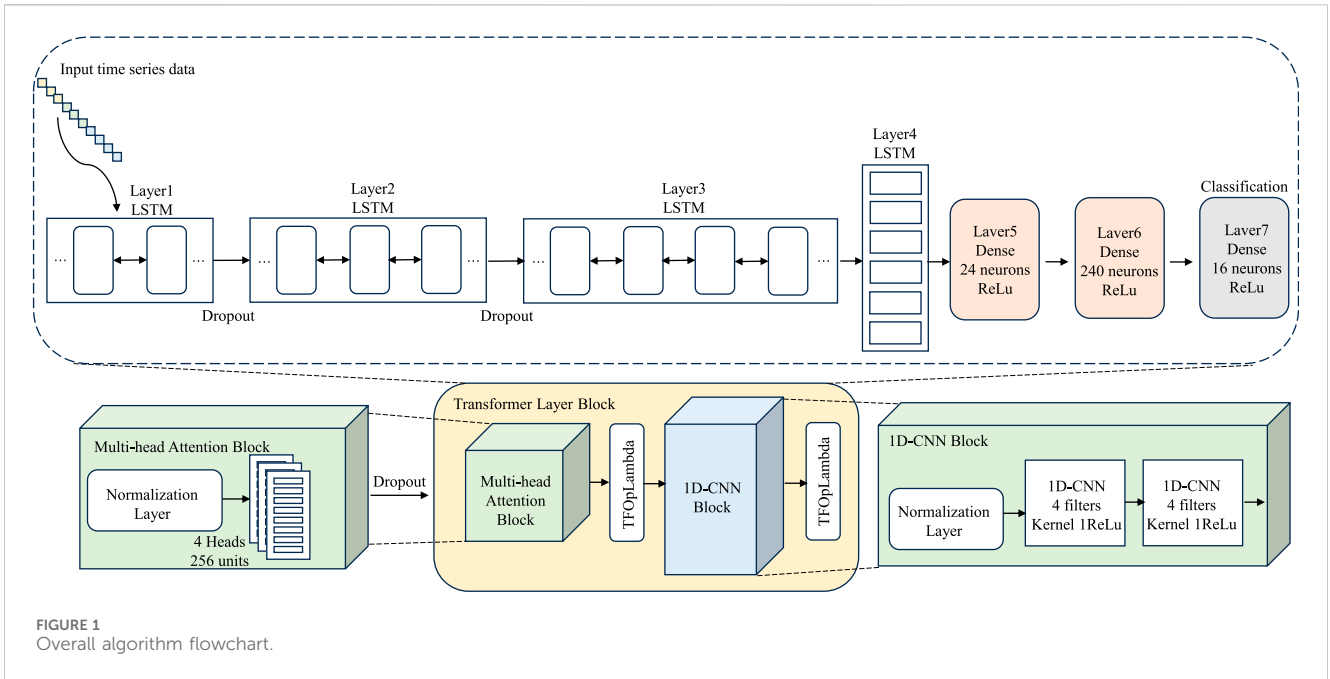
FIGURE 3
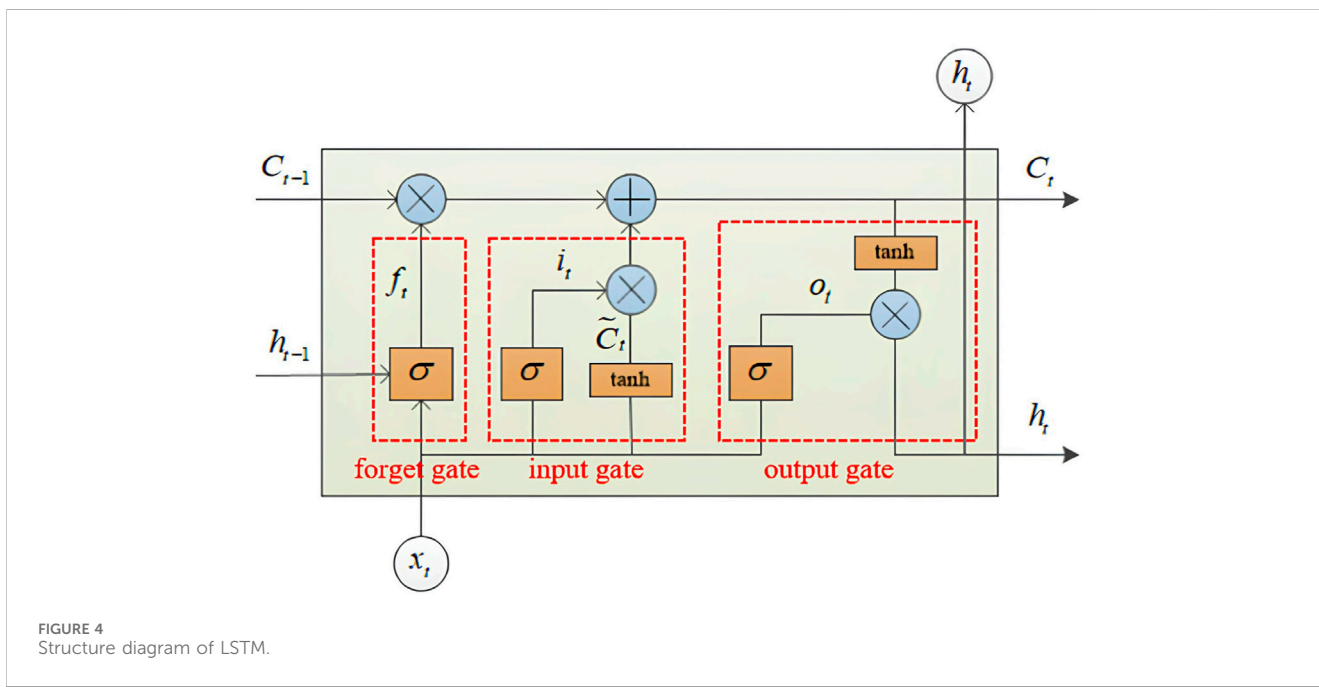Structure diagram of one-dimensional CNN.

FIGURE 4
Structure diagram of LSTM.

the model significantly improved the accuracy and stability of PWV predictions, providing technical support for selecting the timing of artificial rainfall operations. Vo et al. (2023) proposed a hybrid model based on Long Short-Term Memory networks and climate models (LSTM-CM) for drought prediction. The study demonstrated that LSTM-CM showed higher accuracy and lower uncertainty in drought prediction compared to the single LSTM model (LSTM-SA) and the climate prediction model GloSea5 (GS5). The hybrid model combined the low bias of LSTM-SA with the physical process simulation capabilities of GS5, effectively improving predictions for 1, 2, and 3 months. Mukhtar et al. (2024) proposed a GIS-based Multi-Criteria Decision Analysis (MCDA) method combined with large climate data records to assess flood risk in the Hunza-Nagar Valley in northern Pakistan. They developed a comprehensive flood risk map considering nine influencing factors, assigned weights to each factor using the Analytic Hierarchy Process (AHP), and integrated GIS with geospatial data to generate the flood risk map. This method performed excellently in flood risk modeling. Chin and Lloyd

(2024) proposed a hybrid model based on autoregressive Long Short-Term Memory networks (LSTM) for climate change prediction. The study used the ensemble mean version of the ERA5 dataset to develop a baseline machine learning model capable of predicting the overall long-term trends of Earth's climate and weather, accurately capturing seasonal patterns.

3 Methods

The overall flowchart of the algorithm in this article is shown in Figure 1, illustrates our proposed Transformer-CNN-LSTM hybrid model. This model integrates LSTM layers for capturing temporal dependencies, a Transformer layer block for leveraging multi-head attention to enhance feature extraction, and 1D-CNN blocks for refining the extracted features, ultimately leading to a dense classification layer that outputs the final prediction results.The proposed Transformer-CNN-LSTM hybrid model, designed for climate prediction, can be highly useful in several real-world

FIGURE 5
Model training and testing process.
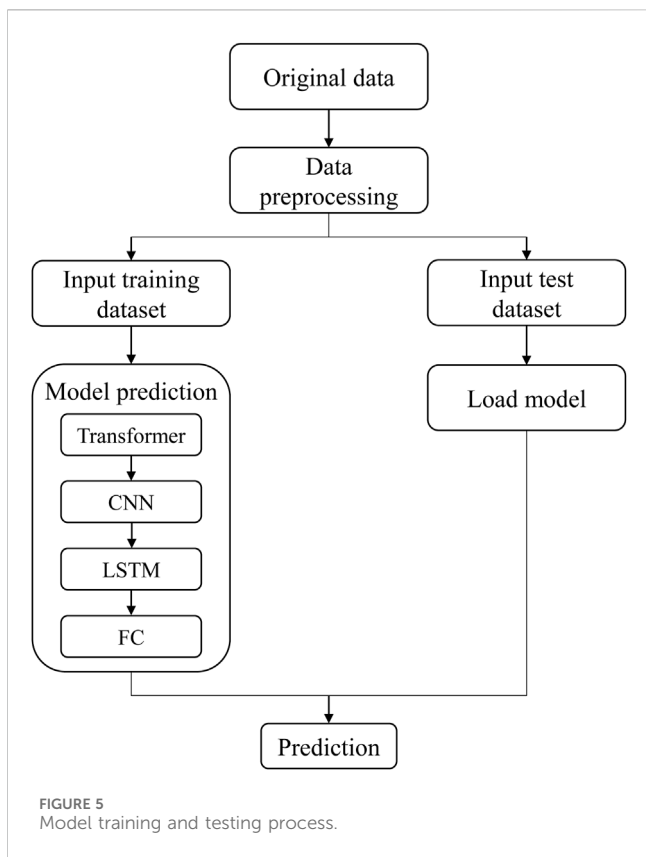
applications, particularly those where accurate climate forecasts are critical for decision-making: Energy Sector: Accurate climate predictions are essential for energy companies, particularly in the renewable energy sector. Wind and solar power generation are highly dependent on weather conditions such as wind speed, solar radiation, and temperature. By using our model, businesses can optimize energy production forecasts, reduce operational costs, and ensure better resource allocation. Agriculture: Farmers rely on precise weather forecasts to make decisions on crop planting, irrigation, and harvesting. Our model can improve agricultural

productivity by predicting extreme weather events, helping farmers minimize damage, reduce water usage, and optimize crop yields, which leads to better planning and reduced financial risks. Disaster Management and Insurance: Insurance companies and disaster management agencies can use our model to assess the risk of weather-related disasters such as floods, hurricanes, or droughts. Improved prediction accuracy helps these organizations set premiums more accurately, improve disaster preparedness, and provide timely warnings to reduce potential losses. Supply Chain and Logistics: Climate conditions can significantly affect supply chains, especially in sectors like transportation and retail. With accurate weather predictions, businesses can better plan routes, adjust inventory, and prevent disruptions due to extreme weather, leading to cost savings and efficiency improvements.

3.1 Transformer architecture

The Transformer architecture is mainly used to process sequence data tasks and relies on the attention mechanism to capture dependencies in the input sequence. Its algorithm architecture diagram is shown in Figure 2. The architecture consists of two parts: an encoder and a decoder. The encoder maps the input sequence to a potential representation space, while the decoder generates an output sequence from this potential representation. Each encoder and decoder is composed of multiple identical layers stacked together Alerskans et al. (2022).

Each encoder layer contains two main parts, a self-attention mechanism and a feedforward neural network. The self-attention mechanism captures global information by calculating the correlation between each position in the input sequence and other positions. Specifically, for the input sequence $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n]$, the self-attention mechanism calculates the attention weights between each pair of positions. First, the query, key, and value matrices are calculated Equation 1:

$$\mathbf{Q} = \mathbf{X}\mathbf{W}_Q, \quad \mathbf{K} = \mathbf{X}\mathbf{W}_K, \quad \mathbf{V} = \mathbf{X}\mathbf{W}_V \tag{1}$$

where $\mathbf{W}_Q$, $\mathbf{W}_K$, and $\mathbf{W}_V$ are trainable weight matrices.

FIGURE 6
Study site and location.
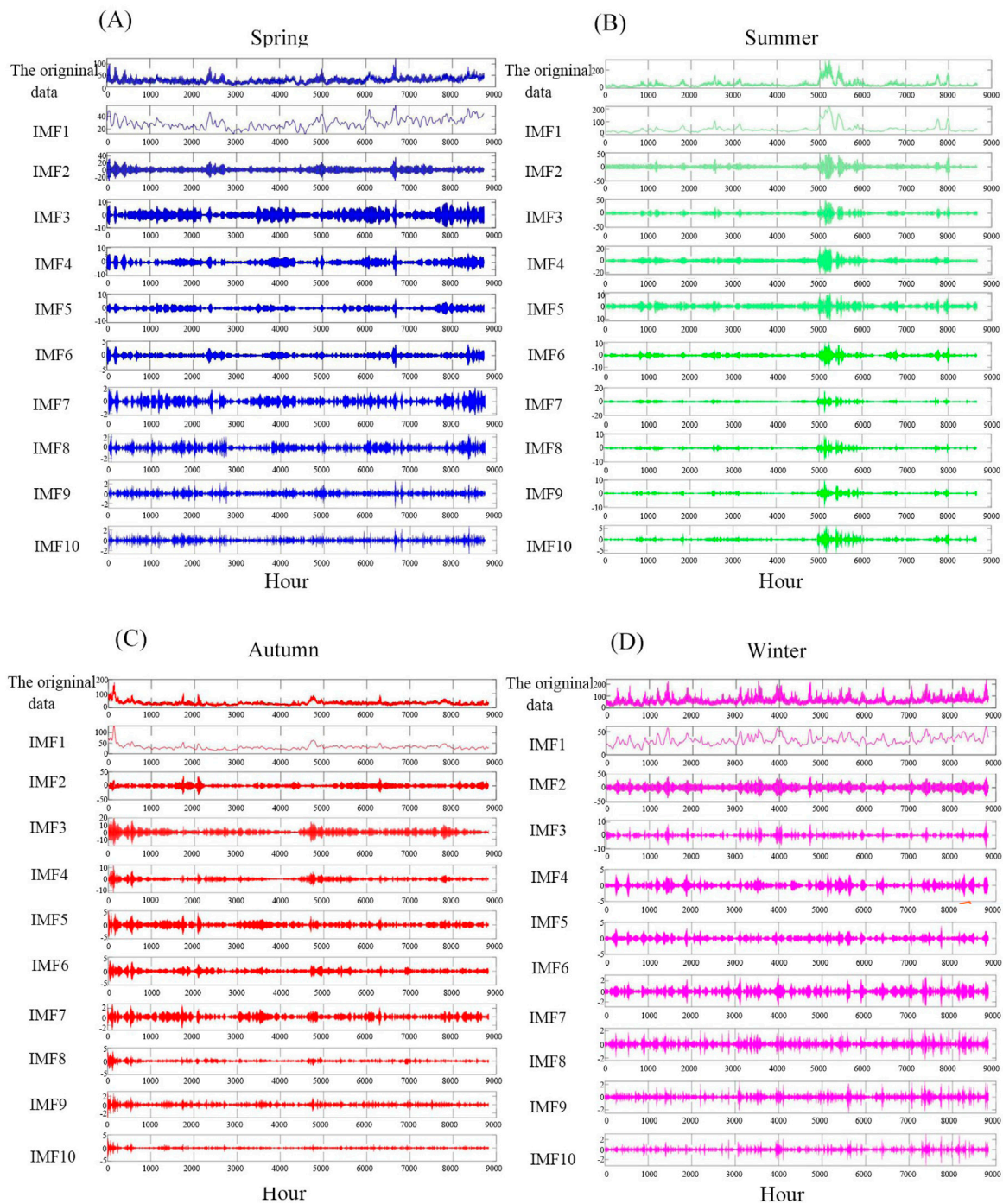
FIGURE 7
The original climate data is decomposed by transformer **(A–D)**.

Next, calculate the attention score Equation 2:

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_k}}\right)\mathbf{V} \tag{2}$$

Here, d_k is the dimension of the key vector, which is used to scale the attention score.

Each encoder layer also contains a feedforward neural network that processes the representation of each position independently Equation 3:

$$\text{FFN}(\mathbf{H}) = \text{ReLU}(\mathbf{H}\mathbf{W}_1 + \mathbf{b}_1)\mathbf{W}_2 + \mathbf{b}_2 \tag{3}$$

In addition to the encoder-like self-attention and feedforward neural network, the decoder layer also adds an attention mechanism that interacts with the encoder output. The encoder-decoder attention mechanism is used to interact the decoder's query with the encoder's keys and values, thereby utilizing the contextual information captured by the encoder Equation 4:

TABLE 1 The comparison forecast results of different hybrid models based on different decomposition methods (with random error range).

Dataset	Method	MAE	MAPE	RMSE	IA	TIC
Dataset 1 (spring)	CNN-LSTM	0.10213 ± 0.027	0.0606 ± 0.011	0.08367 ± 0.024	0.87 ± 0.029	0.1447 ± 0.018
	WT-CNN-LSTM	0.02096 ± 0.020	0.1196 ± 0.019	0.02747 ± 0.022	0.93 ± 0.021	0.0497 ± 0.012
	EMD-CNN-LSTM	0.02341 ± 0.025	0.1024 ± 0.014	0.01185 ± 0.011	0.96 ± 0.017	0.0071 ± 0.028
	VMD-CNN-LSTM	0.01855 ± 0.012	0.1041 ± 0.029	0.01581 ± 0.016	0.94 ± 0.015	0.0272 ± 0.024
	Transformer-CNN-LSTM	0.01061 ± 0.017	0.0977 ± 0.022	0.01842 ± 0.011	0.98 ± 0.013	0.0031 ± 0.026
Dataset 2 (summer)	CNN-LSTM	0.84763 ± 0.024	0.1426 ± 0.021	0.08591 ± 0.017	0.9 ± 0.013	0.1214 ± 0.016
	WT-CNN-LSTM	0.02418 ± 0.013	0.01000 ± 0.012	0.02747 ± 0.011	0.92 ± 0.022	0.0634 ± 0.018
	EMD-CNN-LSTM	0.02261 ± 0.011	0.1024 ± 0.013	0.01185 ± 0.024	0.95 ± 0.016	0.0087 ± 0.019
	VMD-CNN-LSTM	0.01844 ± 0.028	0.1167 ± 0.011	0.01185 ± 0.014	0.94 ± 0.012	0.1028 ± 0.022
	Transformer-CNN-LSTM	0.01724 ± 0.013	0.0950 ± 0.024	0.01138 ± 0.015	0.99 ± 0.019	0.0044 ± 0.017
Dataset 3 (autumn)	CNN-LSTM	0.47633 ± 0.016	0.1773 ± 0.018	0.06724 ± 0.014	0.89 ± 0.025	0.0912 ± 0.028
	WT-CNN-LSTM	0.01878 ± 0.011	0.1197 ± 0.022	0.02747 ± 0.013	0.93 ± 0.015	0.0545 ± 0.024
	EMD-CNN-LSTM	0.01861 ± 0.024	0.1084 ± 0.011	0.01185 ± 0.025	0.96 ± 0.012	0.0032 ± 0.014
	VMD-CNN-LSTM	0.01637 ± 0.017	0.1106 ± 0.019	0.05427 ± 0.021	0.95 ± 0.016	0.0234 ± 0.027
	Transformer-CNN-LSTM	0.01265 ± 0.012	0.0924 ± 0.026	0.01439 ± 0.018	0.99 ± 0.013	0.0026 ± 0.015
Dataset 4 (winter)	CNN-LSTM	0.03612 ± 0.022	0.1623 ± 0.015	0.05126 ± 0.021	0.9 ± 0.018	0.0845 ± 0.019
	WT-CNN-LSTM	0.01419 ± 0.011	0.1186 ± 0.017	0.02747 ± 0.013	0.93 ± 0.011	0.0639 ± 0.012
	EMD-CNN-LSTM	0.01367 ± 0.014	0.1024 ± 0.025	0.01185 ± 0.023	0.95 ± 0.012	0.0073 ± 0.011
	VMD-CNN-LSTM	0.01265 ± 0.017	0.1024 ± 0.014	0.01578 ± 0.024	0.95 ± 0.022	0.0423 ± 0.015
	Transformer-CNN-LSTM	0.00934 ± 0.012	0.0886 ± 0.013	0.02173 ± 0.014	0.98 ± 0.019	0.0058 ± 0.016

$$\text{Attention}\left(\mathbf{Q}_D, \mathbf{K}_E, \mathbf{V}_E\right) = \text{softmax}\left(\frac{\mathbf{Q}_D \mathbf{K}_E^T}{\sqrt{d_k}}\right)\mathbf{V}_E \qquad (4)$$

where $\mathbf{Q}_D$ is the query matrix of the decoder, $\mathbf{K}_E$ and $\mathbf{V}_E$ are the key and value matrices of the encoder, respectively.

The Transformer architecture uses self-attention mechanism and feed-forward neural network to effectively capture complex patterns and long-distance dependencies in sequence data, and has significant advantages in processing time series tasks such as climate prediction.

3.2 Convolutional neural network

CNN has demonstrated remarkable capabilities in computer vision tasks, particularly in image recognition and processing. CNNs are specifically designed to process pixel data, and they are equipped with three key strengths: a local sensing field, weight sharing, and down sampling Kareem et al. (2021). These three strengths can decrease the complexity of the network. The CNN model is composed of a convolutional layer, a pooling layer, and a fully connected layer. In particular, the convolutional layer is the fundamental component of the CNN model for automatic feature extraction. If a two-dimensional feature is provided as input, the convolutional layer can be represented as follows Equation 5:

$$a = \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{pmatrix} \qquad (5)$$

where $m \times n$ is the shape of input vector a. The input two-dimensional feature is convolved with filters W_k at the convolution layer Equation 6.

$$W_k = \begin{pmatrix} W_{11} \\ W_{21} \\ \ldots \\ W_{i1} \end{pmatrix} \qquad (6)$$

After the convolution, output map is formed and the feature map at the given layer is given by Equation 7

$$f(a) = f(W_k \times x + b_k) \qquad (7)$$

where W_k is the weight matrix and b_k is the bias value, k is the filter, which is the total number of filtering in convolutional layer.

Typically, a rectified linear (ReLU) function is used as the activation function after each convolutional layer, owing to its reliability and ability to accelerate convergence. The ReLU function is defined as follows Equation 8:
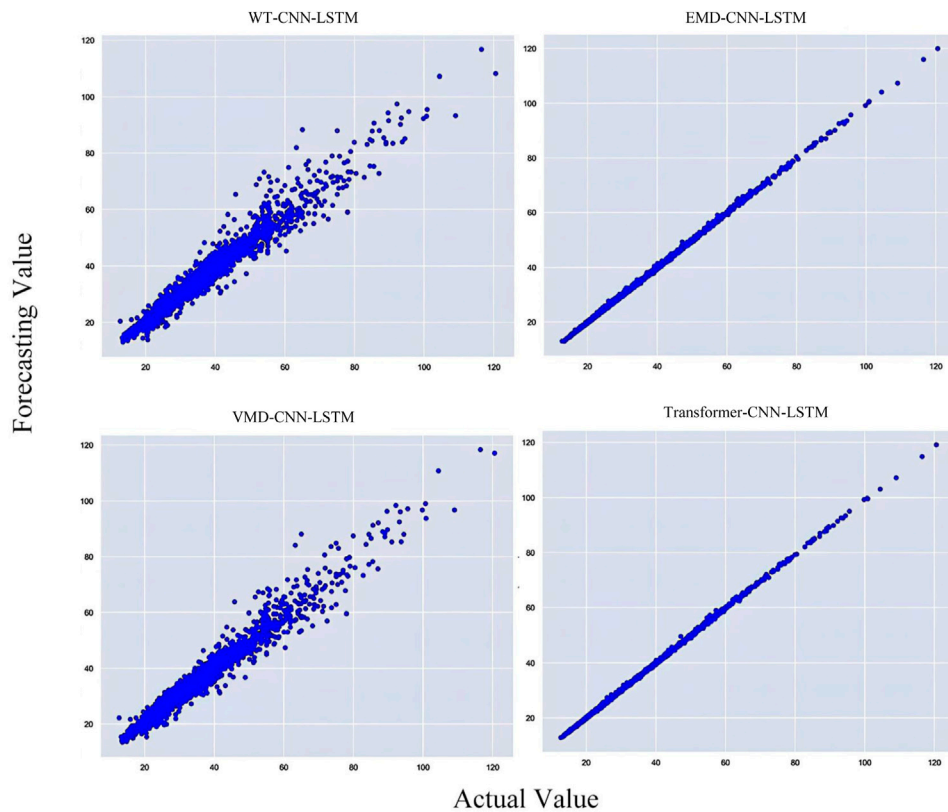
$$f(a) = \max(0, a) \qquad (8)$$

FIGURE 8
(Continued).

The pooling layer performs a down-sampling operation that reduces complexity, enhances efficiency, and mitigates the risk of overfitting. The pooling layer can be expressed as Equation 9

$$W_{\text{output}} = \frac{W_a - P}{S} + 1 \qquad (9)$$

where W_{output} is max pooling output, W_a is the function of the input volume, p is Pooling window size, S is stride.

After multiple convolutional and pooling layers, the output data is fully connected and usually flattened as the final output of the network. CNN uses the concept of weight sharing to provide better accuracy in highly nonlinear climate prediction problems. The one-dimensional convolution and pooling layer is shown in Figure 3.

3.3 Long Short-Term Memory

LSTM is a excellent variant of Recurrent Neural Network (RNN), which is designed to solve the vanishing gradient problem of RNN. The memory cell in the LSTM model is a crucial component that addresses the limitations of traditional RNN Liu et al. (2022). The memory cell is governed by three gates: a forget gate f_t, an input gate i_t, and an output gate o_t. These gates work together to regulate the flow of information through the memory cell. In the LSTM model, the forget gate is particularly important as it determines which information from the

memory cell should be retained or discarded. The forget gate controls the information flow through the cell state and is calculated using the equation below.

$$f_t = \sigma\left(W_i[h_{t-1}, x_t] + b_f\right) \qquad (10)$$

Here, σ denotes the sigmoid function, which maps the input values to the interval (0, 1). h_{t-1} is the previous hidden state. x_t is the current node. W_f represents the weight parameters. b_f represents the bias parameters. The weight and bias parameters are learned during the LSTM training process.

The Input gate in the LSTM model is responsible for incorporating new data into the memory cell by element-wise multiplication with the new input. It comprises the input activation gate and the candidate memory cell gate, which enable the LSTM to update and control the data in the memory. The Input gate determines which values of the cell state should be modified based on the input signal, and it is calculated using the equation below in the LSTM network. This modification is carried out using a sigmoid function and a Hyperbolic tangent (tanh) layer, as shown in Equation 10, resulting in the Equations 11–13

$$i_t = \sigma\left(W_i[h_{t-1}, x_t] + b_f\right) \qquad (11)$$

$$\tanh = \left(\frac{e^x - e^{-x}}{e^x + e^{-x}}\right) \qquad (12)$$

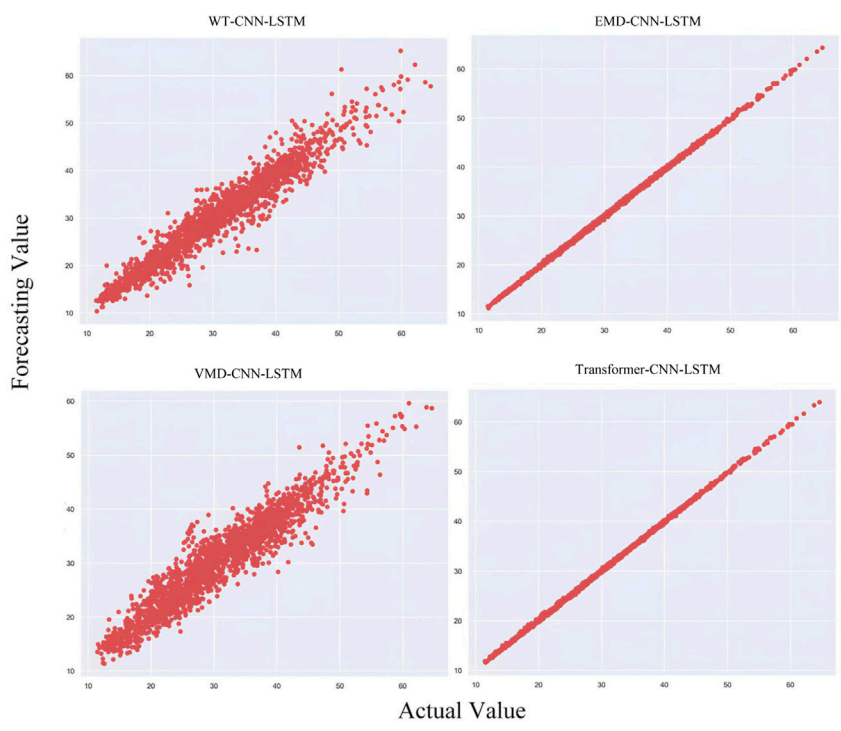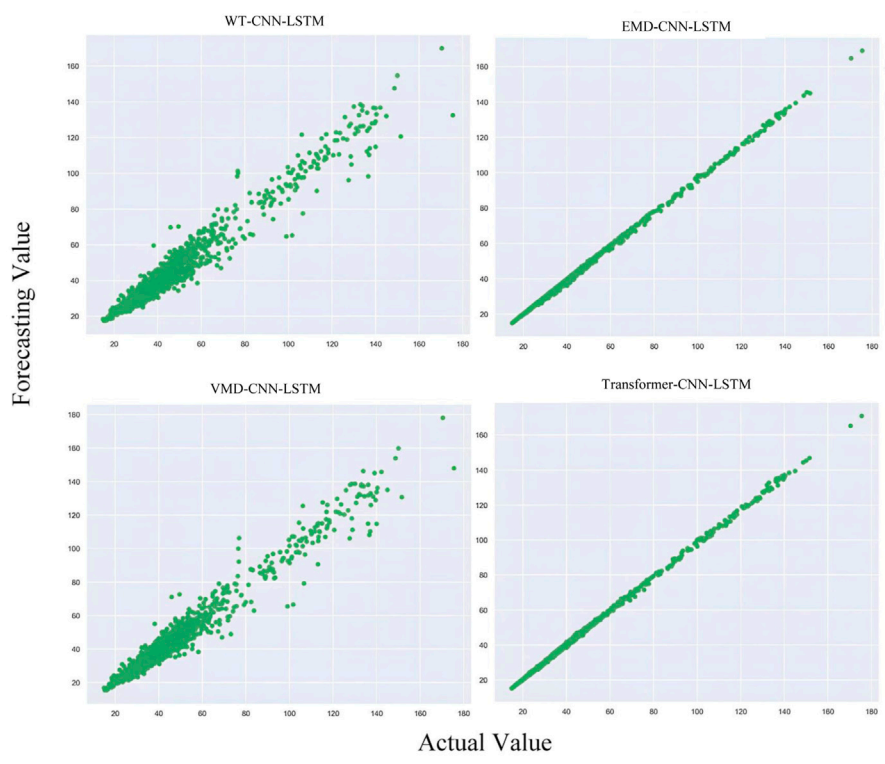$$\tilde{c}_t = \tanh\left(W_c[h_{t-1}, x_t] + b_c\right) \qquad (13)$$

FIGURE 8
(Continued).

To update the previous memory cell c_{t-1}, the input vector and the candidate memory cell vector are combined, which is represented as Equation 14:

$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t \qquad (14)$$

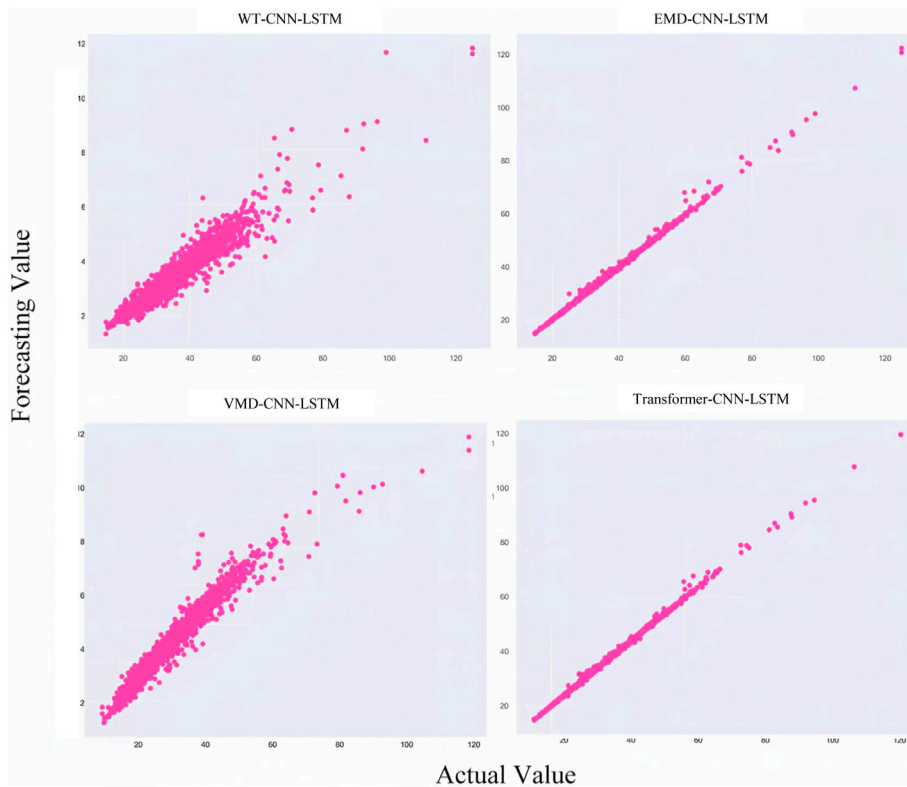where $\odot$ denotes element-wise multiplication.

FIGURE 8
(Continued). Scatter Plot of Actual vs. Forecasted Values for the different models.

The output gate in the given system is utilized to merge with the existing memory cell, resulting in the production of the present hidden state. This current hidden state subsequently affects the output in the subsequent time step. The computation process can be illustrated as follows Equation 15:

$$o_t = \sigma \left(W_o \left[h_{t-1}, x_t, c_t \right] + b_o \right) \tag{15}$$

The structure of LSTM is shown in Figure 4. The LSTM method addresses the problem of long-term dependence in learning, which provides good forecasting results Tzoumpas et al. (2024).

3.4 Model training and testing

Figure 5 illustrates the training and testing process of the model proposed in this paper. First, the processed data is split into training and testing datasets, with 70% and 30% proportions, respectively.

Subsequently, the training dataset is fed into the hybrid model, which consists of different components such as Transformer, CNN, and LSTM. The Transformer component helps decompose the time series data into meaningful components, the CNN extracts spatial features, and the LSTM captures temporal dependencies. Finally, the hybrid model is trained using the training dataset. During this process, once the model reaches the predetermined number of iterations, a predictive model is obtained. In the testing phase, the trained model is loaded, and the testing dataset is input for prediction, resulting in experimental outcomes.

This process involves data preprocessing, where the raw data is cleaned, transformed, and split into training and testing sets. The training dataset comprises the input features for model training, while the testing dataset is used to evaluate model performance. The hybrid model is trained on the training data to predict climate. The model combines the Transformer for time series decomposition, the CNN for capturing spatial or temporal patterns, the LSTM for handling sequential data, and fully connected layers for complex feature interactions. After training is complete, the model is loaded for prediction, generating climate forecasts based on the input features from the testing dataset.

3.5 Prediction results

In the experiment, five different metrics are used to evaluate the predictive performance of climate prediction models. These metrics are Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Root Mean Square Error (RMSE), Index of Agreement (IA), and Theil's Inequality Coefficient (TIC). Specifically, MAE and MAPE are used to measure the accuracy of the model by calculating the percentage difference between the predicted values and the actual values. The formulas for these calculations are as follows Equations 16, 17:

$$MAE = \frac{1}{N} \sum_{i=1}^{N} |F_i - A_i| \tag{16}$$

4 Experiment results

The hybrid deep learning model for climate prediction was validated using data from the Guangdong Meteorological Center, and all experiments were implemented on a computer using MATLAB 2018a.

4.1 Data collection

The climate temperature data for each season (spring, summer, autumn, and winter) were collected from the Guangdong Meteorological Center in China (see Figure 6). This study utilized hourly climate temperature data for the entire year of 2021. Specifically, Dataset A covers 30 days from 1 April 2021, to 30 April 2021 (spring); Dataset B covers 30 days from 1 July 2021, to 30 July 2021 (summer); Dataset C covers 30 days from 1 October 2021, to 30 October 2021 (autumn); and Dataset D covers 30 days from 1 January 2021, to 30 January 2021 (winter). Due to the impact of extreme weather in summer, the range of climate fluctuations is particularly large. To explain seasonal differences, the annual climate data were divided into four seasons. This division can improve the prediction accuracy when using the model.

4.2 Data preprocessing

The proposed Transformer technique was used to decompose the raw climate data to effectively generate decomposition results. Transformer is a data analysis technique used to decompose time series into its Intrinsic Mode Functions (IMFs). Each IMF captures specific oscillatory or trend behaviors within the time series. IMFs are ordered based on their energy content, with the IMF containing the highest energy considered the most significant component. By summing the selected IMFs, the original time series can be reconstructed. As shown in Figure 7, when using the

TABLE 2 Comparison of the Random Forest model with STL decomposition and deep learning-based hybrid models using different decomposition methods.

Model	MAE	RMSE	MAPE	TIC	IA
Random Forest (STL)	0.0421	0.0912	0.0678	0.0324	0.91
CNN-LSTM	0.0184	0.0487	0.0389	0.0147	0.95
WT-CNN-LSTM	0.0209	0.0503	0.0412	0.0152	0.94
EMD-CNN-LSTM	0.0181	0.0465	0.0375	0.0139	0.96
VMD-CNN-LSTM	0.0175	0.0451	0.0368	0.0136	0.96
Transformer-CNN-LSTM	**0.0106**	**0.0315**	**0.0267**	**0.0091**	**0.98**

$$\text{MAPE} = \frac{1}{N} \sum_{i=1}^{N} \left| \frac{A_i - F_i}{A_i} \right| \times 100\% \qquad (17)$$

RMSE is used to evaluate the effectiveness of the proposed method and conventional strategy. The equation of RMSE is as follows Equation 18:

$$\text{RMSE} = \sqrt{\frac{1}{N} \times \sum_{i=1}^{N} (F_i - A_i)^2} \qquad (18)$$

IA is employed to evaluate the model's forecasting ability. The IA equation is Equation 19

$$\text{IA} = 1 - \sum_{i=1}^{N} (F_i - A_i)^2 \Big/ \sum_{i=1}^{N} \left(\left| F_i - \bar{A} \right| + \left| A_i + \bar{A} \right| \right)^2 \qquad (19)$$

TIC is used to assess the generalization ability of the model. The TIC is defined as in equation Equation 20:

$$\text{TIC} = \sqrt{\frac{1}{N} \times \sum_{i=1}^{N} (F_i - A_i)^2} \Big/ \left(\sqrt{\frac{1}{N} \times \sum_{i=1}^{N} A_i^2} + \sqrt{\frac{1}{N} \times \sum_{i=1}^{N} F_i^2} \right)$$
$$(20)$$

TABLE 3 The comparison forecast results of different hybrid models for four datasets.

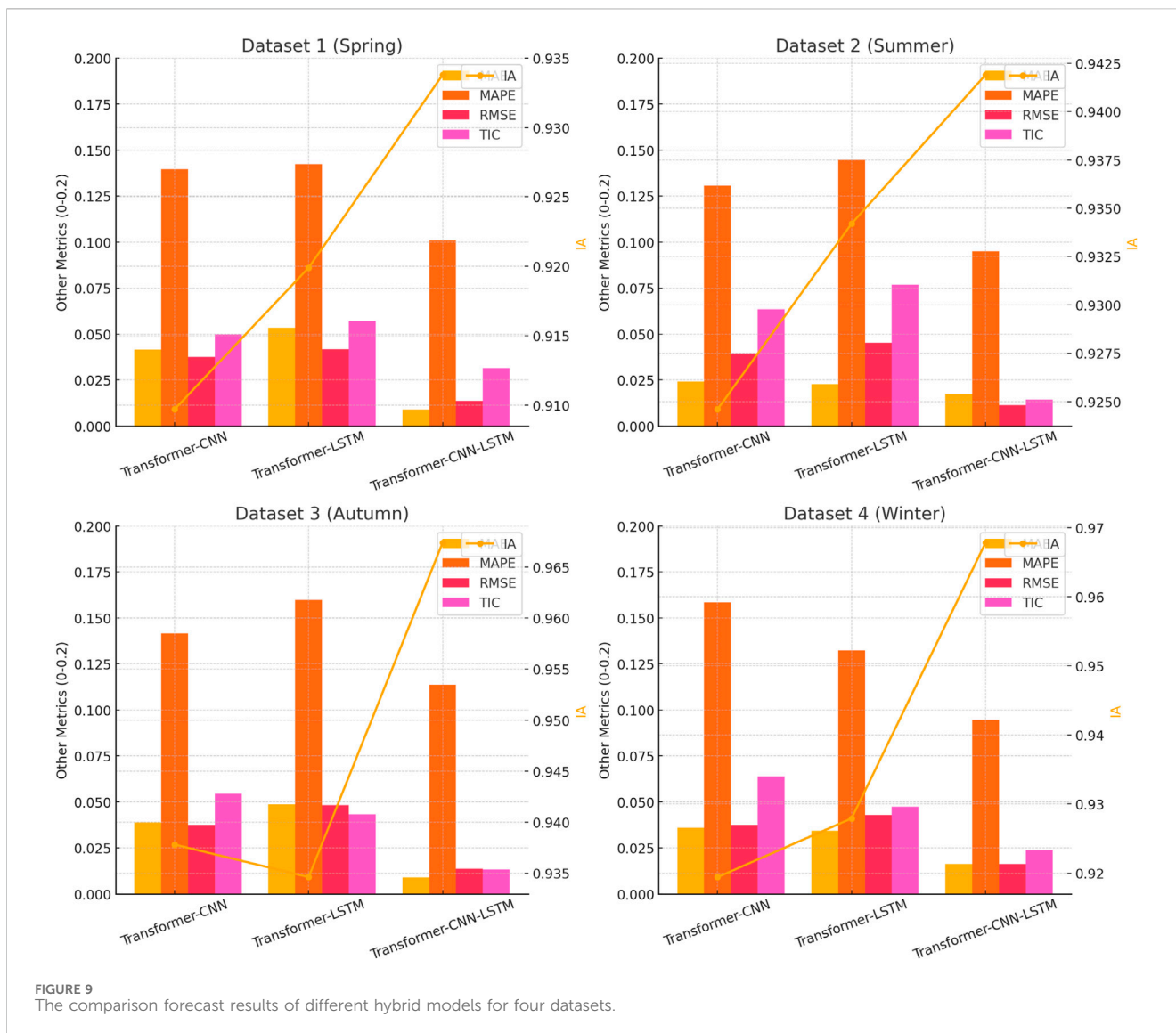| Dataset	Method	MAE	MAPE	RMSE	IA	TIC
Dataset 1 (spring)	Transformer-CNN	0.04158	0.1396	0.03747	0.9097	0.0497
	Transformer-LSTM	0.05341	0.1,424	0.04185	0.9199	0.0571
	Transformer-CNN-LSTM	0.00898	0.1009	0.01359	0.9338	0.0316
Dataset 2 (summer)	Transformer-CNN	0.02418	0.1307	0.03947	0.9246	0.0634
	Transformer-LSTM	0.02261	0.1445	0.04532	0.9342	0.0767
	Transformer-CNN-LSTM	0.01724	0.0950	0.01138	0.9419	0.0144
Dataset 3 (autumn)	Transformer-CNN	0.03878	0.1416	0.03747	0.9378	0.0545
	Transformer-LSTM	0.04861	0.1598	0.04826	0.9346	0.0432
	Transformer-CNN-LSTM	0.00898	0.1137	0.01376	0.9674	0.0132
Dataset 4 (winter)	Transformer-CNN	0.03603	0.1586	0.03747	10.9194	0.0639
	Transformer-LSTM	0.03422	0.1324	0.04285	0.9279	0.0473
	Transformer-CNN-LSTM	0.01637	0.0946	0.01637	0.9678	0.0238

FIGURE 9
The comparison forecast results of different hybrid models for four datasets.

Transformer method to decompose the hourly climate temperature for the four seasons of 2021, 10 IMFs with different central frequencies were obtained. The Transformer can effectively capture the fundamental characteristics of the raw data and significantly contribute to the overall success of the prediction process. In our experiments, the dataset was split into 80% for training and 20% for testing. We employed a random sampling approach rather than a time-based split. This decision was made with the goal of ensuring the model's ability to generalize across different time periods. While climate data inherently possess temporal dependencies, we chose random sampling to allow the model to learn patterns across various points in time, enhancing its generalization capabilities. This approach ensures that both the training and testing datasets contain data points spread across different time periods, reducing the risk of overfitting to a specific time range. Moreover, random splitting helps mitigate any potential temporal bias and prepares the model to better adapt to unseen future data. In our current model setup, we are forecasting one time step ahead at a time, meaning that the model predicts a single value for each future time step based on the input

data. However, the model is flexible and can be adapted to output a vector of future values rather than just a single value. The architecture of models like CNN-LSTM and Transformer-CNN-LSTM inherently supports multi-step forecasting, where instead of predicting just the next time step, the model can be trained to output multiple future values simultaneously. This is referred to as multi-step forecasting or sequence-to-sequence prediction, and it can be achieved by adjusting the output layer to produce a vector of values, as well as modifying the training process accordingly. So, while the current experiments focus on single-step forecasting, the model can be extended to forecast multiple steps ahead (vector output) with some adjustments to the architecture and training approach.

4.3 Scenario I

In Scenario 1, the prediction results of CNN-LSTM, WT-CNN-LSTM, EMD-CNN-LSTM, VMD-CNN-LSTM, and Transformer-CNN-LSTM were compared to study the impact of different decomposition methods on the climate prediction of hybrid

TABLE 4 The comparison forecast results of the proposed hybrid model and other popular models for four datasets.

Dataset	Method	MAE	MAPE	RMSE	IA	TIC
Dataset 1 (spring)	LSTM	0.01844	0.1186	0.02096	0.9531	0.0497
	BiLSTM	0.01724	0.140,970	0.02341	0.9637	0.0571
	Random Forest	0.02341	0.1584	0.01855	0.9755	0.0273
	Decision Trees	0.03855	0.1606	0.01523	0.9507	0.0116
	RNN	0.02637	0.1383	0.01747	0.9743	0.0098
	ANN	0.02265	0.1034	0.01123	0.9858	0.0044
	Ours	0.01061	0.0977	0.00549	0.9912	0.0036
Dataset 2 (summer)	LSTM	0.02418	0.1784	0.01778	0.9346	0.0545
	BiLSTM	0.02261	0.1497	0.01412	0.9554	0.0639
	Random Forest	0.01899	0.1584	0.01591	0.9246	0.0473
	Decision Trees	0.01624	0.1606	0.01285	0.9642	0.0723
	RNN	0.01185	0.1167	0.01138	0.9656	0.0138
	ANN	0.01427	0.0950	0.00939	0.9706	0.0096
	Ours	0.00536	0.0897	0.00732	0.9996	0.0044
Dataset 3 (autumn)	LSTM	0.02747	0.1567	0.01466	0.9279	0.0245
	BiLSTM	0.01185	0.1396	0.02627	0.9065	0.0182
	Random Forest	0.02427	0.1424	0.01276	0.9378	0.0178
	Decision Trees	0.01439	0.2141	0.01429	0.9346	0.0145
	RNN	0.02747	0.0924	0.00778	0.9054	0.0081
	ANN	0.01185	0.1106	0.01172	0.9674	0.0084
	Ours	0.00434	0.0624	0.00549	0.9956	0.0026
Dataset 4 (winter)	LSTM	0.02747	0.1197	0.01919	0.9279	0.0145
	BiLSTM	0.01419	0.1416	0.02167	0.9065	0.0234
	Random Forest	0.01367	0.1598	0.01165	0.9194	0.0139
	Decision Trees	0.02265	0.2020	0.01178	0.9279	0.0073
	RNN	0.00934	0.0886	0.00973	0.9065	0.0123
	ANN	0.01578	0.0924	0.00785	0.9678	0.0158
	Ours	0.00128	0.0663	0.00277	0.9965	0.0058

models. As shown in Table 1, introducing decomposition methods such as WT, EMD, VMD, and Transformer into the hybrid model can improve the performance of CNN-LSTM-based prediction results. According to the results of MAE, RMSE, MAPE, and TIC, the proposed Transformer-CNN-LSTM hybrid model outperforms the aforementioned hybrid models, and the IA results of the proposed hybrid model are also superior. Figure 8 shows the correlation between the predicted values of CNN-LSTM models with different decomposition methods and the actual values for each season. It was found that the Transformer-CNN-LSTM model is almost a linear function, indicating that the proposed hybrid model is more effective than other decomposition methods. Therefore, we conclude that the proposed Transformer method can be used to effectively decompose raw climate data. The results

presented in Table 2 demonstrate a clear performance advantage of the deep learning-based hybrid models over the Random Forest model, which used STL decomposition. While the Random Forest model performed reasonably well, achieving a MAE of 0.0421 and an IA of 0.91, the deep learning models, particularly the Transformer-CNN-LSTM hybrid, consistently outperformed it across all metrics. The Transformer-CNN-LSTM model achieved the lowest MAE (0.0106), RMSE (0.0315), and TIC (0.0091), indicating that it is not only more accurate but also more stable in its predictions. Moreover, the IA (0.98) of the Transformer-CNN-LSTM model is notably higher, showcasing its superior ability to fit the data compared to the Random Forest model and other hybrid models. The other deep learning models, such as VMD-CNN-LSTM and EMD-CNN-LSTM, also performed better than the Random Forest
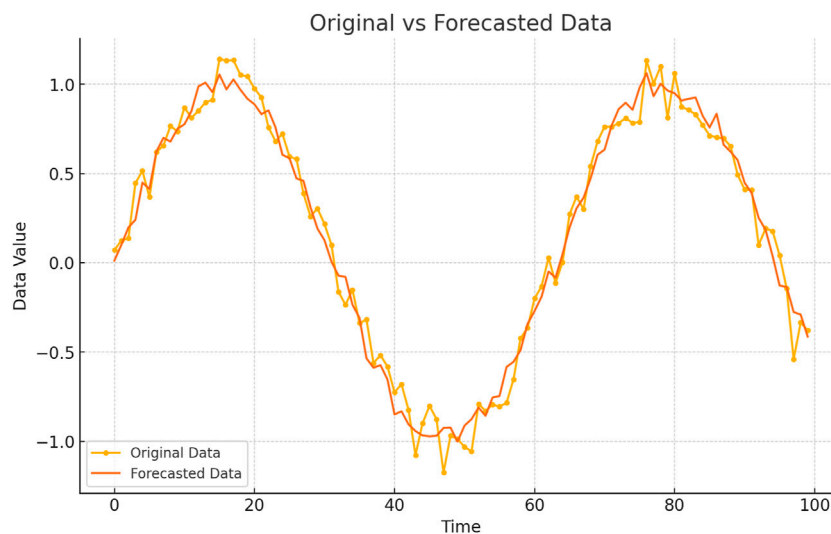
FIGURE 10
Comparison of original and predicted values.

model, though their results were slightly less impressive than the Transformer-based model. These findings highlight the effectiveness of decomposition methods, especially when integrated with advanced deep learning architectures like the Transformer, in improving the accuracy and generalization of climate predictions. This comparative analysis confirms the superiority of the Transformer-CNN-LSTM hybrid model in handling complex climate data, validating its use in scenarios where precise and reliable predictions are required.

4.4 Scenario II

From the results of Scenario 1, the Transformer method outperforms other decomposition methods. In Scenario 2, the Transformer method was combined with different comparison prediction algorithms to compare the climate prediction results of Transformer-CNN, Transformer-LSTM, and Transformer-CNN-LSTM. The comparative prediction results of the aforementioned hybrid models for the four datasets are shown in Table 3 and Figure 9. Table 3 indicates that the Transformer-CNN-LSTM hybrid climate prediction model outperforms other hybrid models across five evaluation metrics, demonstrating more accurate prediction effects and more stable fitting results.

4.5 Scenario III

To compare with different popular climate prediction methods, Scenario IV was designed, including BiLSTM, LSTM, random forest, decision tree, ANN, RNN, and our hybrid method. Table 4 shows the comparative prediction results of the aforementioned models for the four datasets. As seen in Table 4, the Transformer-CNN-LSTM hybrid climate prediction

model outperforms other models in the five evaluation metrics. Specifically, the MAE values of the prediction results using the proposed hybrid model for the four datasets are 0.01061, 0.00536, 0.00434, and 0.00128, respectively. The MAPE values for the four datasets are 0.0977, 0.0897, 0.0624, and 0.0663, respectively. The RMSE values for the four datasets are 0.00549, 0.00732, 0.00549, and 0.00277, respectively. Lower MAE, MAPE, and RMSE values indicate higher accuracy and that the predicted values are closer to the actual values. Among these models, the Transformer-CNN-LSTM has the lowest MAE, MAPE, and RMSE values. The IA values of the prediction results using the proposed hybrid model for the four datasets are the highest among all models, at 0.9912, 0.9996, 0.9956, and 0.9965, respectively. The TIC values for the four datasets using the proposed model are 0.00549, 0.00732, 0.00549, and 0.00277, respectively. A TIC value of 0 indicates complete equality, meaning there is no inequality between groups or time periods. According to these five evaluation metrics, the proposed model demonstrates more accurate prediction effects and more stable fitting results.For the climate prediction task, the primary features included Temperature (as the target variable), Humidity, Wind Speed, Precipitation, and Atmospheric Pressure. To capture temporal dependencies in the data, we incorporated a range of lagged observations from 1 to 12 h, with intervals of 1 h between each lag. This approach enabled the models to account for both short-term and longer-term dependencies. In the proposed hybrid model (Transformer-CNN-LSTM), these lagged values were used to capture both local and global temporal patterns. For comparison models like Random Forest, LSTM, and CNN, the same set of lagged values was applied to ensure consistency across experiments. This uniform feature set allowed for a direct comparison of model performance. Figure10 is an example of the original and forecasted data plotted using a line plot. The plot shows patterns in both the original data and the forecasted values, helping to visualize their similarities and differences.

This allows readers to better understand the model's predictive performance.

5 Disucssion and conclusion

This study developed a Transformer-CNN-LSTM hybrid prediction model to forecast future climate temperatures. There are five steps to selecting the optimal model with the best predictive performance. First, climate data from Guangdong, China, was collected and organized into datasets. Then, the Transformer method was used to decompose the raw feature time series matrix into a set of IMF components with different frequencies. Third, the proposed CNN-LSTM hybrid model was used to independently predict the multidimensional feature matrix. Fourth, the performance of the proposed model was evaluated using the MAE, MAPE, RMSE, IA, and TIC metrics. Finally, the proposed hybrid model was compared with popular climate prediction models using the evaluation metrics. The experimental results indicate that the Transformer decomposition method in the proposed hybrid framework performed best in Scenario I. Our results also show that the Transformer-CNN-LSTM hybrid climate prediction model outperforms other hybrid models in the five evaluation metrics, demonstrating more accurate prediction effects and more stable fitting results. Despite the numerous advantages of the Transformer-CNN-LSTM hybrid model, there are some limitations to this study. The first limitation is the use of only one set of climate data from Guangdong, China. To verify the generalizability of our hybrid deep learning model, future testing on different climate datasets and expansion of our dataset is necessary. Finally, several other influencing factors may significantly impact climate temperature in this study, such as greenhouse gas emissions, land use changes, ocean temperature and current changes, variations in solar radiation, and human activities. In the future, we intend to address these limitations and rectify these shortcomings.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

References

Author contributions

KL: Conceptualization, Formal Analysis, Investigation, Methodology, Project administration, Resources, Supervision, Visualization, Writing–original draft, Writing–review and editing. KL: Conceptualization, Formal Analysis, Investigation, Methodology, Project administration, Resources, Supervision, Visualization, Writing–review and editing. ZW: Conceptualization, Formal Analysis, Methodology, Project administration, Resources, Writing–review and editing. YL: Conceptualization, Formal Analysis, Investigation, Methodology, Project administration, Resources, Supervision, Visualization, Writing–review and editing. BB: Conceptualization, Investigation, Methodology, Project administration, Resources, Writing–original draft. RZ: Formal Analysis, Investigation, Methodology, Resources, Writing–original draft, Writing–review and editing.

Conflict of interest

Authors SL, ZW, YL, BB, and RZ were employed by State Grid Ningxia Electric Power Co. Ltd., Eco-tech Research Institute. Author KL was employed by Shandong Chengxin Engineering Construction & Consulting Co. Ltd.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Akpuokwe, C. U., Adeniyi, A. O., Bakare, S. S., and Eneh, N. E. (2024). Legislative responses to climate change: a global review of policies and their effectiveness. *Int. J. Appl. Res. Soc. Sci.* 6, 225–239. doi:10.51594/ijarss.v6i3.852

Alerskans, E., Nyborg, J., Birk, M., and Kaas, E. (2022). A transformer neural network for predicting near-surface temperature. *Meteorol. Appl.* 29, e2098. doi:10.1002/met.2098

Amjad, M., Khan, A., Fatima, K., Ajaz, O., Ali, S., and Main, K. (2022). Analysis of temperature variability, trends and prediction in the karachi region of Pakistan using arima models. *Atmosphere* 14, 88. doi:10.3390/atmos14010088

Bauer, P., Dueben, P., Chantry, M., Doblas-Reyes, F., Hoefler, T., McGovern, A., et al. (2023). Deep learning and a changing economy in weather and climate prediction. *Nat. Rev. Earth and Environ.* 4, 507–509. doi:10.1038/s43017-023-00468-z

Broszkiewicz-Suwaj, E., and Wyłomańska, A. (2021). Application of non-Gaussian multidimensional autoregressive model for climate data prediction. *Int. J. Adv. Eng. Sci. Appl. Math.* 13, 236–247. doi:10.1007/s12572-021-00300-1

Butt, A. Q., Shangguan, D., Ding, Y., Banerjee, A., Sajjad, W., and Mukhtar, M. A. (2024). Assessing the existing guidelines of environmental impact assessment and mitigation measures for future hydropower projects in Pakistan. *Front. Environ. Sci.* 11, 1342953. doi:10.3389/fenvs.2023.1342953

Chen, R., Wang, X., Zhang, W., Zhu, X., Li, A., and Yang, C. (2019). A hybrid cnn-lstm model for typhoon formation forecasting. *GeoInformatica* 23, 375–396. doi:10.1007/s10707-019-00355-0

Chin, S., and Lloyd, V. (2024). Predicting climate change using an autoregressive long short-term memory model. *Front. Environ. Sci.* 12, 1301343. doi:10.3389/fenvs.2024.1301343

Fahad, S., Su, F., Khan, S. U., Naeem, M. R., and Wei, K. (2023). Implementing a novel deep learning technique for rainfall forecasting via climatic variables: an approach via hierarchical clustering analysis. *Sci. Total Environ.* 854, 158760. doi:10.1016/j.scitotenv.2022.158760

Han, Y., Sun, K., Yan, J., and Dong, C. (2023). The cnn-gru model with frequency analysis module for sea surface temperature prediction. *Soft Comput.* 27, 8711–8720. doi:10.1007/s00500-023-08172-2

Kareem, S., Hamad, Z. J., and Askar, S. (2021). An evaluation of cnn and ann in prediction weather forecasting: a review. *Sustain. Eng. Innovation* 3, 148–159. doi:10.37868/sei.v3i2.id146

Kou, M., Zhang, K., Zhang, W., Ma, J., Ren, J., and Wang, G. (2023). Application research of combined model based on vmd and mohho in precipitable water vapor prediction. *Atmos. Res.* 292, 106841. doi:10.1016/j.atmosres.2023.106841

Liu, X., Wang, S., Lu, S., Yin, Z., Li, X., Yin, L., et al. (2023). Adapting feature selection algorithms for the classification of Chinese texts. *Systems* 11, 483. doi:10.3390/systems11090483

Liu, Y., Li, D., Wan, S., Wang, F., Dou, W., Xu, X., et al. (2022). A long short-term memory-based model for greenhouse climate prediction. *Int. J. Intelligent Syst.* 37, 135–151. doi:10.1002/int.22620

Mukhtar, M. A., Shangguan, D., Ding, Y., Anjum, M. N., Banerjee, A., Butt, A. Q., et al. (2024). Integrated flood risk assessment in hunza-nagar, Pakistan: unifying big climate data analytics and multi-criteria decision-making with gis. *Front. Environ. Sci.* 12, 1337081. doi:10.3389/fenvs.2024.1337081

Niksersht, A., and Amindavar, H. (2023). Hourly solar irradiance forecasting based on statistical methods and a stochastic modeling approach for residual error compensation. *Stoch. Environ. Res. Risk Assess.* 37, 4857–4892. doi:10.1007/s00477-023-02539-5

Pande, C. B., Kushwaha, N., Orimoloye, I. R., Kumar, R., Abdo, H. G., Tolche, A. D., et al. (2023). Comparative assessment of improved svm method under different kernel functions for predicting multi-scale drought index. *Water Resour. Manag.* 37, 1367–1399. doi:10.1007/s11269-023-03440-0

Pandey, P., Tripura, H., and Pandey, V. (2019). Improving prediction accuracy of rainfall time series by hybrid sarima–garch modeling. *Nat. Resour. Res.* 28, 1125–1138. doi:10.1007/s11053-018-9442-z

Peng, W., and Ni, Q. (2020). "A hybrid svm-lstm temperature prediction model based on empirical mode decomposition and residual prediction," in 2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Toronto, ON, Canada, 11-14 October 2020 (IEEE), 1616–1621.

Rahman, S., Olausson, M., Vitucci, C., and Avgouleas, I. (2023). "Ambient temperature prediction for embedded systems using machine learning," in *International conference on engineering of computer-based systems* (Springer), 12–25.

Ruigar, H., and Golian, S. (2016). Prediction of precipitation in golestan dam watershed using climate signals. *Theor. Appl. Climatol.* 123, 671–682. doi:10.1007/s00704-015-1377-2

Sang, J., Hou, B., Wang, H., and Ding, X. (2023). Prediction of water resources change trend in the three gorges reservoir area under future climate change. *J. Hydrology* 617, 128881. doi:10.1016/j.jhydrol.2022.128881

Sarhadi, A., Kelly, R., and Modarres, R. (2014). Snow water equivalent time-series forecasting in ontario, Canada, in link to large atmospheric circulations. *Hydrol. Process.* 28, 4640–4653. doi:10.1002/hyp.10184

Sreehari, E., and Ghantasala, P. G. (2019). Climate changes prediction using simple linear regression. *J. Comput. Theor. Nanosci.* 16, 655–658. doi:10.1166/jctn.2019.7785

Tao, H., Awadh, S. M., Salih, S. Q., Shafik, S. S., and Yaseen, Z. M. (2022). Integration of extreme gradient boosting feature selection approach with machine learning models: application of weather relative humidity prediction. *Neural Comput. Appl.* 34, 515–533. doi:10.1007/s00521-021-06362-3

Tzoumpas, K., Estrada, A., Miraglio, P., and Zambelli, P. (2024). A data filling methodology for time series based on cnn and (bi) lstm neural networks. *IEEE Access* 12, 31443–31460. doi:10.1109/access.2024.3369891

Villia, M. M., Tsagkatakis, G., Moghaddam, M., and Tsakalides, P. (2022). Embedded temporal convolutional networks for essential climate variables forecasting. *Sensors* 22, 1851. doi:10.3390/s22051851

Vo, T. Q., Kim, S.-H., Nguyen, D. H., and Bae, D.-H. (2023). Lstm-cm: a hybrid approach for natural drought prediction based on deep learning and climate models. *Stoch. Environ. Res. Risk Assess.* 37, 2035–2051. doi:10.1007/s00477-022-02378-w

Weirich-Benet, E., Pyrina, M., Jiménez-Esteve, B., Fraenkel, E., Cohen, J., and Domeisen, D. I. (2023). Subseasonal prediction of central european summer heatwaves with linear and random forest machine learning models. *Artif. Intell. Earth Syst.* 2, e220038. doi:10.1175/aies-d-22-0038.1

Ye, F., Hu, J., Huang, T.-Q., You, L.-J., Weng, B., and Gao, J.-Y. (2021). Transformer for ei niño-southern oscillation prediction. *IEEE Geoscience Remote Sens. Lett.* 19, 1–5. doi:10.1109/lgrs.2021.3100485