electronics

MDPI

Article

Hybrid Deep Learning and Stacking Ensemble Model for Time Series-Based Global Temperature Forecasting

Gökalp Çınarer

Department of Computer Engineering, Faculty of Engineering-Architecture, Yozgat Bozok University, 66100 Yozgat, Turkey; gokalp.cinarer@bozok.edu.tr

Abstract

Variations in global surface temperatures serve as critical indicators of climate change, and making accurate predictions regarding these patterns is essential for designing effective mitigation strategies. This study utilized a time series prediction methodology, leveraging annual temperature anomaly records from 1880 to 2022 provided by NASA's GISTEMP v4 dataset. Following an extensive preprocessing phase, multiple deep learning models, namely, LSTM, DNN, CNN, and Transformer, were trained and analyzed separately. The individual model outputs were subsequently combined using a weighted averaging strategy grounded in linear regression, forming a novel LSTM and Transformer-based hybrid forecasting model. Model performance was assessed through widely recognized metrics including MSE, MAE, RMSE, and R^2. By integrating the distinct advantages of each model, the ensemble framework aimed to improve the overall predictive capability. The findings revealed that this hybrid design delivered more stable and accurate forecasts compared to any single model. The integration of diverse neural network structures proved effective in boosting predictive reliability and underscored the viability of deep learning-based hybrid modeling for climate trend forecasting.

Keywords: global temperature anomaly; deep learning; hybrid time series forecasting; artificial intelligence

check for updates

Academic Editor: Xianzhi Wang

Received: 21 July 2025
Revised: 5 August 2025
Accepted: 7 August 2025
Published: 13 August 2025

Citation: Çınarer, G. Hybrid Deep Learning and Stacking Ensemble Model for Time Series-Based Global Temperature Forecasting. *Electronics* **2025**, *14*, 3213. https://doi.org/10.3390/electronics14163213

Copyright: © 2025 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Global shifts in climate patterns have emerged as one of most pressing issues of the last century, with significant impacts across environmental, economic, and social domains worldwide. Several driving forces, such as the increased consumption of fossil fuels, advancements in technology, and rising living standards, have led to a notable rise in global average temperatures [1]. A major contributing factor to global warming is the intensifying, elevated presence of climate-forcing gases in the atmosphere, largely driven by human activities. These gases, known for their heat-trapping capability, accumulate in the atmosphere and act as a primary catalyst for global temperature rises [2]. Consequently, the long-term monitoring and forecasting of global surface temperatures are of great importance for both environmental planning and disaster risk management. Climate change is known to directly affect several core sectors such as agriculture, fisheries, livestock, forest ecosystems, tourism, healthcare, construction, climate control, and logistics, producing both beneficial and adverse effects on global economic systems [3].

A temperature anomaly is defined as the deviation of an observed temperature value from a long-term reference mean over a specific time period. This methodology facilitates

the direct comparison of different regions and time spans, enabling a more accurate analysis of global temperature trends [4].

Long-term temperature trends play a vital role in tracking changes within climate systems and evaluating the consequences of such changes. Figure 1 presents the GISTEMP v4 dataset provided by NASA [4]. A consistent upward trend in global average surface temperatures is evident from the late 19th century to the present. While negative anomalies were dominant in the early 20th century, there has been a rapid increase in anomalies since the 1980s. Moving average curves further validate this long-term warming trend, highlighting a steady rise beyond short-term fluctuations. In recent years, anomalies have approached 1 °C, marking the highest values recorded in historical data. The graphical analysis reinforces the link between persistent climate warming and the underlying causes of global warming.

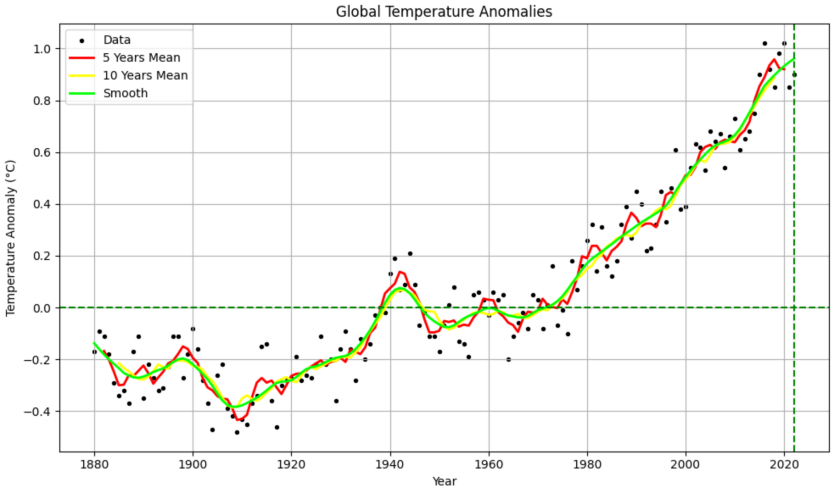


Figure 1. Annual global temperature anomalies and corresponding moving average trends from 1880 to 2022.

Accurately forecasting future global temperatures plays a pivotal role in assessing the evolving influence of climate variability and formulating effective climate risk reduction methods. Traditional statistical techniques often fall short when it comes to capturing long-term patterns and modeling the complex relationships within climatic datasets. As a result, deep learning-based models have gained significant traction in recent years for analyzing climate data and forecasting time series with long-range dependencies [5].

In particular, modern artificial neural network architectures such as LSTM (Long Short-Term Memory), DNNs (Deep Neural Networks), CNNs (Convolutional Neural Networks), and the Transformer model have shown strong performance in time series prediction tasks involving long-term dependencies [6]. Empirical studies in the literature confirm that these models generally outperform conventional approaches in terms of predictive accuracy and generalization capability.

In the current analysis, a time series analysis of global temperature anomalies was conducted using the NASA GISTEMP v4 dataset and future temperature anomalies were forecasted using various deep learning models. Forecasts based on historical data were generated using state-of-the-art neural network architectures including LSTM, CNN, DNN, and Transformer. The models were evaluated using standard error metrics such as RMSE (Root Mean Squared Error), MAE (Mean Absolute Error), MSE (Mean Squared Error), and R^2. To enhance prediction accuracy, a hybrid method based on linear regression was also adopted, combining the individual forecasts using a weighted averaging technique to produce final outputs. This approach overcomes the limitations of existing single-model

frameworks, offering advantages in more effectively capturing long-term climate variability, minimizing prediction errors, and enhancing generalization capability.

In this respect, this study combines the strengths of model-based approaches with real-world climate data, addressing a methodological gap while offering innovative solutions for environmental and geophysical time series applications. The main contributions to the literature can be summarized as follows: a novel hybrid deep learning model integrating LSTM and Transformer architectures is developed; outputs from different models are combined using a linear regression-based weighted averaging method to optimize forecasting performance; the effectiveness of deep learning-based approaches in predicting long-term climate trends is demonstrated using temperature anomaly data; individually trained LSTM, DNN, CNN, and Transformer models are analyzed with respect to their advantages and limitations. Furthermore, the applicability of deep learning techniques to environmental time series is showcased, providing a methodological foundation for future research in this field.

A flowchart illustrating the proposed hybrid model is presented in Figure 2. This study ultimately aims to contribute to scientific decision support systems by generating actionable insights from climate data.

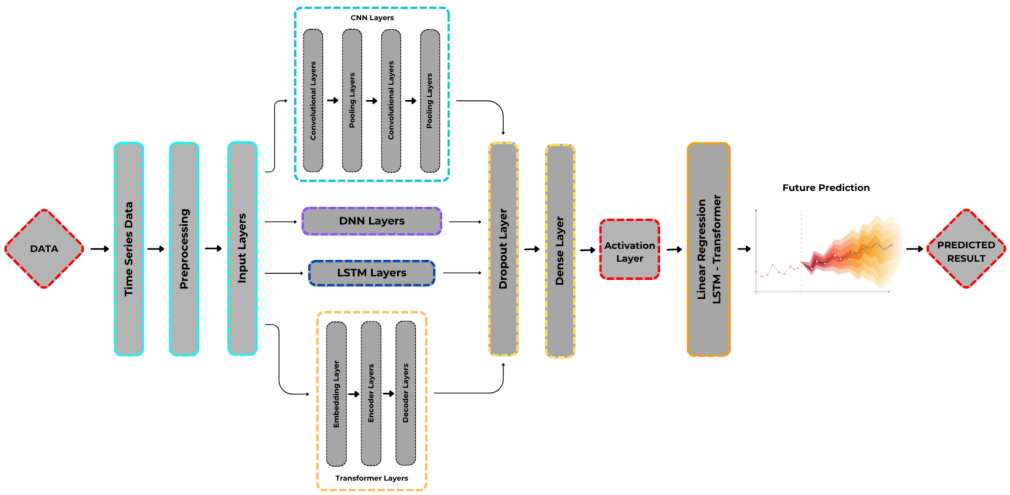


Figure 2. Overall architecture of the deep learning models employed for forecasting global temperature anomalies.

2. Literature Review

A review of the relevant literature reveals several noteworthy contributions in the field. Qingchun Guo et al. [7] employed multiple models including RNN, LSTM, CNN, ANN, and hybrid CNN-LSTM architecture to forecast six climatic parameters on a monthly basis. Their models were trained using data from 1951 to 2022, taking the previous 12 months as input. Among these, the CNN-LSTM hybrid outperformed other models by offering the highest accuracy and the lowest error rates. The authors highlighted its potential for improving climate prediction and contributing to disaster preparedness and resilience planning. Emy Alerskans and colleagues [8] proposed a statistical regression-based SST prediction algorithm using satellite data obtained from AMSR-E and AMSR2, within the framework of the ESA Climate Change Initiative. Their two-stage method estimated wind speed and SST using localized models and demonstrated strong predictive performance. The algorithm's accuracy was validated through comparison with in situ observations. Another study [9] introduced a hybrid deep learning framework combining TCN and Transformers. The TCN component extracted temporal local and global features, while the Transformer component used an attention mechanism to contextualize these features for

improved long-term dependency learning. The hybrid model outperformed standalone models across several benchmark time series datasets in both accuracy and generalization. Taylor et al. [10] developed a Unet-LSTM-based model to forecast global sea surface temperature anomalies up 24 months ahead, using monthly SST and 2 m air historical temperature measurements collected between 1950 and 2021. The model performed exceptionally well in climatologically critical regions such as the northeastern Pacific. It successfully predicted climate events like the 2019–2020 El Niño and 2016–2018 La Niña, although it showed reduced accuracy for the extreme El Niño event of 2015–2016. The study confirmed that data-driven methods offer strong potential for long-term SST anomaly forecasting. However, its reliance on only SST and air temperature limited its predictive scope. Xu et al. [11] proposed a regionally dynamic data processing approach for short-term SST prediction and introduced two LSTM-based architectures: MR-EDLSTM and MR-EDConvLSTM. Using OISST data, they demonstrated that MR-EDLSTM performed better in coastal zones, whereas MR-EDConvLSTM achieved greater accuracy in equatorial regions. The findings emphasized the superior accuracy and reduced error margins of deep learning models over traditional oceanographic methods. To enable high-resolution urban air temperature forecasts, Manzhu Yu et al. [12] developed an LSTM-based model trained on IoT sensor data collected along major transportation routes in New York City. Comparative analysis against ARIMA and FNN models showed that LSTM outperformed both, providing more accurate and reliable forecasts. Ridzna et al. [13] focused on the Cilacap region to predict surface temperature one year ahead, employing DES for trend estimation technique. Their study utilized data from the NASA POWER platform (2015–2025), incorporating surface temperature, solar radiation, and maximum wind speed at 10 m. The study highlighted DES as a low-cost, effective method for modeling seasonal temperature variations and emphasized its applicability in developing data-driven climate policy strategies. Fahad and colleagues [14] proposed a CNN-GRU-RNN hybrid deep learning model aimed at predicting climate variables through 2050. The model targeted four environmental indicators in Al-Kharj, Saudi Arabia: climatic variables such as temperature, dew point readings, visibility conditions, and sea-level atmospheric pressure. To address data imbalance, they applied Synthetic Minority Oversampling and Gaussian noise techniques. Evaluation was conducted using multiple metrics and the results showed that the hybrid model significantly outperformed traditional regression methods in long-term climate forecasting. Ayşegül et al. [15] investigated the impact of building design energy efficiency, emphasizing the importance of CDD in estimating cooling energy consumption. They used average CDD data from 1991 to 2022 provided by the Turkish State Meteorological Service and employed the Seasonal ARIMA model to forecast CDD values for period 2023–2040. Their results offer valuable insights for climate-responsive architectural planning. Xiaoxin and co-authors [16] developed a hybrid forecasting model based on a linear combination of GM and ARIMA approaches to improve the prediction of global temperature variations. They tested four weighting methods, with the standard deviation method yielding the most accurate results. Comparative experiments confirmed that the S-GM-ARIMA model provided higher accuracy and reliability, establishing it as a potentially valuable tool for climate policy-making. Xinxing et al. [17] highlighted the importance of accurately forecasting microclimatic conditions in greenhouses to enhance agricultural productivity and pest management. To address this, they proposed multi-step time series prediction-based Attention-LSTM model. Using roughly 48 h of past environmental data, the model was able to predict air and soil temperatures up to 480 min ahead with high accuracy. The results indicated that the model was highly effective in short-term forecasting of environmental variables and useful in optimizing greenhouse operations. Edward Appau and colleagues [18] developed multivariate time series models using the UCI database, drawing

on meteorological data from five Chinese cities between 2010 and 2015. Variables included meteorological parameters encompassing atmospheric heat levels, dew concentration, humidity percentages, barometric force, and combined wind flow speeds. Five RNN-based model configurations were tested, and the LSTM-RNN variant produced the lowest temperature prediction error among all. Uluocak et al. [19] proposed hybrid deep learning strategies for daily temperature forecasting, developing GRU–CNN and LSTM –CNN models. One-day-ahead forecasts were evaluated using statistical metrics and visual analysis. Both hybrid models outperformed other methods in short-term temperature prediction. Finally, Bilgili et al. [20] applied LSTM, SARIMA, and GRU for time series forecasting based on global sea surface temperature data. Their performance was evaluated using metrics, with all three models delivering high prediction accuracy. The results validated the effectiveness of these methods for practical forecasting applications.

3. Materials and Methods

In this study, various deep learning models were employed to analyze global temperature anomalies and forecast future trends. The dataset utilized consisted of annual "J-D" temperature anomaly values from NASA's GISTEMP v4, spanning the period from 1880 to 2022. State-of-the-art deep learning architectures including LSTM, GRU, DNN, CNN, and Transformer were implemented for time series forecasting. Additionally, several hybrid configurations combining these models were incorporated into this study. Each model was specifically structured for time series prediction tasks. Model performance was evaluated using statistical error metrics such as RMSE, MAE, R^2, and MSE. Furthermore, comparisons were made based on training time and the number of trainable parameters.

All training and evaluation procedures for the deep learning models were conducted using the Python 3.12 programming language. The modeling pipeline relied primarily on the TensorFlow and Keras libraries for model construction, training, and testing. Additional tasks such as data preprocessing, normalization, and visualization were carried out using widely adopted open-source libraries, including NumPy, Pandas, Matplotlib 3.10, and Seaborn. Performance metrics were computed using the metrics module from the Scikit-learn library.

The entire model development and testing process was executed on a high-performance workstation located in a university laboratory. This system was equipped with an Intel Core i7, NVIDIA GeForce RTX 4090 GPU, and high-capacity RAM, which significantly reduced training durations and enhanced the overall efficiency of model validation workflows.

3.1. Dataset

For the purpose of forecasting global temperature anomalies using time series analysis, this study utilized the GISTEMP v4 dataset provided by NASA's Goddard Institute for Space Studies [4]. GISTEMP is a globally recognized climate dataset that reports surface temperature anomalies on both annual and monthly scales. It is constructed by merging observations from GHCN and ERSST, curated by NASA. The anomalies are computed relative to a baseline reference period from 1951 to 1980 and global means are derived using a gridded parceling system.

The annual anomaly values labeled "J-D" were selected for analysis as they offer a clearer depiction of year-to-year global climate trends. The "J-D" series represents the average of monthly temperature anomalies from January through December for each year. To increase the number of data points, the annual data were converted to monthly resolutions after removing missing values, and linear interpolation was preferred as it ensured a stable transition that preserved the trend structure without introducing artificial fluctuations. The data were scaled to the [0, 1] range using Min–Max normalization to

eliminate the effects of different scales and improve the model's learning process. Although the dataset includes values from 1880 to 2023, only data from 1880 to 2022 were used in this analysis, as the 2023 data were incomplete and only reflected the first half of the year. A time series plot of these annual anomalies is presented in Figure 3.

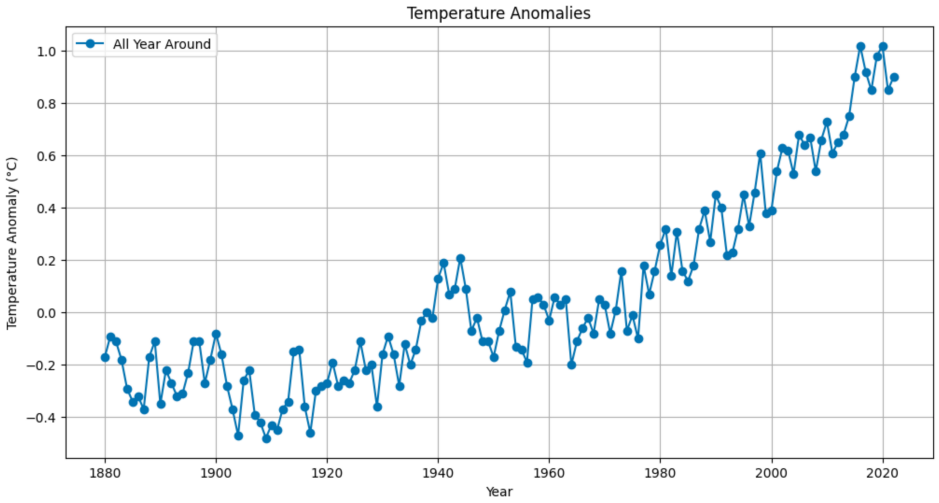


Figure 3. Time series graph illustrating monthly global temperature anomalies from 1880 to 2022.

Figure 4 illustrates the global surface temperature anomalies for the year 2025, relative to the 1951–1980 reference period, based on NASA GISTEMP analysis. The map highlights a pronounced warming trend in recent years, particularly across the Arctic, North America, and large portions of Europe and Asia.

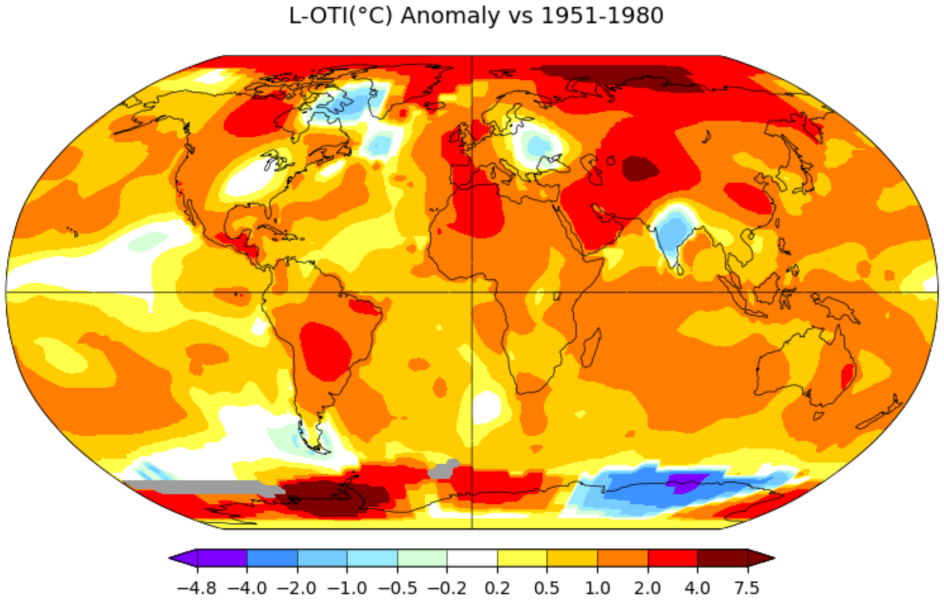


Figure 4. Heat map illustrating global surface temperature anomalies for the year 2025 based on NASA GISTEMP data [4].

Figure 5 presents a graph of annual global temperature anomalies along with the corresponding linear trend. The trend exhibits a positive slope, quantified as 0.00781 °C per year, indicating that the global mean temperature anomaly has been increasing by approximately 0.00781 °C annually. Notably, the graph reveals an acceleration in temperature rise following the 1980s, providing statistically significant evidence of ongoing global warming.

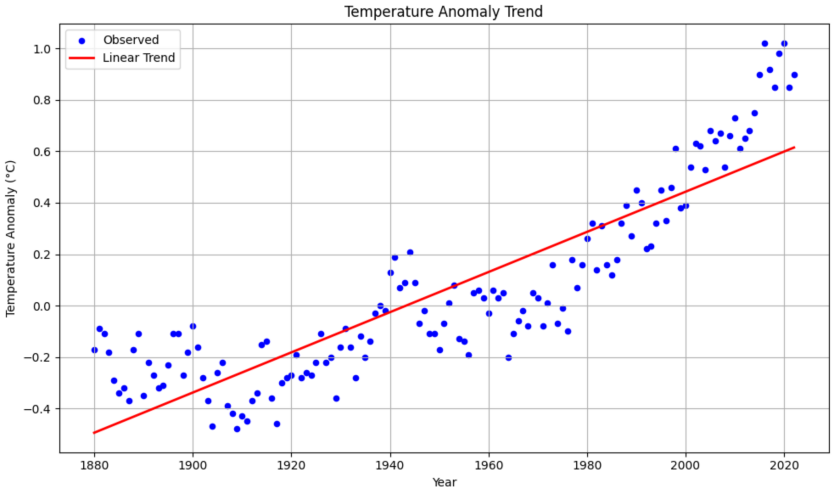


Figure 5. Visualization of global temperature anomalies from 1880 to 2022 with a linear trend line.

Figure 6 displays the results of a change point detection analysis, performed using the Binary Segmentation method, applied to the annual temperature anomaly time series. The vertical dashed lines indicate statistically significant structural shifts within the series. These identified change points correspond to the years 1930, 1940, 1945, 1980, and 2000, each marking potential periods of structural change, trend shifts, or accelerations in the climate system.

In the change point detection analysis, the *p*-value was found to be less than 0.01, indicating that the identified change points are statistically significant. The Z-value, which reflects the strength of the change point, showed a notably high value of 12.7, suggesting a strong level of significance. Furthermore, the slope value was calculated as 0.007 and, when evaluated together with the *p*-value, this result confirms that the detected change points are both meaningful and have a substantial impact.

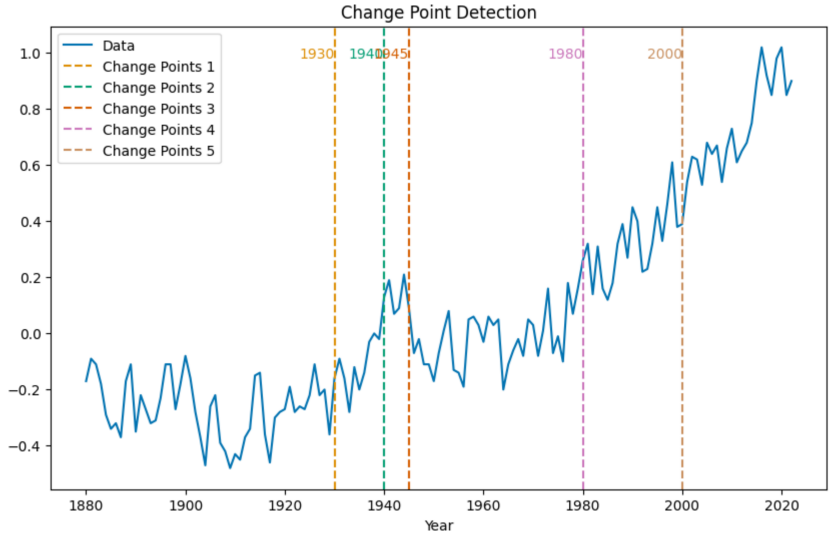


Figure 6. Visualization of structural breakpoints identified in the annual temperature anomaly time series.

3.2. Hyperparameters

Table 1 summarizes the key hyperparameters utilized during the training phase for developed model. The effect of each parameter on the model's performance was determined using Grid Search. The influence of each parameter on model performance was determined

through empirical experimentation. The architecture consisted of hidden layers comprising 128 and 64 neurons, respectively. This design was selected to effectively control the model's parameter count. ReLU was employed as an activation function between layers, which helped mitigate the issue of diminishing gradients and improved the stability of the deep learning process. MSE was chosen as the loss function, owing to its sensitivity to large errors and its widespread use in time series regression tasks. The Adam optimizer was adopted for model training, with the learning rate set to 0.001. The training process was configured with 100 epochs and a batch size of 16. The dataset was divided into 70% training, 15% validation, and 15% test sets and the results were evaluated based on the test data. Additionally, the data was subjected to Min–Max normalization before being fed into the model. This configuration aimed to strike a balance between model adaptability to training data and the ability to generalize unseen data. Overall, these hyperparameter settings were selected to improve prediction accuracy while maintaining robustness across different datasets. The hyperparameters presented in Table 1 were kept constant for each model. In this way, it was assumed that the observed performance differences stemmed solely from the architectural structures, and any inconsistencies in training conditions were avoided. Additionally, for CNN model, the kernel size was set to 3 to capture short-term temporal patterns effectively, while the pool size was set to 2 to reduce dimensionality and extract the most significant features, thereby improving computational efficiency.

Table 1. Hyperparameter values used during the training of deep learning models.

Parameter	Value
Number of Layers	2
Layer Dimensions	256, 128
Activation Function	ReLU
Loss Function	MSE
Optimization Algorithm	Adam
Learning Rate	0.001
Number of Epochs	50
Batch Size	64
Dropout Rate	0.2
Time Step	60

3.3. Deep Learning Models

3.3.1. LSTM

LSTM architecture is a variant of RNN specifically crafted to model long-term dependencies within sequential data. This design was introduced to overcome the limitations of standard RNNs, particularly their tendency to forget information over extended time intervals. The LSTM unit integrates three essential gates—forget gate (f_t), input gate (i_t), and output gate (o_t)—which collectively manage the information flow and update the internal memory cell (C_t) [21]. These gating mechanisms allow the network to selectively preserve or discard historical data when necessary [22]. Due to this dynamic control mechanism, LSTM networks have been widely utilized in areas such as time series forecasting, the processing of natural language, and climate-related data modeling. An illustration of the LSTM framework is provided in Figure 7.

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (1)$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (2)$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (3)$$

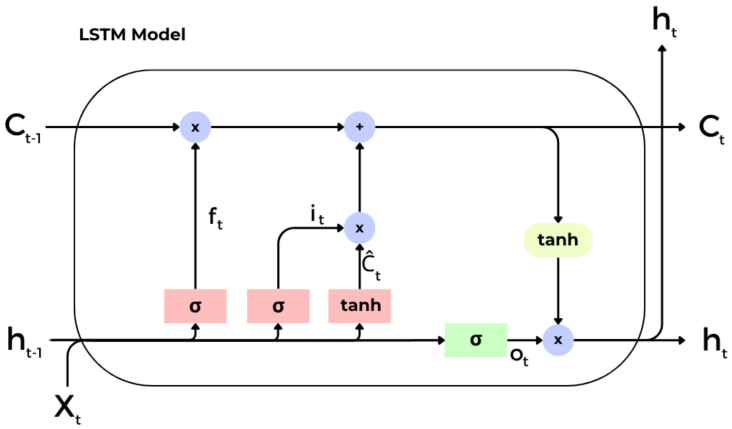


Figure 7. Architectural diagram of the LSTM model illustrating cell structure and information flow.

3.3.2. DNN

DNN is a hierarchical machine learning model composed of multiple interconnected layers of artificial neurons [23]. Each neuron in a layer receives inputs from the preceding layer, performs weighted summation, incorporates bias terms, and then passes the result through a nonlinear activation function to generate the layer's output. These models are particularly effective in time series prediction tasks as they can capture intricate and high-dimensional relationships between past observations and future values [24]. A schematic representation of the DNN architecture is provided in Figure 8.

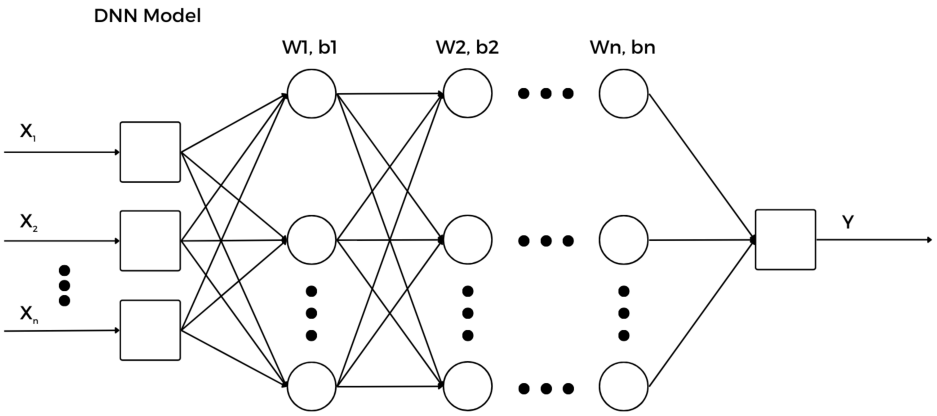


Figure 8. Architectural diagram of the DNN model showing the layer structure.

3.3.3. Transformer

The Transformer model, originally proposed by Vaswani et al. [25], was primarily designed for natural language processing tasks. In recent years, however, it has also been successfully applied to various domains, including time series forecasting. Unlike traditional RNN-based architectures, the Transformer processes all time steps in parallel rather than sequentially. This parallelization significantly accelerates training on large datasets and enhances the model's ability to capture long-term dependencies more effectively. The structural design of the Transformer model is illustrated in Figure 9.

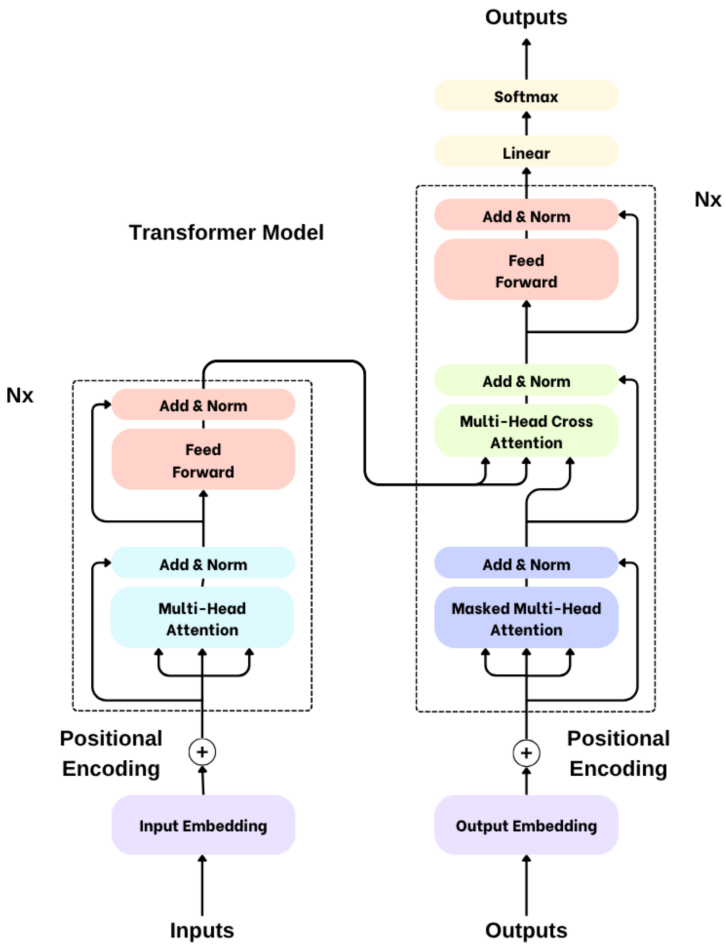


Figure 9. Overview of the Transformer architecture showing the attention mechanism and layer components.

The model computes how much attention each query should pay to other inputs. The similarity score, obtained via dot product, is scaled by $\sqrt{d_k}$ and then normalized using the softmaxfunction. The resulting attention weights are subsequently multiplied by the value matrix (V) to generate output representations. To facilitate this, each input vector is projected into three distinct representations: (Query, Q), Key, K ve Value, V [25]. These projections are computed by multiplying the input vector X with corresponding learnable parameter matrices W^Q, W^K, and W^V, respectively.

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^\top}{\sqrt{d_k}}\right)V \tag{4}$$

Multi-head attention enables the model to learn diverse representations. Rather than relying on a single attention mechanism, multiple attention heads are utilized concurrently. The architecture enables each head to independently emphasize unique components of the input, allowing richer parallel representation, thereby enhancing the model's expressiveness and flexibility.

$$\text{head}_i = \text{Attention}(QW_i^Q, KW_i^K, VW_i^V) \tag{5}$$

$$\text{MultiHead}(Q, K, V) = \text{Concat}(\text{head}_1, \ldots, \text{head}_h)W^O \tag{6}$$

Unlike RNN, the Transformer architecture does not inherently encode sequential information [26]. To address this, positional encoding is added to input embeddings using

sinusoidal and cosine functions, allowing the model to capture the order and structure of data.

$$PE_{(pos,2i)} = \sin\left(\frac{pos}{10000^{\frac{2i}{d_{\text{model}}}}}\right) \tag{7}$$

$$PE_{(pos,2i+1)} = \cos\left(\frac{pos}{10000^{\frac{2i}{d_{\text{model}}}}}\right) \tag{8}$$

The positional encoding layer was set with D_Model 256 to incorporate temporal information into the model. In addition, the multi-head attention mechanism used Num_Heads 4 and Key_Dim 32 to learn relationships across different representation subspaces in parallel.

3.4. Performance Metrics

This approach ensured consistency between the data used during model training and the evaluation process, enabling meaningful comparisons across different models. However, it also limited the direct interpretability of these metrics in terms of physical units. Therefore, this context should be taken into account when interpreting results.

3.4.1. RMSE

RMSE is a widely used metric in regression analysis. Root mean square error measures the typical magnitude of prediction errors by squaring the residuals, finding their mean, and taking the square root [27]. A lower value indicates better model performance as this reflects smaller average deviations between predictions.

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{9}$$

In the above formula, y_i denotes actual values, $\hat{y}_i$ represents predicted values produced by the model and n refers to the total number of observations. Since RMSE directly reflects the magnitude of prediction errors, it is frequently used in the literature for comparative model evaluations.

3.4.2. MSE

MSE is a fundamental error metric used in regression models, calculated by averaging squared differences between predicted, actual values. Due to its sensitivity to large errors, MSE is employed to assess the accuracy of predictions. An increase in its value indicates a decline in a model's performance [28].

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{10}$$

3.4.3. MAE

MAE is an error metric calculated by taking the average of absolute differences between predicted, actual values. It is often favored for its straightforward interpretation, as it assigns equal weights to all errors, regardless of their magnitude [29].

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{11}$$

3.4.4. R^2

R^2 is a commonly used metric that quantifies how much of the total variance in the dependent variable is explained by the regression model. It usually falls within the interval from 0 to 1, with values nearer to 1 suggesting that the model accounts for a larger proportion of the data's variance [30].

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}i)^2}{\sum i = 1^n (y_i - \bar{y})^2} \tag{12}$$

4. Proposed Model

In this research, a deep learning-based hybrid forecasting model is proposed to enhance the accuracy of global temperature anomaly predictions using the GISTEMP v4 dataset. The model follows a multi-stage architecture comprising individual model training, hyperparameter optimization, ensemble integration, and final evaluation. The overall framework, illustrated in Figure 10, encompasses all steps from data preprocessing to the final prediction output.

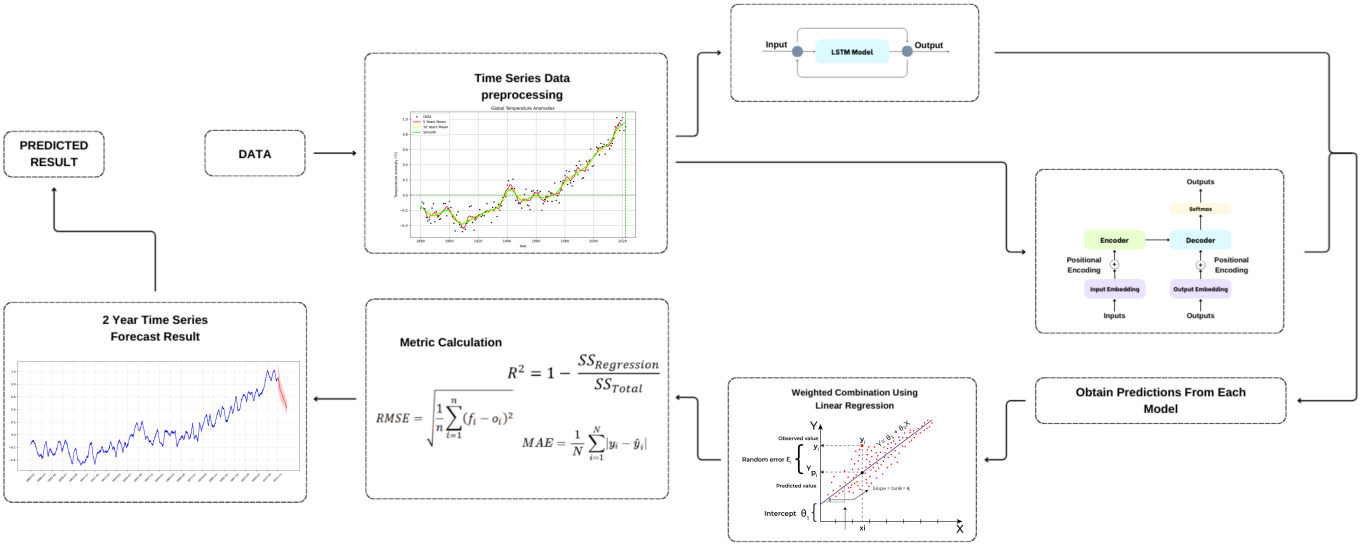


Figure 10. Proposed hybrid model architecture encompassing all steps from data processing to modeling and forecasting.

The primary reason for employing both LSTM and Transformer architectures in this study is their complementary ability to effectively capture both long-term trends and short-term fluctuations in time series data. LSTM excels at learning long-term dependencies through its gating mechanisms, making it a powerful tool for modeling slowly changing patterns in climate data. Transformer, on the other hand, introduces an innovative multi-head attention mechanism that not only addresses sequential dependencies but also focuses on long-range relationships and key features across the entire sequence. This innovative design enables more flexible and parallelizable modeling compared to LSTM, particularly for complex and multi-scale climate dynamics. By combining the long-term dependency modeling strength of LSTM with Transformer's innovative attention mechanism, the proposed approach achieves more accurate prediction of both trends and sudden changes.

The architectures, training processes, and hyperparameter tuning of the four core models of LSTM and Transformer were carefully designed. Each model was optimized specifically for time series forecasting, and their complementary strengths were integrated to construct a robust ensemble framework.

To maximize model performance, hyperparameter tuning was conducted using both Grid Search and Bayesian Optimization techniques. Grid Search, in a structured manner, explores all possible combinations within predefined hyperparameter ranges to identify the optimal configuration [31]. However, due to its computational intensity, especially with complex models, Bayesian Optimization was also employed. This approach leverages prior evaluations to guide the search process more efficiently, aiming to reach optimal solutions with fewer evaluations [32].

- Hidden layer size: $h \in \{128, 256\}$
- Number of layers: $L \in \{2\}$
- Learning rate: $\eta \in \{0.001\}$

In the Transformer architecture, the *softmax* function plays a pivotal role within the self attention mechanism. This mechanism allows the model to capture dependencies between each element in an input sequence and all other elements.

$$Q = XW^Q, \quad K = XW^K, \quad V = XW^V \tag{13}$$

This operation is performed to compute the similarity scores between query and key matrices.

$$\text{Score}(Q, K) = QK^\top \tag{14}$$

Since the resulting similarity scores can become numerically large, they are scaled by the square root of dimensionality of key vectors in order to stabilize model's learning process.

$$\text{Scaled Score}(Q, K) = \frac{QK^\top}{\sqrt{d_k}} \tag{15}$$

At this stage, the *softmax* function is employed to compute attention distribution for each position relative to all others. By normalizing the scaled similarity scores, the softmax transforms them into a probability distribution [33]. This allows the model to determine on a probabilistic basis how much attention should be allocated to each input position.

Afterward, the individual predictions are evaluated using accuracy metrics to ensure quality control. Following this step, the Stacking Ensemble method is applied to combine the outputs of multiple models in a more effective manner. Stacking is an advanced technique that aggregates the strengths of individual models to produce more accurate, robust, and generalizable predictions. It has shown notable success in time series forecasting, as well as in classification and regression tasks. In this approach, predictions from each base model are combined using a linear regression-based stacking mechanism. In this method regression learns optimal weight coefficients for each model's output based on the validation data, and the final prediction is computed on the test data as follows:

$$\hat{y}_{final,t} = \beta_1 \cdot \hat{y}_{LSTM,t} + \beta_4 \cdot \hat{y}_{Transformer,t} \tag{16}$$

- $\hat{y}$: Represents the final combined prediction.
- $x_1, x_2, \ldots, x_n$: Denote the outputs from each individual deep learning model.
- β_0: Acts as the intercept term in the regression equation.
- $\beta_1, \beta_2, \ldots, \beta_n$: Represent the regression coefficients corresponding to each model's prediction.

In this way, information from each model's prediction is integrated in a weighted manner according to its performance, leading to more robust, generalized, and balanced results.

In the final stage, the output of this hybrid system is used to forecast future temperature anomalies. The forecasting process follows an iterative prediction approach. The model

is trained to produce one-step-ahead predictions, where each forecasted value serves as the input for the subsequent year. Based on this strategy, the model first used the data up to the year 2022 to generate a prediction for 2023. Then, the predicted value for 2023 was used to estimate the value for 2024, and so on. This process was repeated step by step until projections were obtained up to the year 2047. This approach aligns with single-step learning and is a widely adopted technique in time series forecasting. The proposed framework aims to deliver more reliable and stable predictions by leveraging the strengths of diverse model architectures.

5. Results

Within the scope of this study, deep learning-based models were implemented to forecast global temperature anomalies using time series data. After training artificial neural network models with different architectures both individually and in a hybrid manner, their performances were directly compared on the test set, and the results are presented in Table 2. To assess the contribution of each model within the ensemble, individual models and selected model combinations were executed and ablation experiments were conducted to evaluate performance differences.

Table 2. Performance comparison of the implemented models in terms of RMSE, MSE, MAE, R^2, number of parameters, and training time.

MODELS	RMSE	MSE	MAE	R^2	PARAM	SEC
LSTM	0.0228	0.0005	0.0183	0.9764	461441	491.73
DNN	0.0464	0.0021	0.0425	0.9023	8129	12.58
CNN	0.0257	0.0007	0.0216	0.9699	951553	36.68
Transformer	0.0264	0.0007	0.0208	0.9683	165761	223.07
LSTM CNN	0.0240	0.0006	0.0194	0.9739	49985	67.99
LSTM DNN	0.0371	0.0014	0.0338	0.9375	74881	128.43
DNN CNN	0.0302	0.0009	0.0263	0.9586	61697	17.25
LSTM DNN CNN	0.0264	0.0007	0.0221	0.9683	35393	50.03
Proposed Model	0.0219	0.0004	0.0171	0.9783	313601	357.40

When the performance comparisons of various deep learning models for forecasting global temperature anomalies were evaluated based on standard metrics, it was observed that the models exhibited relatively close performances overall. The proposed hybrid model outperformed all others, achieving the lowest error rates and highest accuracy according to the RMSE (0.0219), MSE (0.0004), MAE (0.0171), and R^2 (0.9783). These results demonstrate that the linear regression-based weighted combination of outputs from the LSTM and Transformer models performed more effectively than any individual model.

Among the non-hybrid models, the DNN model achieved the lowest RMSE (0.0228), while the CNN–LSTM hybrid yielded the lowest MSE (0.0006). Although the Transformer model showed slightly reduced accuracy, as indicated by a lower R^2 value of 0.9683, it offers practical advantages due to its relatively smaller number of parameters and reduced computational cost. Conversely, the LSTM and CNN architectures were found to be the most resource-intensive in terms of parameter size and training duration.

Figure 11 presents the prediction performance of the proposed Stacking Ensemble model on the test dataset, where the period from 2001 to 2021 represents the test data. In the graph, the blue line represents actual temperature anomaly values, while the green line indicates the model's predictions during the test period. Focusing on the post-2000 era, the figure illustrates that the model effectively captured the overall warming trend. Even during periods of sudden fluctuations, gaps between predicted and actual values remained minimal. This suggests that the model successfully learns meaningful patterns and is

capable of generalizing to future data. The obtained results confirm that the hybrid structure formed by linearly combining the outputs of the LSTM and Transformer models offers a robust and accurate solution. In this context, the hybrid model not only integrates the strengths of individual architectures but also contributes to generating reliable predictions in complex time series tasks such as climate forecasting.

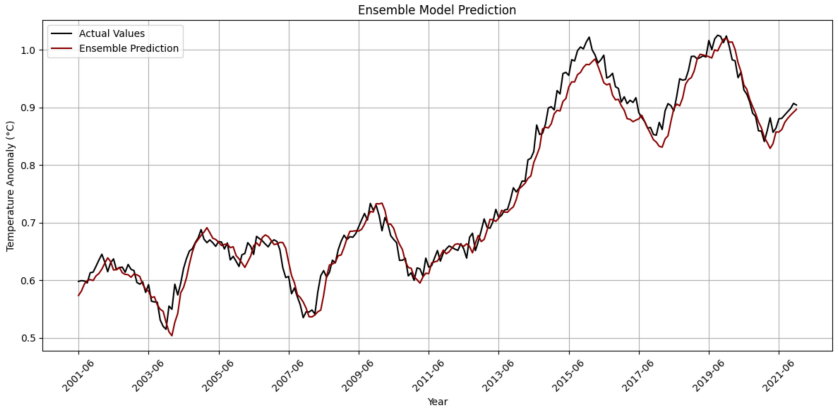


Figure 11. Comparison between actual and predicted values on the test data using the proposed hybrid model.

Figure 12 presents the 24-month temperature anomaly forecasts. The blue line represents the observed historical temperature anomalies, while the red line indicates the model's forecasted values. The shaded red area denotes the ±20% confidence band around the predictions. As observed in the graph, the model predicted a declining trend in temperature anomalies over the short term. Furthermore, the widening of the confidence band over time reflects the increasing uncertainty associated with long-term forecasts. Overall, the model's predictions and the associated uncertainty range suggest ongoing variability in the climate system, with the projections remaining within a reasonable confidence interval.

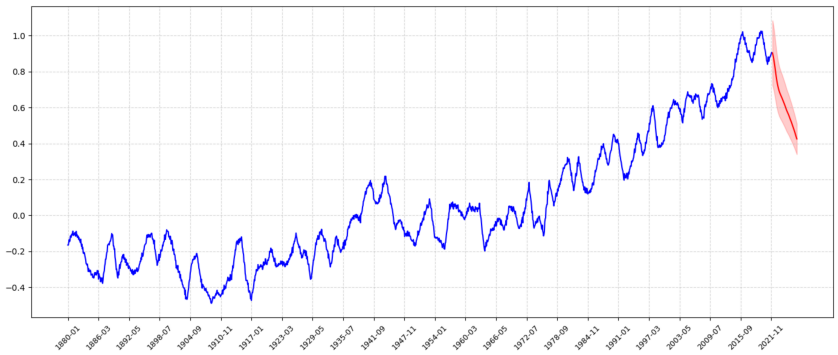


Figure 12. Future projection graph of global temperature anomaly forecasts from 2025 to 2027 generated by the proposed hybrid model.

Figure 13 illustrates the relationship between the predicted values generated by the proposed model and actual observed temperature anomalies in the test dataset. The horizontal axis represents actual temperature anomaly values, while the vertical axis corresponds to the model's predictions. The distribution of data points enables an assessment of the model's prediction accuracy. The dashed black line in the plot represents the ideal scenario, where predicted values perfectly match observed values. The proximity of points to this line is indicative of the model's accuracy. A majority of the points are closely clustered around the ideal line, suggesting that the model yields highly accurate predictions.

Furthermore, the R^2 value achieved by the model serves as a strong indicator of its overall performance. In the context of environmental datasets, such a high R^2 score is particularly significant, reinforcing the reliability and robustness of the model's forecasting capabilities.

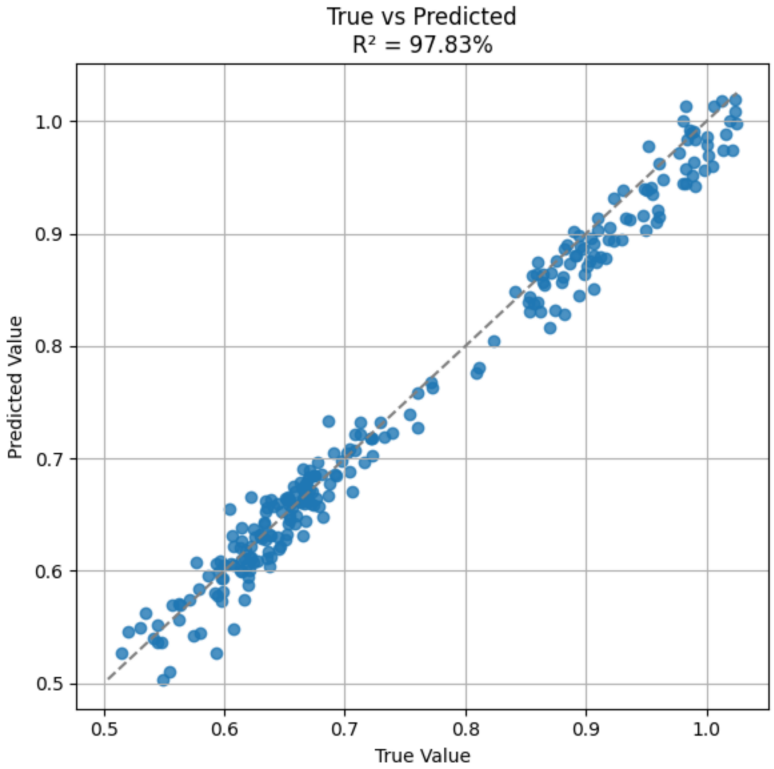


Figure 13. R^2 regression plot illustrating the relationship between actual and predicted values by the proposed model. The blue points represent individual data samples, while the dashed line denotes the ideal 1:1 correspondence between true and predicted values.

Figure 14 presents a comparative visualization of the predictions generated by the individual LSTM and Transformer models during the test period and for future years. The black line represents the observed temperature anomaly values, while the colored lines correspond to the forecast trajectories of each respective model.

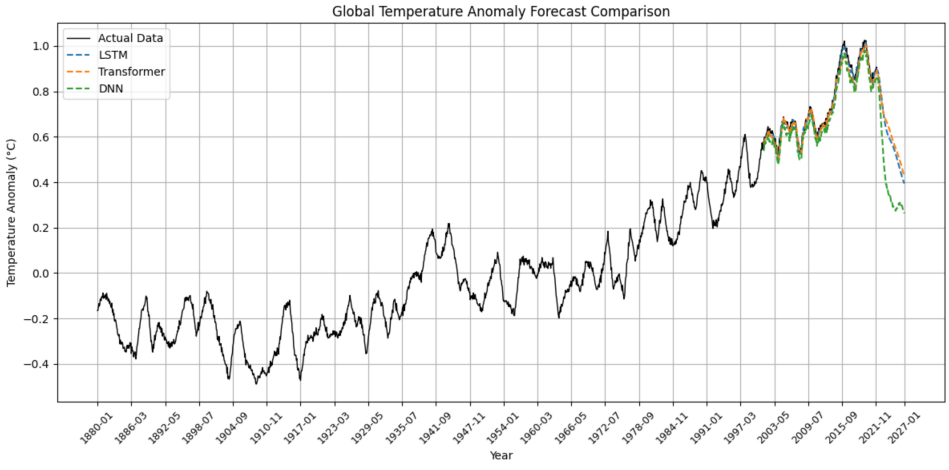


Figure 14. Comparison of temperature anomaly forecasts from LSTM and Transformer models for test data and future years.

The first half of Figure 14 illustrates the performance of the models on the test dataset, while the second half displays their forecasts for future years. Notably, during the test period, the predicted curves of the LSTM and DNN models closely aligned with the observed data, suggesting that these models exhibit superior generalization capabilities based on historical trends.

In the forecasting segment, more pronounced divergences between models are observed. The LSTM model predicted a steeper increase in temperature anomalies, whereas the other models suggested more moderate warming patterns. These variations offer valuable insights into how each architecture interprets future temporal dynamics. The CNN and Transformer models, owing to their sophisticated structures, tended to capture long-term trends more effectively; however, this could occasionally lead to overestimations.

Overall, this figure provides a comprehensive view of both short-term predictive accuracy and long-term forecasting tendencies across different models. It also highlights why the proposed hybrid model, which integrates outputs from these individual architectures, demonstrates enhanced predictive performance.

All models converged to low loss values within a short number of training epochs, indicating that the preprocessing steps and hyperparameter configurations were well-optimized. The LSTM and Transformer models showed smoother and more stable learning curves, while CNN exhibited rapid convergence. DNN, in contrast, displayed slight fluctuations during validation. Figure 15 offers a valuable visualization for analyzing the learning behavior of each component model within the proposed hybrid framework.

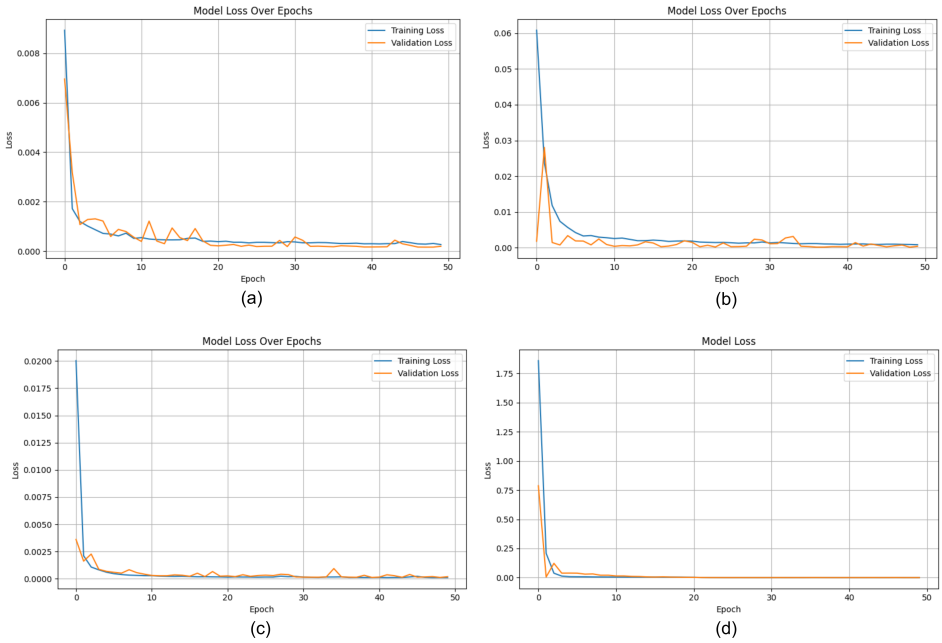


Figure 15. (**a**) LSTM, (**b**) DNN, (**c**) CNN, and (**d**) Transformer models' loss value changes during the training process.

Figure 16 demonstrates that, overall, all models achieved a meaningful level of accuracy in forecasting temperature anomalies. Among them, the CNN model stood out with slightly superior performance in terms of predictive accuracy, while the other models also produced comparably reliable results. This comparative analysis reinforces the rationale behind constructing the proposed hybrid architecture from these individual models, as it consistently outperforms the standalone approaches.

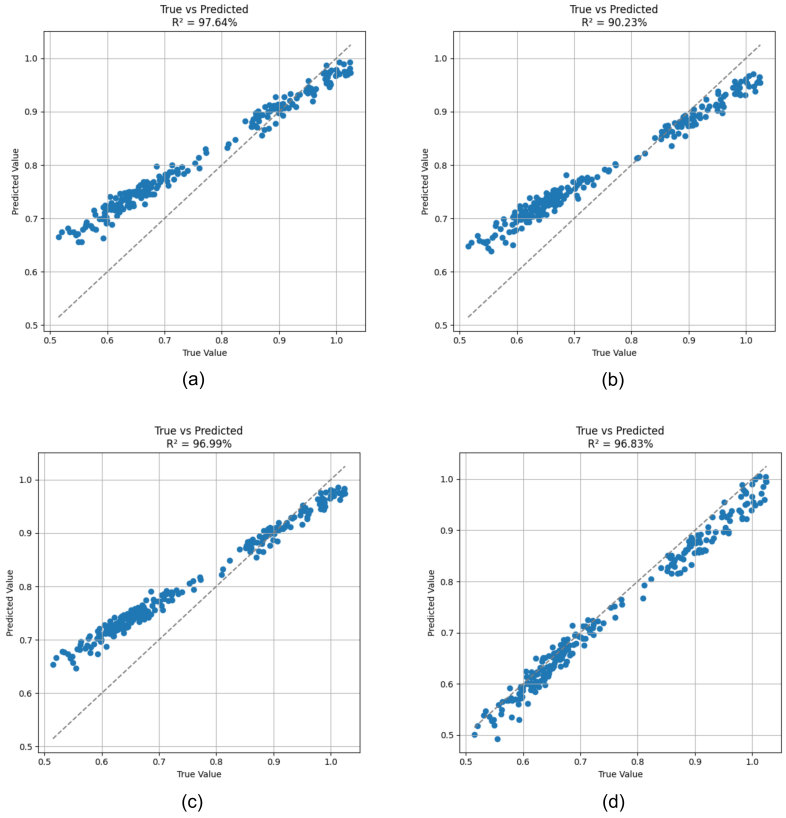


Figure 16. (**a**) LSTM, (**b**) DNN, (**c**) CNN, and (**d**) Transformer models' R^2 accuracy plots calculated for the test dataset.

In conclusion, the findings indicate that the proposed hybrid approach effectively integrates the strengths of the individual models, thereby enhancing its generalization capability. Accordingly, the proposed architecture demonstrated its potential to serve as a reliable tool for the long-term forecasting of climate data.

6. Discussion

As part of this study, a selection of recent AI-based research focusing on forecasting tasks is summarized in Table 3. In the table, a "-" symbol indicates that the corresponding performance metric was not explicitly reported in the referenced study.

Table 3. Performance comparison of temperature forecasting studies in the literature using different datasets, methods, and models.

Reference	Dataset	Method	Model	RMSE	MAE	MSE	R^2
[34]	Sea Surface Temperature Anomaly Data	Short-Term Temperature Forecasting	LSTM–AdaBoost	0.39	0.29	-	-
[35]	Daily Temperature Data	Short-Term Temperature Forecasting	CNN–LSTM	1.97	1.02	3.88	0.72
[36]	Daily Maximum Temperature Data	Long-Term Temperature Forecasting	ARIMA	1.56	1.14	-	0.86
[37]	Monthly Maximum and Minimum Temperature Data	Temperature Forecasting	CNN–GRU	1.82	-	-	0.95

Table 3. *Cont.*

Reference	Dataset	Method	Model	RMSE	MAE	MSE	R^2
[38]	Weekly Temperature Data	Short-Term Temperature Forecasting	GRU	0.97	0.71	-	0.61
[39]	Meteorological Data	Short-Term Weather Forecasting	GNN–TCNN	1.21	-	-	-
			GNN–LSTM	1.25	-	-	-
[40]	Daily Minimum and Maximum Temperature Values	Short-Term Temperature Prediction	T–LSTM	2.70	2.20	7.0	-
[41]	Urban Temperature Projections	Long-Term Urban Temperature Distribution	WRF–Temporal Fusion Transformer	0.53	0.49	-	-
[42]	Meteorological Data	Short-Term Forecast	LSTM	0.99	0.82	-	0.63
[Ours]	**NASA GISTEMP v4**	**Long-Term Temperature Forecasting**	**Proposed Model**	**0.0219**	**0.0171**	**0.0004**	**0.9783**

Several recent studies in the literature have explored the use of deep learning techniques for temperature prediction. Nair et al. [43] proposed an LSTM-based framework for global surface temperature forecasting, demonstrating improved accuracy over decision tree models. However, their work solely focused on the LSTM model without benchmarking against other deep learning architectures. In contrast, the present study evaluated a range of architectures, including LSTM, DNN, CNN, and Transformer, both individually and within a hybrid framework constructed using a linear regression-based weighted averaging method. Hou et al. [35] developed a hybrid CNN–LSTM model to predict hourly air temperature, reporting performance metrics of RMSE = 1.97 and R^2 = 0.72. Compared to this, our proposed model achieved significantly better accuracy, with an RMSE of 0.0219 and an R^2 score of 0.9783. Siddique et al. [36] used ARIMA-based models to forecast temperature and precipitation in the Mymensingh region of Bangladesh to support fishery planning. Although their study provided localized insights into climate impacts on agriculture and aquaculture, it did not include comparative analyses or deep learning-based modeling, which our study incorporated using GISTEMP v4 and advanced neural architectures. Haghrahmani et al. [37] focused on monthly maximum and minimum temperature forecasting in the United Arab Emirates using DNN and CNN–GRU hybrid models. Their results demonstrated the capacity of deep models to capture seasonal and temporal patterns in temperature series. Similarly, our study leveraged LSTM and Transformer models to extract latent patterns in long-term temperature anomaly sequences and combined their outputs to enhance overall prediction performance. Shahriar et al. [39] applied deep learning to forecast the Fire Weather Index (FWI) across the United States using meteorological features such as temperature, humidity, wind speed, and precipitation. They compared several models and found that the GNN-TCNN hybrid model performed best for short- to mid-term forecasts. While their work addressed climate-related hazard prediction, our model emphasizes annual-scale temperature anomaly prediction with an emphasis on computational efficiency and generalizability through hybrid learning. Elshewey et al. [44] proposed the CNN–ResNet50–LSTM hybrid model for short-term wind and temperature forecasting, achieving high accuracy. However, their study lacked temporal consistency and generalization across different regions due to reliance on heterogeneous datasets. In contrast, our approach focuses on long-term, globally representative forecasts by integrating the strengths of the LSTM, DNN, CNN, and Transformer models

into a unified ensemble. Zhu et al. [41] introduced a hybrid model combining the WRF physical model and the Temporal Fusion Transformer (TFT) to predict urban temperatures with high accuracy while reducing computational cost. Similarly, the hybrid framework presented in this study effectively integrates multiple deep learning models to produce accurate annual temperature anomaly forecasts from time series data.

7. Conclusions

Using the proposed hybrid model, this study presented 25-year forecasts of global annual temperature anomalies based on the GISTEMP v4 dataset. The dataset, published by NASA, covers the period from 1880 to 2022 and provides comprehensive records of global surface temperature anomalies. According to the model forecasts, a short-term decline in temperature anomaly is observed after 2025, decreasing from approximately 1.02 °C to around 0.42 °C by the end of 2027. Although this represents an estimated 59% reduction, it is not indicative of a reversal in the long-term global warming trend. Rather, it reflects natural variability within the climate system and falls within the model's uncertainty range. Temporary cooling patterns should not be misinterpreted as a sign of diminishing climate change impacts.

Future values were generated using an iterative prediction strategy. In this approach, the model was trained to forecast one time step ahead, and each prediction was subsequently used as input for the next step. Through this process, sequential estimates were made for the years following 2023.

Although temperature anomaly forecasts are subject to uncertainty due to inherent variability in the climate system such as regional effects and atmospheric dynamics, the hybrid model's architecture, which integrates the strengths of various model types, produced more stable and generalizable outcomes. For future research, extending the framework to multivariate time series that incorporate additional variables such as greenhouse gas concentrations, ocean currents, and solar activity could further improve predictive accuracy. Moreover, enhancing spatial resolution and modeling regional climate patterns would enable more localized forecasting. These advancements would strengthen climate-related decision support systems and contribute meaningfully to mitigation planning efforts.

Funding: This research received no external funding.

Data Availability Statement: Data are contained within the article.

Conflicts of Interest: The author declares no conflicts of interest.

Abbreviations

The following abbreviations are used throughout this study.

ANN	Artificial Neural Network
CDDs	Cooling Degree Days
CNN	Convolutional Neural Network
DES	Double Exponential Smoothing
DL	Deep Learning
DNN	Deep Neural Network
FNN	Feedforward Neural Network
FWI	Fire Weather Index
GRU	Gated Recurrent Unit
GISTEMP	Goddard Institute for Space Studies Surface Temperature Analysis
LR	Linear Regression
LSTM	Long Short-Term Memory
MAE	Mean Absolute Error

ML	Machine Learning
MSE	Mean Squared Error
PE	Positional Encoding
ReLU	Rectified Linear Unit
RMSE	Root Mean Squared Error
RNN	Recurrent Neural Network
R^2	Coefficient of Determination
SST	Sea Surface Temperature
TCN	Temporal Convolutional Network
TFT	Temporal Fusion Transformer

References

1. Akın, G. Küresel ısınma, nedenleri ve sonuçları. *Ank. üNiversitesi Dil-Tar.-CoğRafya FakülLtesi Derg.* **2006**, *46*, 29–43. [CrossRef]
2. Özmen, M.T. Sera gazı-küresel ısınma ve Kyoto Protokolü. *İMO Derg.* **2009**, *453*, 42–46.
3. Bayraç, H.N. Enerji kullanımının küresel ısınmaya etkisi ve önleyici politikalar. *EskişEhir Osman. üNiversitesi Sos. Bilim. Derg.* **2010**, *11*, 229–259.
4. GISTEMP Team. *GISS Surface Temperature Analysis (GISTEMP)*, Version 4; NASA Goddard Institute for Space Studies: New York, NY, USA, 2025. Available online: https://data.giss.nasa.gov/gistemp/ (accessed on 15 January 2025).
5. Chen, C.; Dong, J. Deep learning approaches for time series prediction in climate resilience applications. *Front. Environ. Sci.* **2025**, *13*, 1574981. [CrossRef]
6. Bandara, K.; Bergmeir, C.; Smyl, S. Forecasting across time series databases using recurrent neural networks on groups of similar series: A clustering approach. *Expert Syst. Appl.* **2020**, *140*, 112896. [CrossRef]
7. Guo, Q.; He, Z.; Wang, Z. Monthly climate prediction using deep convolutional neural network and long short-term memory. *Sci. Rep.* **2024**, *14*, 17748. [CrossRef]
8. Alerskans, E.; Høyer, J.L.; Gentemann, C.L.; Pedersen, L.T.; Nielsen-Englyst, P.; Donlon, C. Construction of a climate data record of sea surface temperature from passive microwave measurements. *Remote Sens. Environ.* **2020**, *236*, 111485. [CrossRef]
9. Isreal, O.; Alonge, M. Combining Temporal Convolutional Networks and Transformer Models for Time Series Forecasting. Preprint. 2025. Available online: https://www.researchgate.net/publication/390494107_Combining_Temporal_Convolutional_Networks_and_Transformer_Models_for_Time_Series_Forecasting (accessed on 15 May 2025).
10. Taylor, J.; Feng, M. A deep learning model for forecasting global monthly mean sea surface temperature anomalies. *Front. Clim.* **2022**, *4*, 932932. [CrossRef]
11. Xu, T.; Zhou, Z.; Li, Y.; Wang, C.; Liu, Y.; Rong, T. Short-term prediction of global sea surface temperature using deep learning networks. *J. Mar. Sci. Eng.* **2023**, *11*, 1352. [CrossRef]
12. Yu, M.; Xu, F.; Hu, W.; Sun, J.; Cervone, G. Using long short-term memory (LSTM) and Internet of Things (IoT) for localized surface temperature forecasting in an urban environment. *IEEE Access* **2021**, *9*, 137406–137418. [CrossRef]
13. Purwanto, R.A.; Hakim, D.K. Optimization of Double Exponential Smoothing Model for Daily Earth Temperature Forecasting in Dayeuhluhur, Cilacap. *J. E-Komtek* **2025**, *9*, 61–73.
14. Aljuaydi, F.; Zidan, M.; Elshewey, A.M. A Deep Learning CNN-GRU-RNN Model for Sustainable Development Prediction in Al-Kharj City. *Eng. Technol. Appl. Sci. Res.* **2025**, *15*, 20321–20327. [CrossRef]
15. Bilgili, A.; Çelik, K.; Bilgili, M. Analysis of historical and future cooling degree days over Türkiye for facade design and energy efficiency in buildings. *J. Therm. Anal. Calorim.* **2024**, *149*, 7413–7431. [CrossRef]
16. Chen, X.; Jiang, Z.; Cheng, H.; Zheng, H.; Cai, D.; Feng, Y. A novel global average temperature prediction model—Based on GM-ARIMA combination model. *Earth Sci. Inform.* **2024**, *17*, 853–866. [CrossRef]
17. Li, X.; Zhang, L.; Wang, X.; Liang, B. Forecasting greenhouse air and soil temperatures: A multi-step time series approach employing attention-based LSTM network. *Comput. Electron. Agric.* **2024**, *217*, 108602. [CrossRef]
18. Nketiah, E.A.; Chenlong, L.; Yingchuan, J.; Aram, S.A. Recurrent neural network modeling of multivariate time series and its application in temperature forecasting. *PLoS ONE* **2023**, *18*, e0285713. [CrossRef]
19. Uluocak, I.; Bilgili, M. Daily air temperature forecasting using LSTM-CNN and GRU-CNN models. *Acta Geophys.* **2024**, *72*, 2107–2126. [CrossRef]
20. Bilgili, M.; Pinar, E.; Durhasan, T. Global monthly sea surface temperature forecasting using the SARIMA, LSTM, and GRU models. *Earth Sci. Inform.* **2025**, *18*, 10. [CrossRef]
21. Greff, K.; Srivastava, R.K.; Koutník, J.; Steunebrink, B.R.; Schmidhuber, J. LSTM: A search space odyssey. *IEEE Trans. Neural Netw. Learn. Syst.* **2016**, *28*, 2222–2232. [CrossRef]
22. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef]

23. Bhanja, S.; Das, A. Impact of data normalization on deep neural network for time series forecasting. *arXiv* **2018**, arXiv:1812.05519.
24. Gopali, S.; Abri, F.; Siami-Namini, S.; Namin, A.S. A comparative study of detecting anomalies in time series data using LSTM and TCN models. *arXiv* **2021**, arXiv:2112.09293. [CrossRef]
25. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention Is All You Need. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 5998–6008.
26. Zhang, Q.; Lu, H.; Sak, H.; Tripathi, A.; McDermott, E.; Koo, S.; Kumar, S. Transformer transducer: A streamable speech recognition model with transformer encoders and rnn-t loss. In Proceedings of the ICASSP 2020—2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Barcelona, Spain, 4–9 May 2020; pp. 7829–7833.
27. Chai, T.; Draxler, R.R. Root mean square error (RMSE) or mean absolute error (MAE). *Geosci. Model Dev. Discuss.* **2014**, *7*, 1525–1534.
28. Willmott, C.J.; Matsuura, K. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Clim. Res.* **2005**, *30*, 79–82. [CrossRef]
29. Hyndman, R.J.; Athanasopoulos, G. *Forecasting: Principles and Practice*, 2nd ed.; OTexts: Melbourne, Australia, 2018. Available online: https://otexts.com/fpp2/ (accessed on 16 May 2025).
30. Chicco, D.; Warrens, M.J.; Jurman, G. The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Comput. Sci.* **2021**, *7*, e623. [CrossRef]
31. Liashchynskyi, P.; Liashchynskyi, P. Grid search, random search, genetic algorithm: A big comparison for NAS. *arXiv* **2019**, arXiv:1912.06059. [CrossRef]
32. Shahriari, B.; Swersky, K.; Wang, Z.; Adams, R.P.; De Freitas, N. Taking the human out of the loop: A review of Bayesian optimization. *Proc. IEEE* **2015**, *104*, 148–175. [CrossRef]
33. Hajra, S.; Alam, M.; Saha, S.; Picek, S.; Mukhopadhyay, D. On the instability of softmax attention-based deep learning models in side-channel analysis. *IEEE Trans. Inf. Forensics Secur.* **2023**, *19*, 514–528. [CrossRef]
34. Xiao, C.; Chen, N.; Hu, C.; Wang, K.; Gong, J.; Chen, Z. Short and mid-term sea surface temperature prediction using time-series satellite data and LSTM-AdaBoost combination approach. *Remote Sens. Environ.* **2019**, *233*, 111358. [CrossRef]
35. Hou, J.; Wang, Y.; Zhou, J.; Tian, Q. Prediction of hourly air temperature based on CNN–LSTM. *Geomat. Nat. Hazards Risk* **2022**, *13*, 1962–1986. [CrossRef]
36. Siddique, M.A.B.; Mahalder, B.; Haque, M.M.; Ahammad, A.K.S. Forecasting air temperature and rainfall in Mymensingh, Bangladesh with ARIMA: Implications for aquaculture management. *Egypt. J. Aquat. Res.* **2025**, *in press*. [CrossRef]
37. Haghrahmani, S. Enhancing temperature prediction in the UAE: A process-driven framework for adaptive learning with GRU-CNN hybrid models. *Model. Earth Syst. Environ.* **2025**, *11*, 100. [CrossRef]
38. Zhang, J.; Ding, Y.; Zhu, L.; Wan, Y.; Chai, M.; Ding, P. Estimating and forecasting daily reference crop evapotranspiration in China with temperature-driven deep learning models. *Agric. Water Manag.* **2025**, *307*, 109268. [CrossRef]
39. Shahriar, S.A.; Choi, Y.; Islam, R. Advanced deep learning approaches for forecasting High-Resolution fire weather index (FWI) over CONUS: Integration of GNN-LSTM, GNN-TCNN, and GNN-DeepAR. *Remote Sens.* **2025**, *17*, 515. [CrossRef]
40. Karevan, Z.; Suykens, J.A. Transductive LSTM for time-series prediction: An application to weather forecasting. *Neural Netw.* **2020**, *125*, 1–9. [CrossRef]
41. Zhu, H.C.; Ren, C.; Wang, J.; Feng, Z.; Haghighat, F.; Cao, S.J. Fast prediction of spatial temperature distributions in urban areas with WRF and temporal fusion transformers. *Sustain. Cities Soc.* **2024**, *103*, 105249. [CrossRef]
42. Long, X.; Wang, J.; Gong, S.; Li, G.; Ju, H. Reference evapotranspiration estimation using long short-term memory network and wavelet-coupled long short-term memory network. *Irrig. Drain.* **2022**, *71*, 855–881. [CrossRef]
43. Nair, R.R.; Ebin, P.M.; Babu, T. Advancing climate change forecasting: LSTM-based earth surface temperature prediction across 88 global cities. In Proceedings of the 2024 International BIT Conference (BITCON), Dhanbad, India, 7–8 December 2024; pp. 1–6.
44. Elshewey, A.M.; Jamjoom, M.M.; Alkhammash, E.H. An enhanced CNN with ResNet50 and LSTM deep learning forecasting model for climate change decision making. *Sci. Rep.* **2025**, *15*, 14372. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.