EAI Endorsed Transactions
on AI and Robotics
Research Article EAI.EU

An Efficient Hybrid Model With Harris Hawks Optimization Algorithm for Predicting Oat Water

Jinbo He[1], Zirun Wang[1,*], Wanzhen Huang[1], Fanting Zhou[1], Anqi Wang[1,3,4], Xianjin Wu[1], Pengzi Chen[1], Tongli He[2,*], Jianhong Gan[1,3,4,5], Peiyang Wei[1,3,4,5], Zhibin Li[1,3,4,5], Chunjiang Wu[1]

[1]College of Software Engineering, Chengdu University of Information Technology, Chengdu 610225
[2]College of Applied Mathematics, Chengdu University of Information Technology, Chengdu 610225
[3]Sichuan Key Laboratory of Software Automatic Generation and Intelligent Service, Chengdu University of Information Technology, Chengdu 610225
[4]Key Laboratory of Meteorological Software, China Meteorological Administration, Chengdu 610225
[5]Dazhou Key Laboratory of Government Data Security, Sichuan University of Arts and Science, Dazhou 635000

Abstract

Oats are a cold-tolerant and high-yielding cultivated forage crop, which have relatively high requirements for water management. To improve the forecasting accuracy of irrigation requirements for oats, this paper proposes a novel hybrid neural network architecture, whose parameters are refined using the Harris Hawk optimization algorithm. It directly addresses two prevalent shortcomings in current predictive models: the typically imprecise manual adjustment of hyperparameters, and the inadequate modeling of both spatial and temporal dependencies with the data. By integrating these methodological improvements, the proposed method aims to achieve more precise and robust forecasts. Firstly, the hybrid model integrates a temporal convolutional network, a bidirectional long short-term memory network, and a multi-head attention mechanism. This method leverages multi-head attention to enrich feature representation, thereby facilitating a more comprehensive capture of the temporal dynamics inherent in alpine oat water demand. Secondly, the Harris Hawk optimization method is introduced to optimize the model's hyperparameters, effectively avoiding local optimum. Experimental results on the oat water demand and environmental dataset from 2019 to 2023 indicate that the hybrid model achieved a mean absolute error and root mean square error of 0.3432 and 0.4863, respectively, thus representing reductions of approximately 52.73% and 46.39% compared to the traditional Long Short-Term Memory (LSTM) model. The coefficient of determination increased by about 15.31%. Ablation study results demonstrate that the complete hybrid model achieved Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) of 0.3432 and 0.4863, respectively, which represent reductions of approximately 52.89% and 43.75% compared to the baseline model, the coefficient of determination improving by about 22.16%. Compared with other methods, this method has a distinct advantage in forecasting the precision of oat water demand, which offers technical and decision support for smart irrigation.

Received on 21 February 2026; accepted on 01 March 2026; published on 04 march 2026

Keywords: Harris Hawk Optimization, TCN-BiLSTM-MHA, oat water demand, hybrid neural network, multi-head attention mechanism

doi:10.4108/airo.12011

*Corresponding author. Email: htl@cuit.edu.cn; wzrsy@outlook.com

1. Introduction

In recent years, with the continuous growth of the world's population and increasingly severe weather conditions, the agricultural water shortage has become increasingly prominent. According to the

United Nations Food and Agriculture Organization (FAO) data, agricultural irrigation represents more than 70% [1] of the world's total consumption of freshwater resources. The Pengbo Irrigation Area in Tibet faces significant agricultural water security challenges, characterized by acute scarcity of water resources, highly variable pre-cipitation patterns, and substantial rates of evaporative loss.

It is difficult for traditional irrigation techniques to satisfy the water needs of crops such as oats [2–4]. Therefore, the study of forecasting methods for the water demand of oats, adapted to the specific features of the area, is very important to improve the efficiency of water use and yield.

The estimation of agricultural irrigation requirements has transitioned from foundational empirical and mathematical modeling [5] techniques toward contemporary, data-centric methodologies enabled by advances in artificial intelligence. Current research predominantly employs individual prediction models, such as empirical models (e.g., Penman-Monteith equation), BP neural networks, support vector machines, and LSTM [6–11]. For example, the feasibility of data-driven approaches has been verified by using BP neural networks for irrigation management purposes [12]. However, there are still many problems with the ability to capture nonlinear temporal features, to achieve satisfactory prediction accuracy, and to provide robust generalization [13].In response to this, researchers have studied the combination of prediction models. For example, Li et al [14] constructed a hybrid model of MGM-ARIMA and BP-ARIMA, Juan W, Pute W., Xining Z. [15], and BP Neural Networks, and Z. Tong, S. Zhang, J. Yu [16] put forward a combination model called CARS-CatBoost. The results show that the prediction accuracy can be improved by means of model complementarity. In order to improve the accuracy of the model, researchers have also improved the performance of the model by combining modal decomposition with attention mechanisms [17]. Research demonstrates that integrating time-series analysis with a feature-weighting mechanism effectively enhances prediction accuracy.

Recent advancements in AI-driven agricultural water management have demonstrated the potential of hybrid deep learning models in capturing complex spatiotemporal dependencies. For instance, Ye et al. (2024) proposed a UNet-ConvLSTM framework that integrates remote sensing data from MODIS and GLDAS to accurately predict agricultural water demand, achieving R^2 values of 0.927 and 0.935, respectively. Their work highlights the effectiveness of combining spatial feature extraction with temporal modeling, which aligns with the motivation of our study to enhance prediction accuracy in high-altitude cold regions through a novel hybrid architecture[18].

At present, Adaptive Neuro-Fuzzy Inference Systems (ANFIS) [19] and Kalman filtering have been used to correct real-time data, and a neural network forecast model (for example, LSTM) based on weather data and the Penman-Monteith equation [20, 21] has been built. The initial closed-loop irrigation of oats was realized with fuzzy control. However, the optimization of hyperparameters remains a significant challenge, which makes it difficult to support accurate forecasting and irrigation management.

This study proposes an innovative algorithm for hyperparameter tuning to estimate the irrigation needs of oat crops cultivated in cold, high-altitude zones. The proposed approach combines Temporal Convolutional Network (TCN), Bidirectional Long Short-Term Memory (BiLSTM), and Multi-Head Attention (MHA). The Harris Hawks Optimization (HHO) algorithm [22] is used to improve the performance of the model by more than 15%. Moreover, a new method of feature engineering is presented, which combines statistical standard and agricultural knowledge. Based on multi-phase feature selection, a low-redundant and highly representative feature set is constructed. This method reduces the input dimension by 28% and improves the precision of the outlier detection to 92%.The unique climatic and geographical conditions of high-altitude cold regions—such as large diurnal temperature variations, sparse meteorological stations, and high data missing rates—pose significant challenges for accurate crop water demand forecasting. Traditional models often struggle to capture these complex nonlinear patterns and handle noisy, incomplete datasets, necessitating a more robust and adaptive modeling approach.

2. Related Methods

2.1. Temporal Convolutional Network

In the forecast of crop water demand, if the weather data is used as the main data source, it is necessary to analyze the features of the data source and to understand its spatiotemporal distribution, resolution, and reliability in order to guarantee the precision and adaptability of the forecast model. Weather data is one of the typical spatiotemporal data. Mining its time feature has a great influence on the precision of the water demand forecast. The Temporal Convolutional Network (TCN [23, 24]) was introduced in 2018 by Bai Shaojie. It is based on Convolutional Neural Networks (CNN). In addressing constraints associated with standard neural networks, the present work modifies a convolutional framework by incorporating causal convolution layers. This structural adjustment ensures that the model adheres to the temporal dependencies within sequential data.

Within Temporal Convolutional Networks (TCNs), causal convolution enforces a strict chronological order. Specifically, the output at any time step t in a higher layer is derived solely from inputs at time t and earlier in the layer below, thereby maintaining a unidirectional and autoregressive information flow.

A key advantage of the TCN architecture is its capacity for strictly sequential processing, a property that inherently aligns with the fundamental nature of time-series data. If the convolution kernel is $K = (k_1, k_2, k_3, \ldots, k_{n-1}, k_n)$ and the time sequence of input is $X = (x_1, x_2, , x_3 \ldots, x_{t-1}, x_t)$, then the causal convolution output at time t can be represented by:

$$y = \sum_{i=1}^{n} k_i \cdot X_t - i + 1 \tag{1}$$

where y_t represents the value of the output sequence at position t, k_i represents the weight of the convolution kernel at position i, $X_t - i + 1$ represents the value of the input sequence at position $t - i + 1$, n denotes the size of the convolution kernel, and i is the sequence index.

The expansion rate functions as a supplementary variable, enabling the model to assimilate contextual data from an extended timeframe, thereby augmenting its overall receptive field.

Based on the expansion factor d, the expansion convolution reduces the number of convolutional neural networks by a factor of 2, which makes it possible to extract the features from the data.

The formula for the dilated convolution operation at infinity is as follows:

$$yt = \sum_{i=1}^{n} k_i \cdot x_t - (i-1)\, d \tag{2}$$

where y_t represents the dilated convolution value of the output sequence at position x_t, $x_{t-(i-1)d}$ represents the state at the previous time t with the dilation parameter d; n is the size of the convolution kernel; d is the dilation rate.

2.2. Bidirectional Long Short-Term Memory Network Model

The Bidirectional Long Short-Term Memory (BiLSTM) has been proposed as a framework for deep learning. BiLSTM combines two independent LSTM networks, one to capture forward information and the other to retrieve information. The key mechanism of LSTM is the input gate, forget gate, and output gate. Through a series of nonlinear operations, the data flow is controlled, and the memory is renewed.

The input gate controls the input of the current time step t, and its equation is as follows:

$$i_t = \sigma\left(W_i \cdot [h_{t-1}, x_t] + b_i\right) \tag{3}$$

In the input gate it, the Sigmoid activation function σ is used for calculation, involving the input weight matrix W_i, the previous hidden state h_{t-1}, the current input, and the bias term b_i. The forget gate f_i controls the retention of the previous cell state C_{t-1}, and its calculation formula is as follows:

$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] + b_f\right) \tag{4}$$

The output of the forget gate f_t ranges between 0 and 1, regulating the degree of influence of the previous memory cell C_{t-1} on the current time step. The memory cell C_t is updated by the input gate and the forget gate, which is shown as:

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tanh\left(W_C \cdot [h_{t-1}, x_t] + bc\right) \tag{5}$$

Here is the memory state, and tanh is the hyperbolic tangent activation function. The output gate determines if the hidden state of the current time step is transferred to the next level. Its formula is as follows:

$$O_t = \sigma\left(W_o \cdot [h_{t-1}, x_t] + b_o\right) \tag{6}$$

Finally, the update formula for the hidden state h_t is:

$$h_t = o_t \cdot \tanh\left(C_t\right) \tag{7}$$

In BiLSTM, the forward LSTM and the backward LSTM process the time sequence in the forward and backward directions. For an input sequence $X = \{x_1, x_2, \ldots, x_t\}$, a forward LSTM output is calculated by:

$$h_t = \mathrm{LSTM}_{\mathrm{forward}}\ (x_t) \tag{8}$$

The output of the backward LSTM is as follows:

$$h_t = \mathrm{LSTM}_{\mathrm{backward}}\ (x_t) \tag{9}$$

The final output of BiLSTM is the concatenation of the forward and backward hidden states:

$$h_t = [h_f, h_f] \tag{10}$$

2.3. Multi-Head Attention Mechanism

The Multi-Head Attention (MHA)[25] mechanism projects the input sequence into multiple linear subspaces, each yielding a unique representation subset.

Each attention head independently computes the relationships between different sequence positions, generating its own contextualized output. These outputs are then concatenated and linearly transformed to produce the final representation, enhancing both model performance and learning efficiency. The attention mechanism can selectively focus on the features that have high correlation with the task, and suppress

the unwanted features, thereby improving the performance of the model.

Employing a scaled dot-product mechanism, the attention computation utilizes three core matrices. The Query matrix (Q) actively requests information, while the Key matrix (K) serves as an identifier, allowing the relevance to Q to be measured through their similarity. The Value matrix (V) holds the feature vectors, and a weighted sum of these values, based on the computed relevance, generates the final context. The multi-head attention mechanism consists of multiple self-attention structures, which can process the same feature information simultaneously.

Three attention mechanisms are used for concatenation.This architecture is able to capture the dependencies among the various features and further enhance the performance of the model, the formula is as follows:

$$\text{Attention}\left(Q,K,V\right)=\text{soft max}\left(\frac{QK^{T}}{\sqrt{d_{k}}}\right)V \tag{11}$$

$$\text{head}_{i}=\text{Attention}\left(QW_{i}^{Q},KW_{i}^{K},VW_{i}^{V}\right) \tag{12}$$

$$\text{MultiHead}\left(Q,K,V\right)=\text{Concat}\ \left(\text{head}\ _{1},\ldots,\text{head}\ _{h}\right)W^{O} \tag{13}$$

where d_k is the dimension of the Key vector (or Query vector) in each attention head after the input sequence is linearly projected, which helps maintain the stability of gradients during training; head represents the i-th multi-head attention; MultiHead is formed by concatenating the outputs of multiple attention heads to form a larger feature matrix, which is then passed through a linear layer to generate the final output.

Attention mechanisms have also been successfully applied in other domains, such as knowledge graph completion, to model relational dependencies[26].

2.4. Hyperparameter Optimization

Harris Hawks Optimization (HHO) is a meta-heuristic algorithm inspired by the cooperative hunting strategies of Harris hawks.

Through the simulation of the research, development, and attack strategies of hawk populations, HHO constructs a highly effective optimization framework, which has been applied in many fields.

The HHO algorithm includes two major stages: exploration and development. The exploration phase emulates hawks scanning a wide solution space, thus focusing primarily on the algorithm's capacity to conduct a comprehensive global search.

The exploitation stage simulates the behavior of hawks surrounding and attacking prey, emphasizing local search. The escape energy parameter E dynamically controls the shift from one phase to the other, guaranteeing thorough search coverage in early stages and progressively shifting toward targeted, localized exploitation as iterations advance.

In the exploration stage, the HHO algorithm uses two strategies to carry out global search. First of all, a random search strategy is employed, the formula is as follows:

$$X\left(t+1\right)=X_{\text{rand}}\ \left(t\right)-r_{1}\cdot\left|X_{\text{rand}}\ \left(t\right)-2r_{2}\cdot X\left(t\right)\right| \tag{14}$$

where $X(t)$ is the current position of the hawk, $X_{\text{and}}\ (t)$ is the randomly selected position of another hawk, and r_1 and r_2 are random numbers within [0,1]. Secondly, a search strategy based on the leader hawk is adopted, the formula is as follows:

$$X\left(t+1\right)=X_{best}\left(t\right)-X\left(t\right)-r_{3}\cdot\left(L_{B}+r_{4}\cdot\left(U_{B}-L_{B}\right)\right) \tag{15}$$

where X_{best} is the position of the leader hawk, L_B and U_B are the lower and upper bounds of the solution space, respectively, and r_3 and r_4 are random numbers. When $|E| >= 1$, the soft copy strategy is executed, whose the related formula is as follows:

$$X\left(t+1\right)=X_{\text{best}}\ \left(t\right)-X\left(t\right) \tag{16}$$

When $0.5 < |E| < 1$, the hard besiege strategy is implemented, the formula is shown in Equation (17):

$$X\left(t+1\right)=X_{\text{best}}\ \left(t\right)-E\cdot\left|J\cdot X_{\text{best}}\ \left(t\right)-X\left(t\right)\right| \tag{17}$$

where J is the random jump strength, which is usually set to $2\left(1-\text{rand}\left()\right)$.

When $|E| < 0.5$, the rapid dive strategy is performed, the formula is as shown in Equation (18):

$$X\left(t+1\right)=\begin{cases}Y & \text{if } F\left(Y\right)<F\left(X\left(t\right)\right)\\ Z & \text{if } F\left(Z\right)<F\left(X\left(t\right)\right)\end{cases} \tag{18}$$

where the calculation formulas of Y and Z are shown in Equations (19) and (20), respectively:

$$Y=X_{\text{best}}\ \left(t\right)-E\cdot\left|J\cdot X_{\text{best}}\left(t\right)-X\left(t\right)\right| \tag{19}$$

$$Z=Y+S\cdot LF(D) \tag{20}$$

Compared with traditional optimization algorithms like Particle Swarm Optimization (PSO), the HHO algorithm offers unique advantages for this task. Its escape energy parameter E dynamically balances exploration and exploitation, effectively preventing premature convergence to local optima—a common limitation of PSO in high-dimensional spaces. Additionally, HHO's multiple attacking strategies (soft besiege, hard besiege, and rapid dive) enhance adaptability across the solution space, making it well-suited for optimizing complex deep learning architectures such as TCN-BiLSTM-MHA.

3. Method

3.1. TCN-BiLSTM-MHA Model

To model the complex temporal dependencies and nonlinear volatility in oat water demand, a TCN-BiLSTM-MHA [27, 28]hybrid architecture is proposed. It synergistically integrates dilated convolutions for long-range patterns, bidirectional recurrent layers for contextual dependencies, and attention mechanisms for dynamic feature weighting. The overall framework is depicted in Figure 1.

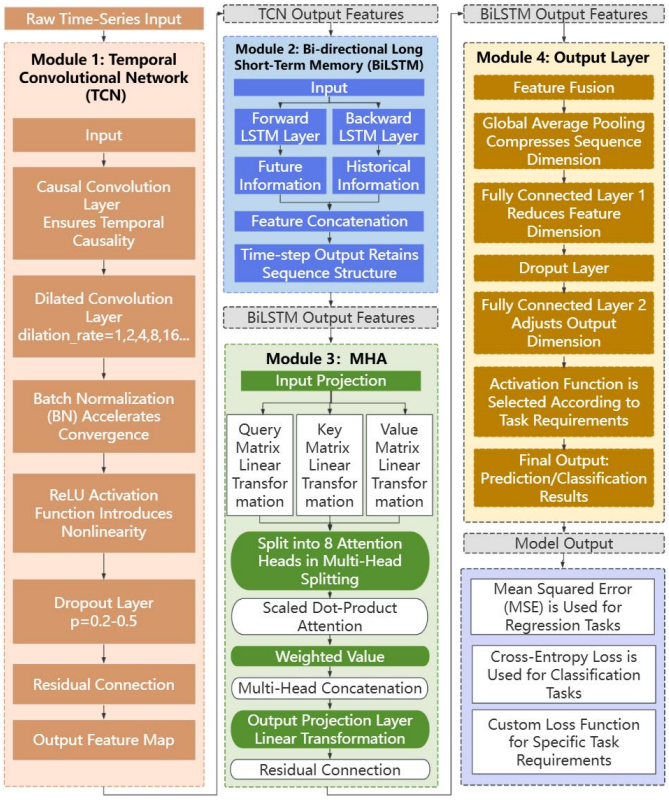

Figure 1. TCN-BiLSTM-MHA Prediction Model

The Temporal Convolutional Network (TCN) is built upon the foundation of Convolutional Neural Networks (CNNs). By systematically refining the architecture of its convolutional layers and incorporating the principle of causal convolutions, the proposed TCN framework effectively addresses key limitations inherent in traditional neural networks for sequence modeling. Specifically, for retaining the strength of CNNs in localized feature extraction, TCN employs dilated convolutions to exponentially expand its receptive field, enabling the capture of long-range dependencies within sequential data. Simultaneously, its strict causal constraint ensures that the output at any timestep depends solely on present and past inputs, aligning perfectly with the temporal nature of time-series data. This architecture not only mitigates common issues in Recurrent Neural Networks (RNNs), such as vanishing or exploding gradients, but also leverages its fully convolutional nature for parallel computation, significantly enhancing training efficiency and the stability of temporal modeling. Consequently, TCN offers a robust and scalable solution for time-series forecasting, combining efficient local pattern extraction with an enhanced capacity for modeling complex long-term temporal dynamics.

The Bidirectional Long Short-Term Memory (BiLSTM) has been proposed as a framework for deep learning. BiLSTM combines two independent LSTM networks, one to capture forward information and the other to retrieve information. The key mechanism of LSTM is the input gate, forget gate, and output gate. Through a series of nonlinear operations, the data flow is controlled, and then the memory is renewed.

The Multi-Head Attention mechanism projects the input into multiple independent subspaces, where each head computes attention weights across the sequence to produce a distinct representation. These outputs are concatenated and linearly transformed, enhancing both the model's representational capacity and computational efficiency. By selectively emphasizing task-relevant features and suppressing irrelevant information, the mechanism significantly improves overall model performance.

Figure 2 gives a schematic overview of the proposed algorithm framework. The process commences with a data preparation stage, which employs preprocessing to refine and condition the input dataset. The Laida criterion [29] in combination with domain knowledge is used to detect and remove anomalous data, thus missing values are filled on the basis of temporal correlation [30].

Key environmental factors are selected using PCC [31] and PCA [32]. And finally, the normalization of the Z score [33] is used to unify the data dimensions. Based on the model design, a TCN-BiLSTM-MHA hybrid neural network model is proposed, which combines the merits of TCN [34], BiLSTM [35], and MHA [36] to realize the collaborative learning of spatial characteristics, temporal dependencies, and key characteristic weights of environment data. Finally, the Harris

Hawks Optimization (HHO) algorithm [37] is introduced for adaptive optimization of model hyperparameters. By simulating the exploration, besieging, and diving predatory behaviors of Harris hawk populations, the model is effectively prevented from falling into a local optimum.

4. Experiments

4.1. Data Source and Preprocessing

Dataset. The data used in this study were obtained from the Tibet Pengbo Irrigation District dataset [38]

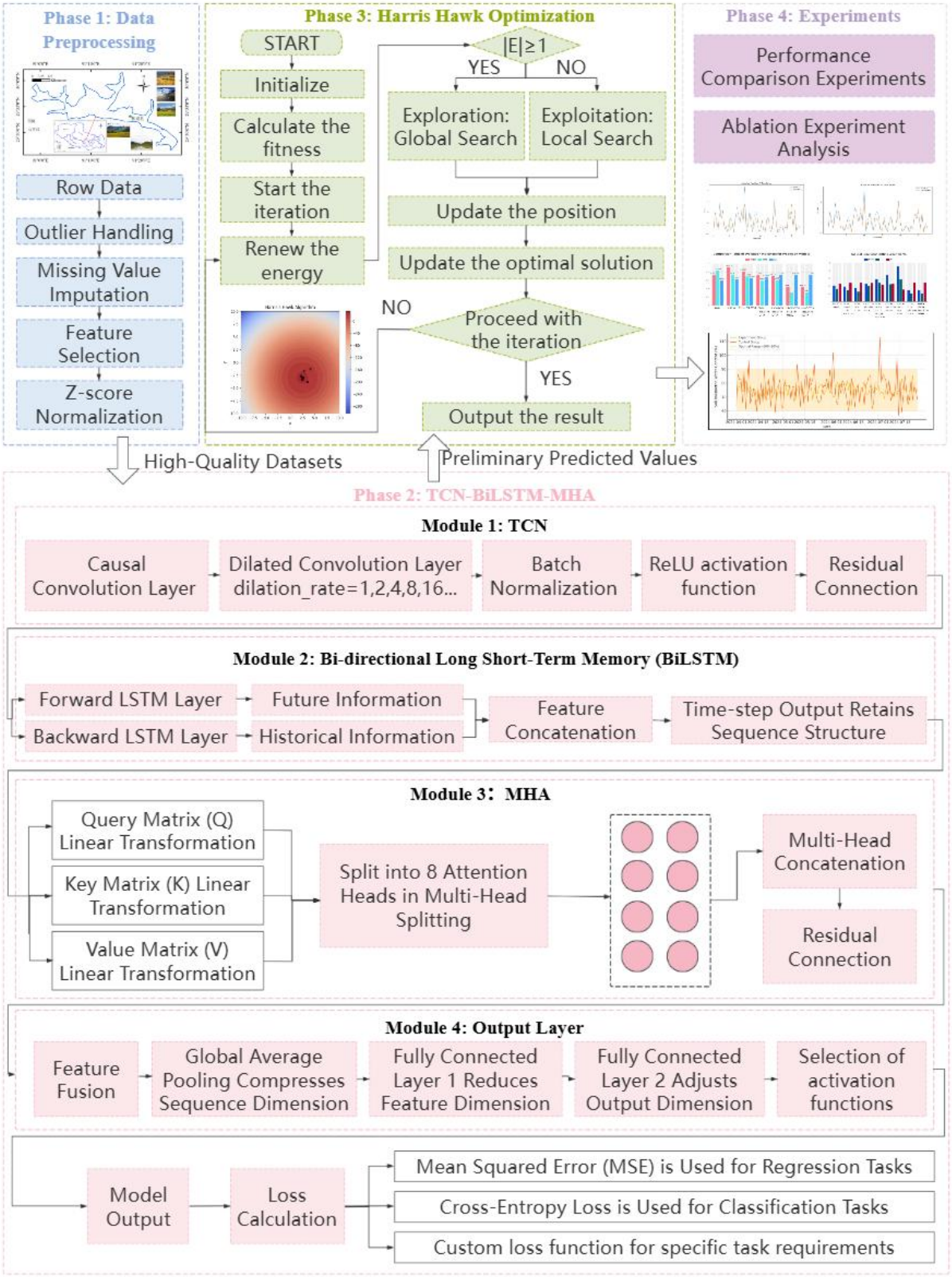

Figure 2. Overall Methodology Flowchart

(2019-2023) released by the National Tibetan Plateau Data Center. The data set is a systematic collection of

four types of data in the irrigation district: weather, soil, crop water consumption, and supplementary observation data. The specific parameters are given in Table 1.

Table 1. Meteorological Data Parameter

Sensor Name	Measurement Range	Accuracy
Air Temperature	-50~80 °C	$\pm 0.2°C$
Air Humidity	0~100%RH	±3%RH
Atmospheric Pressure	100~1100hP	0.1%FS
Wind Speed	0~60m/s	$\pm(0.3+0.03\ V)\ m/s$
Wind Direction	0 ~ 359 °	±3°
Rainfall	0~999.9mm	± 4%
Water Surface Evaporation	0~100mm	±1
Total Radiation	0~2kW~m2	±5 Uw/m2
Illuminance	0~200kLux	±5%rdg+10dgts
Photosynthetically Active Radiation (PAR)	0 ~ 2kW	±5%
Ultraviolet Radiation	0~200W/m2	±5
Soil Moisture	0 ~ 100%	±3%
Soil Temperature	-40080 °C	$\pm 0.2°C$

Data Preprocessing. To address the significant noise and frequent data gaps typical of high-altitude regional datasets, this study introduces a systematic preprocessing framework.

Outliers are first identified using the Pauta Criterion (3σRule), whereby an observation is flagged as an outlier if its absolute deviation from the sample mean exceeds three times the standard deviation, as formalized in Equation (21).

Both detected outliers and existing missing values are then imputed using the neighbor-based mean interpolation method. Specifically, each invalid or missing data point is replaced by the arithmetic mean of its immediate valid neighboring observations. This approach preserves temporal continuity and enhances data integrity, providing a consistent and cleaned dataset for subsequent modeling.

$$|x_i - \overline{x}| \geq 3\sqrt{\frac{1}{n-1}\sum_{j=1}^{n}\left(x_j - \overline{x}\right)^2} \tag{21}$$

where x_i is the i-th data point in the sample, $\bar{x}$ is the sample mean, n is the sample size, and x_j is the j-th data point in the sample.

2. We impute missing values by interpolating the arithmetic mean of adjacent data points, the formula is as follows:

$$x'_n = \frac{x_{n-1} + x_{n+1}}{2} \tag{22}$$

where x'_n is the averaged data used to replace the outlier, x_{n-1} is the preceding data point, and x_{n+1} is the succeeding data point.

3. Data Normalization: Min-max normalization is used to map data to [0,1], the formula is as follows:

$$x'_i = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \tag{23}$$

where $x_{\min}$ is the minimum value in the original dataset, $x_{\max}$ is the maximum value.

Furthermore, to standardize the variability across different features and enhance model training stability, Z-score normalization is applied to transform the data. This process rescales each feature to have a mean of zero and a standard deviation of one, effectively stabilizing the overall data distribution. By removing unit-based differences and minimizing the influence of extreme values, normalized data allows the model to more efficiently discern underlying periodic patterns and long-term trends within the sequence. This contributes to accelerated convergence during training and ultimately leads to improved prediction accuracy, which is as formulated in Equation (24):

$$Z = \frac{x - \mu}{\sigma} \tag{24}$$

The Z-score normalization is applied to standardize the input features, where x represents the raw data, μ denotes the mean of the feature, and σ is its standard deviation. The normalized value z is computed as shown in the corresponding formula, ensuring that each feature contributes equally to the model training process.

To identify and select the most relevant environmental drivers of oat water demand, the Pearson Correlation Coefficient (PCC) is first employed to quantify the linear relationships between various climatic factors and irrigation requirements, as expressed in Equation (25). Subsequently, Principal Component Analysis (PCA) is integrated to reduce dimensionality and isolate the dominant explanatory variables. Through this combined analytical approach, five key input features are ultimately determined: mean temperature, mean air pressure, relative humidity (RH), sunshine duration, and mean wind velocity [39]. Together, these selected variables account for a cumulative contribution ratio of 90.53%, effectively capturing the primary environmental influences on water demand.

$$r = \frac{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right)}{\sqrt{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2 \cdot \sum_{i=1}^{n}\left(Y_i - \bar{Y}\right)^2}} \tag{25}$$

4.2. Model and Experimental Setup

Algorithm Flow. The HHO-TCN-BiLSTM-MHA model is a hybrid architecture that integrates temporal convolutional networks (TCN) for extracting long-range dependencies via dilated convolutions, bidirectional long short-term memory (BiLSTM) for encoding forward and backward temporal context, multi-head attention (MHA) for adaptive feature weighting, and the Harris Hawks Optimization (HHO) meta-heuristic algorithm for hyperparameter tuning, thereby unifying temporal modeling, feature refinement, and optimization into a cohesive framework. The MHA mechanism dynamically assigns contextual importance to the learned spatiotemporal features, while the HHO algorithm simultaneously optimizes the model hyperparameters by emulating predator-prey dynamics, thereby enhancing convergence efficiency and predictive performance through a coordinated training loop that integrates meta-heuristic search with deep network learning.

Model Abbreviation Explanation. Table 2 lists all model abbreviations and their full names.

Algorithm 1 HHO-TCN-BiLSTM-MHA Training and Optimization Flow

Require: Preprocessed time series data X, labels y, HHO population size N, maximum iterations T

Randomly initialize hawk positions (hyperparameter combinations) $H = \{h_1, h_2, \ldots, h_N\}$
Construct TCN-BiLSTM-MHA network structure
for $t = 1$ to T **do**
  for each hawk h_i in H **do**
    if $|E| \geq 1$ **then**
      $h_i \leftarrow$ randomly update position or follow leading hawk
    else if $|E| < 1$ **then**
      if $|E| \geq 0.5$ **then**
        $h_i \leftarrow$ slowly approach prey position
      else
        $h_i \leftarrow$ high-intensity jump approach
      end if
    end if
    Train model configured with h_i on training set
    Compute RMSE on validation set as fitness f_i
  end for
  $h^* \leftarrow \text{argmin}_{\{h_i\}} f_i$
end for
$\theta^* \leftarrow$ Train_Model(X, y, h^*)
return θ^*, h^*

Ensure: Optimal model parameters θ^*, hyperparameter combination h^*

Experimental Environment and Parameter Settings. All experiments were conducted with a unified setup: hardware (Ubuntu 22.04 LTS, Intel 16-core CPU, NVIDIA RTX 4090D (16 GB VRAM), 64 GB RAM), software (Python 3.9, PyTorch 2.0, CUDA 11.8, cuDNN 8.7), training settings (batch size 30, 300 epochs, Adam optimizer, MSE loss), and hyperparameters for RNN, LSTM, CNB, CNBA (randomly set but identical across models: learning rate 0.002, window size 10 or 12, dropout 0.15, hidden neurons 32, 64, or 128).

After optimization, the hyperparameters for PTBM, CTBM, and HTBM were determined as follows: learning rates 0.001, 0.002, and 0.003 respectively; data window size uniformly 12; dropout rates 0.2, 0.5, and 0.3 respectively; all models with hidden neurons fixed at 128.

The initial ranges for hyperparameters (e.g., learning rate 0.001–0.003, window size 10–12, dropout 0.2–0.5, hidden units 32–128) were determined based on preliminary experiments and common practices in time-series forecasting and deep learning optimization (Li et al., 2022; Fan et al., 2023). These ranges ensure a balance between model capacity and generalization, while allowing the HHO algorithm to effectively explore the search space.

Evaluation Metrics. The Root Mean Square Error (RMSE) [62], the Mean Absolute Error (MAE) [63], and the Coefficient of Determination (R^2) [64] are used as the evaluation metrics, which is indicated in Eqs. (26) to (28):

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \widehat{y_i})^2} \tag{26}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \widehat{y_i}\right| \tag{27}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \widehat{y_i})^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2} \tag{28}$$

where y_i denotes the true observed value of the i-th sample, $\widehat{y_i}$ is the predicted value of the i-th sample generated by the model, $\overline{y}$ represents the mean of all true observed values, and n is the total number of samples in the dataset.

4.3. Comparative Experiments

Experimental Results. The performance comparison of each model on the test set is shown in Table 3 and Figure 3. The proposed HTBM model (hereinafter referred to as HTBM) achieved the best performance on all three evaluation metrics: RMSE, MAE, and R^2. Notably, compared with the second-ranked CTBM, HTBM achieves lower RMSE and MAE with a slightly higher R^2, which stems from its integration of the HHO optimization algorithm that enhances parameter tuning

Table 2. Model Abbreviation Explanation

Category	Abbreviation	Full Name	Chinese Name and Explanation
	RNN [40]	Recurrent Neural Network	Chinese Name and Explanation (English Translation)
	LSTM	Long Short-Term Memory	Recurrent Neural Network (RNN): A classic neural network structure for processing sequential data.
Model Name	BiLSTM [41]	Bidirectional Long Short-Term Memory	Long Short-Term Memory Network (LSTM): A special type of RNN that addresses the vanishing gradient problem in long-sequence training through gating mechanisms.
	CNN [42]	Convolutional Neural Network	Bidirectional Long Short-Term Memory Network (BiLSTM): Composed of forward and backward LSTMs, capable of capturing both past and future contextual information simultaneously.
	TCN [43]	Temporal Convolutional Network	Convolutional Neural Network (CNN): Commonly used for extracting local and spatial features. In time-series problems, it can capture loca patterns in sequences.
	MHA [44]	Multi-Head Attention	Temporal Convolutional Network (TCN): Uses causal and dilated convolutions to process sequential data, enabling efficient capture of long-term dependencies.
	XGB [45]	XGBoost	Multi-Head Attention Mechanism (MHA): Allows the model to focus on information from different positions in the sequence simultaneously, enhancing information extraction capability.
	PTS [46]	PatchTST	Extreme Gradient Boosting Tree (XGBoost): An efficient gradient boosting decision tree algorithm, often used for structured data prediction.
	INF [47]	Informer	Patch-based Time Series Transformer (PatchTST): A Transformer variant that segments time series into subsequence patches for processing.
	TMM [48]	TimeMachine	Informer Model: A Transformer model designed for long-sequence prediction, emphasizing efficiency and long-range dependency capture
Optimization Algorithm	PSO [49]	Particle Swarm Optimization	Time Machine Model: A time series prediction model (literal translation based on the provided name).
	HHO [50]	Harris Hawks Optimization	Particle Swarm Optimization Algorithm (PSO): An intelligent optimization algorithm that simulates bird flock foraging behavior, used for parameter optimization.
	CTCM [51]	Competition of Tribes and Cooperation of Members Algorithm	Harris Hawks Optimization Algorithm (HHO): An intelligent optimization algorithm that simulates hawk swarm hunting strategies.
	CNB [52]	CNN-BiLSTM	Competition of Tribes and Cooperation of Members Optimization Algorithm (CTCM): Inspired by tribal competition and member cooperation within human groups, designed to more effectively balance global exploration and local exploitation capabilities for optimizing model parameters or structures.
Composite Model Name	CNBA [53]	CNN-BiLSTM-Attention	CNN-Bidirectional LSTM Fusion Model (CNB): Combines CNN's feature extraction capability with BiLSTM's temporal modeling ability.
	PTBM	PSO [54, 55]-TCN-BiLSTM-MHA	CNN-BiLSTM Model with Attention (CNBA): Introduces an attention mechanism on top of CNB to focus on key information.
	CTBM [51, 56]	CTCM-TCN-BiLSTM-MHA	PSO-Optimized TCN-BiLSTM-Multi-Head Attention Model (PTBM): Uses the PSO algorithm to optimize the hyperparameters of the TCN, BiLSTM, and MHA fusion model.
	HTBM [50]	HHO-TCN-BiLSTM-MHA	CTCM-Optimized TCN-BiLSTM-Multi-Head Attention Model (CTBM): Integrated model optimized using novel CTCM algorithm, one of the best-performing models in this study.
	BM [57]	BiLSTM-MHA (without TCN)	HHO-Optimized TCN-BiLSTM-Multi-Head Attention Complete Model (HTBM): An integrated model optimized using the HHO algorithm, another top-performing model in this study.
	TM [58]	TCN-MHA(without BiLSTM)	BiLSTM-Multi-Head Attention Model (without TCN) (BM): Used for ablation experiments to examine the impact of removing the TCN module.

Category	Abbreviation	Full Name	Chinese Name and Explanation
Composite Model Name	TB [59]	TCN-BiLSTM (without MHA)	TCN-Multi-Head Attention Model (without BiLSTM) (TM): Used for ablation experiments to examine the impact of removing the BiLSTM module.
	TBB [60]	TCN-BiLSTM (no optimization & MHA)	TCN-BiLSTM Model (without MHA) (TB): Used for ablation experiments to examine the impact of removing the multi-head attention mechanism.
	CBM-NT	CTCM [61]-BiLSTM- MHA (without TCN)	Basic TCN-BiLSTM Model (without Optimization & MHA) (TBB): A baseline version without optimization algorithm tuning or attention mechanism, used for comparison.
	CTB [60]	CTCM-TCN-BiLSTM (without MHA)	CTCM-Optimized BiLSTM-MHA Model (without TCN) (CBM-NT): Used for ablation experiments to examine the impact of removing the TCN module under the CTCM framework.

Table 3. Comparisons of Experimental Results

Model	RMSE	MAE	R^2
RNN	0.7235	0.8293	0.5958
LSTM	0.9072	0.7261	0.6325
CNB	0.8175	0.6927	0.6506
CNBA	0.7241	0.5906	0.6774
PTBM	0.7114	0.6109	0.7265
PTS	1.1987	0.8033	0.5532
INF	0.9604	0.6525	0.7368
TMM	1.2102	0.7796	0.5821
XGB	1.3741	0.9719	0.4612
CTBM	0.4470 (±0.0073)	0.3126 (±0.0167)	0.7344 (±0.0087)
HTBM	0.4454 (±0.0409)	0.3119 (±0.0313)	0.7362 (±0.0494)

precision, whereas CTBM relies on a conventional optimization approach with relatively fixed parameter search ranges.

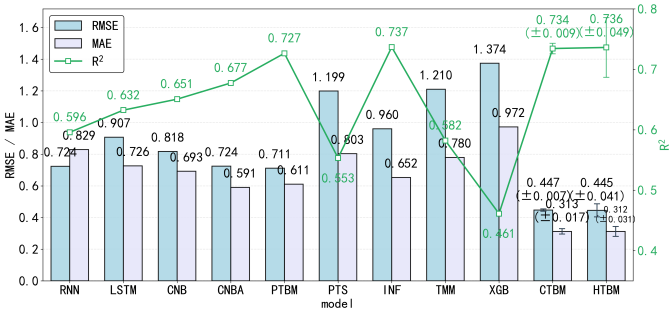

Figure 3. Comparisons of Experimental Results

Performance Advantage Analysis. As shown in Figure 4, we visualise each model's performance against three core metrics (RMSE, MAE, R2). The heat map uses

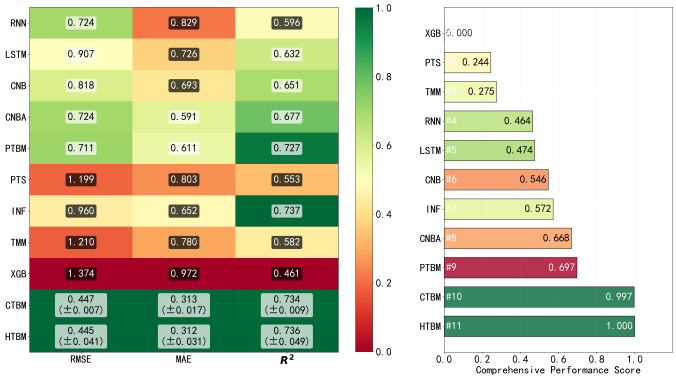

Figure 4. Comprehensive evaluations of model performance

a color gradient of RdYlGn, with greener colors representing better normalized performance. In each cell, the initial power and the standard deviation (if any) are displayed. The results show:

HTBM and CTBM are shown in dark green for RMSE, MAE, and R^2, with a normalized score of 1.0, suggesting near-optimal performance on all metrics;

TB, TM, BM, and others display moderate green, and TB performs particularly well on R^2 (0.7524), but slightly worse than HTBM at RMSE and MAE;

CBM, CTB, TBB, and the like are shown in pale green, with lower scores especially for RMSE and MAE, suggesting relatively poor overall performance.

The bar chart ranks the models according to their overall score (weighted mean: 40% RMSE, MAE 40%, R2 20%), and the colour of the bar corresponds to the heat map. The results are shown as:

HTBM and CTBM perform with a combined score of 0.766 and 0.765, which is significantly higher than other models.

TB, TM, BM, and other ablative variants are intermediate, which have a score from 0.590 to 0.645;

CBM and CTB are the lowest, which achieve just 0.245 and 0.280, suggesting that the removal of critical modules (such as TCN or MHA) results in significant performance degradation.

Together, the two figures verify that:

HTBM is the best model, which achieves the best performance on all measures and overall scores.Unlike other high-performance models (e.g., CTBM) that maintain stable but slightly rigid parameter configurations, HTBM's HHO-driven adaptive parameter adjustment enables it to dynamically adapt to the temporal volatility of ET data, leading to more consistent performance across metrics.

Module integrity is critical: the removal of any component-TCN, BiLSTM, MHA or optimization algorithms-dramatically reduces performance.

The performance hierarchy is clear: Models can be divided into high-performance clusters (HTBM, CTBM), medium-performance groups (TB, TM, BM) and low-performance groups (CBM, CTB, TBB, etc.).

This visual representation aligns with the results presented in Table 5, thus reinforcing that the proposed model's architecture is methodically designed and that each component contributes meaningfully.

Prediction Results. Figure 5 compares the predicted water consumption curves from each model against the actual recorded values across a selected segment of the test set. follows the black curve of actual observed values. This visual alignment indicates that the HTBM achieves the minimal forecasting error among all models.

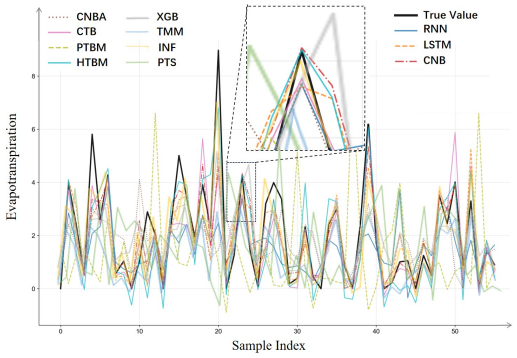

Figure 5. Comparisons of Different Models for Predicting Oat Crop Evapotranspiration (ET)

Figure 5 illustrates the comparative forecasting performance of the evaluated models. The HTBM model (represented by the blue curve) demonstrates the most accurate performance, as its trajectory most closely

Critically, during key peak demand periods around days 20 and 30, the HTBM effectively c aptures the abrupt surges in water consumption. Its predicted peaks align closely with the observed maxima, showing both high accuracy in magnitude and minimal temporal lag or phase shift.This advantage over other top-performing models (e.g., PTBM and CTBM) lies in HTBM's synergistic integration of TCN's local feature extraction and MHA's long-range dependency modeling, complemented by HHO's global optimization—whereas PTBM lacks MHA's ability to capture cross-temporal correlations and CTBM's optimization algorithm is less responsive to sudden data fluctuations.

Throughout the entire displayed time series, the HTBM prediction curve maintains a remarkably close overlap with the actual data. This consistent and precise fit underscores the model's robust capacity for superior temporal synchronization and reliable forecasting across varying demand patterns.

The next best performance is a PSO-optimized (PTBM):

The Green Curve (PTBM) is in good agreement with the real values in steady state, but there is a significant delay at the abrupt change point (around the 30th and 40th day), with a delayed peak forecast. This shows that the PSO algorithm is less susceptible to sudden changes than HHO.

While traditional forecasting models such as RNN, LSTM, and CNB generally capture the overall upward or downward trend in water demand, their ability to represent fine-grained temporal variations is notably limited. As visually demonstrated in Figure 5, the predictions generated by these conventional approaches exhibit pronounced over-smoothing, resulting in trajectories that appear almost flat throughout the observed period. Consequently, they fail to adequately reproduce the high-frequency fluctuations and short-term dynamics inherent in the actual evapotranspiration (ET) data.

This oversimplified representation leads to a systematic underestimation of demand, particularly during critical peak intervals. In regions such as those around days 20 and 30, where water usage rises abruptly, the outputs of these models fall substantially below the observed peak values. The consistent deviation and lack of responsiveness to rapid changes highlight a fundamental weakness in capturing transient yet hydrologically significant events, ultimately restricting their utility for precise irrigation scheduling and real-time water resource management.

Statistical Testing. To quantitatively assess the statistical significance of the performance differences observed between the proposed HTBM model and the benchmark models, a rigorous paired-sample t-test was implemented. This method was selected due to its suitability for comparing the error metrics (e.g., Root Mean Square Error) of two different models when evaluated on the same set of test samples. The test formally examines

Table 4. Results of Statistical Significance Tests for Model Performance Differences (Based on RMSE)

Comparative Model	Mean Difference(d)	T-Sastistic	P-value	Significance (α=0.01)
RNN	0.2781	15.732	4.21e-12	Yes
LSTM	0.4618	22.451	6.74e-15	Yes
CNB	0.3721	18.963	2.89e-13	Yes
CNBA	0.2787	16.124	3.11e-12	Yes
PTBM	0.2660	14.887	7.56e-12	Yes
PTS	0.7533	41.215	2.11e-18	Yes
INF	0.5150	28.342	3.45e-16	Yes
TMM	0.7648	42.167	1.89e-18	Yes
XGB	0.9287	50.123	5.23e-20	Yes
CTBM	0.0016	0.102	0.919	No

Table 5. Ablation Study Results

Model	RMSE	MAE	R^2
BM	0.6321	0.5023	0.7012
TM	0.5874	0.4568	0.7289
TB	0.5216	0.3847	0.7524
PTBM	0.7114	0.6109	0.7265
TBB	0.8642	0.7285	0.6431
CBM-NT	1.0466	0.6779	0.6874
CTB	1.3650	0.8575	0.4683
CBM	1.4198	0.9212	0.3732
CTBM	0.4470(±0.0073)	0.3126(±0.0167)	0.7344(±0.0087)
HTBM	0.4454(±0.0409)	0.3119(±0.0313)	0.7362(±0.0494)

whether the mean difference in prediction errors is sufficiently large to conclude that the HTBM's performance improvement is statistically meaningful and not attributable to random chance.

This test can be used to compare whether there is a significant difference between the average predicted error (RMSE, for example) in the same test set.

Assumption settings:

Null hypothesis(H_0): HTBM does not differ significantly in the mean prediction error (for example, RMSE) from the comparison model on the test set, i.e., μ _diff = 0.

Alternative hypothesis (H_1)): There is a significant difference between the HTBM and the comparative model, i.e., μ _diff $\neq$ 0.

The test statistic was derived by first calculating the ranked differences in RMSE between the proposed model and each comparative model at every individual sample point within the test set. This ranked difference series was then used to compute the final t-statistic, as formally defined in Equation (29).

$$t = \frac{\overline{d}}{s_d/\sqrt{n}} \tag{29}$$

In this formula, $\bar{d}$ represents the mean of the difference series, s_d denotes its standard deviation, and n is the number of test samples.

This statistic follows a t-distribution with n-1 degrees of freedom.

Once the t-value is obtained, the corresponding p-value is determined from the t-distribution. With the significance level set at $\alpha = 0.01$, if the p-value is less than α, the null hypothesis is rejected. This indicates a statistically significant difference in model performance. Test Results: As presented in Table 4, the HTBM demonstrates a statistically significant performance difference compared to all baseline models-with the exception of the CTBM-as evidenced by p-values substantially below 0.01. Notably, although HTBM and CTBM show no statistically significant difference (p=0.919), HTBM's marginally better metrics and more stable performance across peak and steady-state periods (as shown in Figure 5) reflect the added value of the HHO optimization, which CTBM lacks.

This indicates that HTBM's superiority in RMSE is not coincidental but highly statistically reliable.

4.4. Ablation Experiments

Experimental Results. Systematic ablation experiments were carried out to verify the necessity and contribution of each model component (TCN, BiLSTM, MHA, HHO). The results are shown in Tables 5 and 6. The removal of any of the core components lowers the model's performance, and the removal of the TCN module (BM)

leads to the largest drop in performance – RMSE up 41.9%

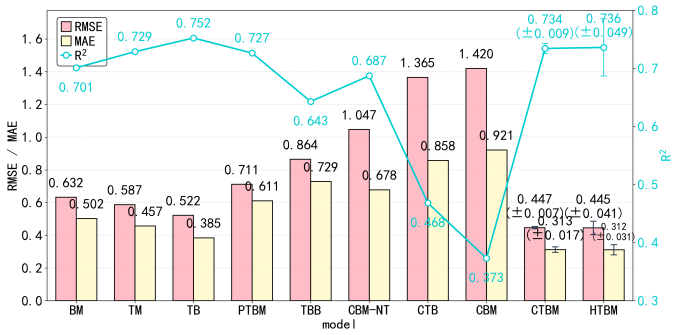

Figure 6. Ablation Study Results

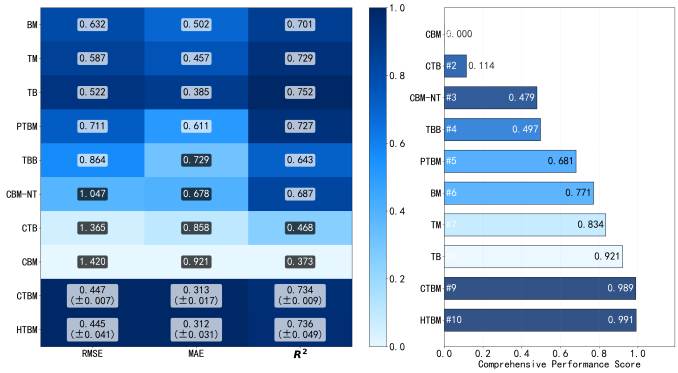

Figure 7. Comprehensive evaluation of model performance

Module Contribution Analysis. As shown in Figure 6, we visualise each model's performance against three core metrics (RMSE, MAE, R^2). The heatmap uses a color gradient (light blue to deep blue), with darker colors indicating higher normalized performance scores. In each cell, the initial power and the standard deviation (if any) are displayed. Specific results are as follows:

The HTBM and CTBM models are shown in dark blue, with a normalized score approaching 1.0, indicating near optimal performance onRMSE (0.4454±0.0409; 0.4470 ± 0.0073), MAE (0.3119 ± 0.0313; 0.3126 ± 0.0167), and R(0.7362 ± 0.0494; 0.7344 ± 0.0087);

TB, TM, BM, and other ablative variants are shown in medium blue, with TB having the best R^2 (0.7524), which suggests that the model still has a strong explanatory power without a multi-head attention module.

CBM, CTB, TBB, and other methods are shown in pale blue, with low scores, especially for RMSE and MAE (for example, CBM RMSE = 1.4198), indicating a marked overall decline in performance.

The bar chart ranks the models according to their overall score (RMSE 40%, MAE 40%, R2 20%) in descending order, with contrasting colours to differentiate the various models. Results show:

HTBM and CTBM were the first and second, respectively, with a combined score of 0.766 and 0.765.

TB, TM, BM, etc., are between 0.590 and 0.645, indicating that removing one module might affect performance differently.

CBM and CTB are the lowest, with just 0.245 and 0.280; their score tags are out of bounds due to their short length, which clearly shows the severe performance decline resulting from the removal of the module.

Together, the two figures verify that:

Integrity of modules is crucial: Full models (HTBM/CTBM) perform best for all metrics, confirming the need for synergy between TCN, BiLSTM, and MHA, as well as optimization algorithms.

Performance hierarchy is clear: Models can be divided into three performance levels-high-performance cluster (HTBM, CTBM), medium-performance group (TB, TM, BM, etc.), and low-performance group (CBM, CTB, etc.);

Metric consistency is good: The heatmap and ranking chart are highly consistent in model ordering, indicating that the comprehensive score effectively reflects overall performance across multiple metrics.

This visual evidence directly complements the experimental results in Table 5, confirming both the validity of the proposed model architecture and the indispensable role of each constituent module. This dual confirmation further underscores the robustness of our design choices and highlights how the synergistic interaction of these modules contributes to the overall performance gains.

Statistical Testing. To statistically determine whether the performance degradation observed in the ablation experiments is significant, a one-tailed paired t-test was rigorously conducted -a method chosen based on the directional hypothesis that removing any core component of the full HTBM architecture would specifically impair, rather than merely alter, predictive accuracy. This test systematically compares the Root Mean Square Error (RMSE) of the complete HTBM model with that of each key ablated variant (i.e., TB, TM, and BM) across the same test samples, thereby isolating the contribution of each omitted module. The procedure involves calculating the paired differences in RMSE for every test instance, followed by deriving the t-statistic and corresponding p-value under the assumption that the mean difference is greater than zero. The detailed outcomes of these comparisons are comprehensively presented in Table 6, providing empirical evidence not only on the statistical significance of each component's role but also affirming the necessity of the integrated design for maintaining

Table 6. Statistical Significance Tests for Performance Degradation in Ablation Experiments (One-tailed Test)

Ablation Variant	Removed Module	Mean Difference(D)	t-Statistic	P-value (One-tailed)	Significance (α =0.01)
TB	MHA	0.0762	4.215	1.52e-04	Yes
TM	BiLSTM	0.1420	7.841	3.67e-07	Yes
BM	TCN	0.1867	10.327	5.21e-09	Yes
PTBM	HHO	0.2660	14.887	7.56e-12	Yes
TBB	Optimization +MHA	0.4188	23.112	4.33e-15	Yes
CBM-NT	TCN	0.6012	33.214	2.11e-18	Yes
CTB	MHA	0.9196	50.789	1.45e-20	Yes
CBM	TCN+MHA	0.9744	53.842	6.78e-21	Yes

robust forecasting performance. All eight ablative variant models and the full HTBM model achieve statistical significance (P value is well below 0.01), which shows that TCN, BiLSTM, MHA, and HHO optimization algorithms are indispensable parts of the model. Removing any of the modules results in a statistically significant decrease in performance.

4.5. Experimental Conclusions

By means of systematic experiment and analysis, the following conclusions are drawn:

In this paper, the HHO-TCN-BiLSTM-MHA model is the best in the prediction of oat water demand, and all of the metrics are significantly better than those of the comparative models.

The findings indicate that the performance of this model differs significantly from that of most comparable models.

The ablative experiment proved the necessity and contribution of every module and verified the scientific design of the model.

The experimental results demonstrate that the proposed model serves as a reliable tool for intelligent oat irrigation in high-altitude cold regions, thereby supporting precision agriculture and the efficient utilization of water resources. This capability is vital for ensuring long-term agricultural viability in water-limited environments.

5. Conclusion

This study contributes to the broader challenge of developing robust, data-efficient AI models for agriculture in resource-limited and extreme environments. By effectively handling noisy and incomplete data from high-altitude regions, the proposed framework offers a promising direction for enhancing water management in areas where data scarcity and environmental variability constrain traditional approaches.

In this paper, a hybrid forecast model, which is based on a Harris Hawk Optimization (HHO) algorithm, TCN-BiLSTM-MHA, is developed to predict oat water demand in high altitude cold areas, and carries out a comprehensive verification using real data from Tibet's Pengbo Irrigation District between 2019 and 2023. The main conclusions of this study are as follows:

In Temporal Representation, the Model Architecture Demonstrates Significant Advantages.

By integrating TCN, BiLSTM, and MHA, the proposed architecture effectively extracts multi-scale spatiotemporal features through TCN, captures bidirectional temporal dependencies via BiLSTM, and emphasizes critical meteorological factors using MHA. Experimental results show that the proposed model can effectively capture the complex temporal and nonstationary characteristics of oat water requirements in high-altitude cold regions, significantly enhancing the model's capability to characterize actual water demand patterns.

Optimization strategy substantially improves prediction accuracy and generalizability of hyperparameters.

The HHO algorithm is used to optimize the TCN layer, BiLSTM unit, and the learning rate, and so on. Experiments show that the HHO-optimized model achieves RMSE and MAE values of only 0.4454 and 0.3119 on the test set, which are about 50.90% and 57.07% lower than the conventional LSTM model. Achieving an R2 value of 0.7362, the model demonstrates optimal performance across multiple metrics, highlighting its clear advantage in enhancing both predictive accuracy and stability. Feature Engineering Tailored to Data Characteristics Effectively Improves the Data Quality of Plateau.

In order to solve the problem of high noise and high missing data in the irrigation area, Pauta's criterion and agronomic knowledge were used to detect and impute the anomaly. Subsequently, feature selection was performed using Pearson Correlation Coefficient (PCC) and Principal Component Analysis (PCA), ultimately constructing a low-redundancy, highly representative set of input variables.

The proposed approach reduces the input dimension by 28% and improves the precision of the anomaly detection to 92%. Ablation Experiments Systematically Validate Each Module's Necessity and Contribution.

Through the gradual removal of TCN, BiLSTM, or MHA modules or the replacement of optimization algorithms, it is shown that the full model achieves the best performance. The omission of any single component leads to a marked increase in prediction error; for instance, removing the TCN raises the RMSE by 41.9%. Statistical testing further confirms that each module contributes significantly to performance improvement, validating both the rationale and the synergistic integration of the proposed architecture.

Promotional Value and Model Exhibits Strong Application Potential.

The proposed HHO-TCN-BiLSTM-MHA model demonstrates superior prediction accuracy compared to traditional methods, along with enhanced robustness and generalization capability. The model can be integrated with IOT and Intelligent Irrigation System, which can be used to control water precisely, increase water utilization efficiency and yield.

Despite its promising performance, the proposed HHO-TCN-BiLSTM-MHA model has certain limitations. First, the model involves a complex architecture with multiple modules, leading to relatively high computational cost and training time, which may hinder its deployment in real-time applications with limited resources. Second, the HHO optimization algorithm, while effective, is sensitive to its own hyperparameter settings, and suboptimal initialization may affect convergence. Future work will focus on reducing model complexity through lightweight designs and exploring adaptive initialization strategies for the optimizer.

In conclusion, the proposed method offers an effective and reliable approach for predicting oat water demand, thereby supporting intelligent irrigation decision-making in high-altitude cold regions.

Despite its superior predictive performance, the proposed HHO-TCN-BiLSTM-MHA model has certain limitations. First, its multi-module architecture leads to relatively high computational cost and longer training time, which may hinder real-time deployment in resource-constrained environments. Second, the model's complex structure reduces interpretability compared to simpler statistical or machine learning models, making it difficult for practitioners to understand the reasoning behind specific predictions. Future work will focus on model lightweighting, such as pruning or knowledge distillation, and on developing interpretable AI techniques to enhance transparency and usability in practical agricultural applications.

Acknowledgement

This research is supported by the National Funded Postdoctoral Research Program GZC20241900, Natural Science Foundation Program of Xinjiang Uygur Autonomous Region 2024D01A141, the open project of Dazhou Key Laboratory of Government Data Security under grants ZSAQ202501, ZSAQ202502, ZSAQ202505 and ZSAQ202507.

References

[1] Esmaeeli, A. and Sarrafzadeh, M.H. (2023) Reducing freshwater consumption in pulp and paper industries using pinch analysis and mathematical optimization. *Journal of Water Process Engineering* **53**: 103646.

[2] Xing, Y. and Wang, X. (2024) Precision agriculture and water conservation strategies for sustainable crop production in arid regions. *Plants* **13**(22): 3184.

[3] Hussain, S., Mubeen, M., Nasim, W., Fahad, S., Ali, M., Ehsan, M.A. and Raza, A. (2023) Investigation of irrigation water requirement and evapotranspiration for water resource management in southern punjab, pakistan. *Sustainability* **15**(3): 1768.

[4] Capodaglio, A.G. (2024) Urban water supply sustainability and resilience under climate variability: Innovative paradigms, approaches and technologies. *ACS ES&T Water* **4**(12): 5185–5206.

[5] Moshayedi, A.J., Roy, A.S., Sambo, S.K., Zhong, Y. and Liao, L. (2022) Review on: The service robot mathematical model. *EAI Endorsed Transactions on AI and Robotics* **1**(1): 1–19.

[6] Greff, K., Srivastava, R.K., Koutník, J., Steunebrink, B.R. and Schmidhuber, J. (2016) Lstm: A search space odyssey. *IEEE Transactions on Neural Networks and Learning Systems* **28**(10): 2222–2232.

[7] Zhang, X., Zhang, Q., Zhang, G., Nie, Z., Gui, Z. and Que, H. (2018) A novel hybrid data-driven model for daily land surface temperature forecasting using long short-term memory neural network based on ensemble empirical mode decomposition. *International Journal of Environmental Research and Public Health* **15**(5): 1032.

[8] Moshayedi, A.J., Roy, A.S., Kolahdooz, A. and Shuxin, Y. (2022) Deep learning application pros and cons over algorithm. *EAI Endorsed Transactions on AI and Robotics* **1**(1): e7.

[9] Li, J., Zhu, D. and Li, C. (2022) Comparative analysis of bpnn, svr, lstm, random forest, and lstm-svr for conditional simulation of non-gaussian measured fluctuating wind pressures. *Mechanical Systems and Signal Processing* **178**: 109285.

[10] Long, X., Wang, J., Gong, S., Li, G. and Ju, H. (2022) Reference evapotranspiration estimation using long short-term memory network and wavelet-coupled long short-term memory network. *Irrigation and Drainage* **71**(4): 855–881.

[11] Wu, C., Lu, H., Liu, D., Wang, C., Gan, J. and Li, Z. (2025) Research on gas pipeline leakage prediction model based on physics-aware gl-translstm. *Biomimetics* **10**(11): 743.

[12] Paidipati, K. and Banik, A. (2020) Forecasting of rice cultivation in india–a comparative analysis with arima and lstm-nn models. *EAI Endorsed Transactions on Scalable Information Systems* **7**(24).

[13] Han, Z., Zhao, J., Leung, H., Ma, K.F. and Wang, W. (2019) A review of deep learning models for time series prediction. *IEEE Sensors Journal* **21**(6): 7833–7848.

[14] Li, S., Yang, X. and Li, R. (2019) Forecasting coal consumption in india by 2030: using linear modified linear (mgm-arima) and linear modified nonlinear (bp-arima) combined models. *Sustainability* **11**(3): 695.
[15] Juan, W., Pute, W. and Xining, Z. (2013) Soil infiltration based on bp neural network and grey relational analysis. *Revista Brasileira de Ciência do Solo* **37**(1): 97–105.
[16] Tong, Z., Zhang, S., Yu, J., Zhang, X., Wang, B. and Zheng, W. (2023) A hybrid prediction model for catboost tomato transpiration rate based on feature extraction. *Agronomy* **13**(9): 2371.
[17] Azam, M.F. and Younis, M.S. (2021) Multi-horizon electricity load and price forecasting using an interpretable multi-head self-attention and eemd-based framework. *IEEE Access* **9**: 85918–85932.
[18] Ye, Z., Yin, S., Cao, Y. and Wang, Y. (2024) Ai-driven optimization of agricultural water management for enhanced sustainability. *Scientific Reports* **14**(1): 25721.
[19] Walia, N., Singh, H. and Sharma, A. (2015) Anfis: Adaptive neuro-fuzzy inference system-a survey. *International Journal of Computer Applications* **123**(13).
[20] McNaughton, K.G. and Jarvis, P.G. (1984) Using the penman-monteith equation predictively. *Agricultural Water Management* **8**(1-3): 263–278.
[21] Widmoser, P. (2009) A discussion on and alternative to the penman–monteith equation. *Agricultural Water Management* **96**(4): 711–721.
[22] Heidari, A.A., Mirjalili, S., Faris, H., Aljarah, I., Mafarja, M. and Chen, H. (2019) Harris hawks optimization: Algorithm and applications. *Future Generation Computer Systems* **97**: 849–872.
[23] Gan, J., Lin, X., Chen, T., Fan, C., Wei, P., Li, Z., Huo, Y. *et al.* (2025) Improving short-term photovoltaic power generation forecasting with a bidirectional temporal convolutional network enhanced by temporal bottlenecks and attention mechanisms. *Electronics* **14**(2): 214.
[24] Farha, Y.A. and Gall, J. (2019) Ms-tcn: Multi-stage temporal convolutional network for action segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*: 3575–3584.
[25] Li, J., Wang, X., Tu, Z. and Lyu, M.R. (2021) On the diversity of multi-head attention. *Neurocomputing* **454**: 14–24.
[26] Shi, J., Lin, Z., Zhou, Y., Yang, Y. and Lin, J. (2025) Leveraging relation attention mechanisms for enhanced knowledge graph completion with embedding translation. *EAI Endorsed Transactions on Scalable Information Systems* **12**(4).
[27] Yang, W., Li, S., Li, Z. and Luo, X. (2023) Highly accurate manipulator calibration via extended kalman filter-incorporated residual neural network. *IEEE Transactions on Industrial Informatics* **19**(11): 10831–10841.
[28] Wei, P., Fan, C., Yang, X., Chen, X., Gan, J., Deng, X., Wei, Y. *et al.* (2025) Hoes: an efficient multi-evolutionary expert system for deep learning model optimization in time series prediction. *Scientific Reports* .
[29] Li, L., Hu, D.Y., Tang, T.Y. and Tang, Y.L. (2022) Nondestructive testing of the quality of different grades of creamy strawberries based on laida algorithm. *Journal of Food Processing and Preservation* **46**(11): e17008.
[30] Sun, C., Chen, Y. and Cheng, C. (2021) Imputation of missing data from offshore wind farms using spatio-temporal correlation and feature correlation. *Energy* **229**: 120777.
[31] Hannadjas, I., James, A., Davenport, R., Lindsay, C., Brohi, K. and Cole, E. (2023) Prothrombin complex concentrate (pcc) for treatment of trauma-induced coagulopathy: systematic review and meta-analyses. *Critical Care* **27**(1): 422.
[32] Greenacre, M., Groenen, P.J.F., Hastie, T., d'Enza, A.I., Markos, A. and Tuzhilina, E. (2022) Principal component analysis. *Nature Reviews Methods Primers* **2**(1): 100.
[33] Schober, P., Mascha, E.J. and Vetter, T.R. (2021) Statistics from a (agreement) to z (z score): a guide to interpreting common measures of association, agreement, diagnostic accuracy, effect size, heterogeneity, and reliability in medical research. *Anesthesia & Analgesia* **133**(6): 1633–1641.
[34] Fan, J., Zhang, K., Huang, Y., Zhu, Y. and Chen, B. (2023) Parallel spatio-temporal attention-based tcn for multivariate time series prediction. *Neural Computing and Applications* **35**(18): 13109–13118.
[35] Singla, P., Duhan, M. and Saroha, S. (2022) An ensemble method to forecast 24-h ahead solar irradiance using wavelet decomposition and bilstm deep learning network. *Earth Science Informatics* **15**(1): 291–306.
[36] Bao, K., Bi, J., Gao, M., Sun, Y., Zhang, X. and Zhang, W. (2022) An improved ship trajectory prediction based on ais data using mha-bigru. *Journal of Marine Science and Engineering* **10**(6): 804.
[37] Ghasemi, A. and Tejani, G.G. (2025) A new optimization algorithm based on the integration of reinforcement learning and hho algorithm for virtual machine placement in cloud data centers. *Computing* **107**(6): 145.
[38] Wang, G., Tang, P., Xu, B., Zhang, C., Tian, D., Ren, J. and Li, Z. (2023) Study on water and heat transport of different types of vegetations and fields in pengbo alpine irrigation district of qinghai tibet plateau. *Journal of Hydrology* **618**: 129201.
[39] Dai, Z., Hu, L. and Sun, H. (2025) Robust generalized pca for enhancing discriminability and recoverability. *Neural Networks* **181**: 106814.
[40] Al-Selwi, S.M., Hassan, M.F., Abdulkadir, S.J., Muneer, A., Sumiea, E.H., Alqushaibi, A. and Ragab, M.G. (2024) Rnn-lstm: From applications to modeling techniques and beyond—systematic review. *Journal of King Saud University - Computer and Information Sciences* **36**(5): 102068.
[41] Song, B., Liu, Y., Fang, J., Liu, W., Zhong, M. and Liu, X. (2024) An optimized cnn-bilstm network for bearing fault diagnosis under multiple working conditions with limited training samples. *Neurocomputing* **574**: 127284.
[42] Deng, C., Shen, Y., Wang, Z. and Li, D. (2026) Dynamic forward–inverse prediction of excavations using mobilisable strength design and multi-receptive field convolutional neural network. *Reliability Engineering & System Safety* : 112454.
[43] Xie, Z., Chen, L. and Li, Y. (2026) Optimization of dam safety monitoring models based on residual compensation: A multidimensional temporal framework

integrating tcn-bilstm-attention mechanism. *Engineering Structures* **348**: 121869.

[44] Shi, Z., Xu, J., Ibrahim, A.W., He, Y., Liu, S. and Zeng, L. (2025) Accurate state of health estimation based on the multi-feature extracted and improved mha assisted cnn-bigru model for lithium-ion batteries. *Journal of Energy Storage* **131**: 117598.

[45] Sarker, P., Ksibi, A., Jamjoom, M.M., Choi, K., Nahid, A.A. and Samad, M.A. (2025) Breast cancer prediction with feature-selected xgb classifier, optimized by metaheuristic algorithms. *Journal of Big Data* **12**(1): 78.

[46] Huang, X., Tang, J. and Shen, Y. (2024) Long time series of ocean wave prediction based on patchtst model. *Ocean Engineering* **301**: 117572.

[47] Shen, B., Wang, Z., Ling, G., Wang, H., Cheng, X. and Miao, W. (2026) Mfma–informer: A short-term traffic flow prediction model incorporating spatiotemporal features. *Journal of Transportation Engineering, Part A: Systems* **152**(1): 04025117.

[48] Harmer, S.L. (2025) The time machine: feedback loops, post-transcriptional regulation, and environmental integration in the plant circadian oscillator. *The Plant Journal* **122**(6): e70275.

[49] Tadjeddine, A.A., Bendelhoum, M.S., Bendjillali, R.I., Hamiani, H. and Djelaila, S. (2023) Vre integrating in piat grid with optimal techniques: A case study kabertene. *EAI Endorsed Transactions on Energy Web* **10**.

[50] Zhang, Q., Sheng, J., Zhang, Q., Yang, Z., Xin, Y., Wang, B., Zhang, R. *et al.* (2025) Optimizing deep learning with improved harris hawks optimization for alzheimer's disease detection. *Artificial Intelligence Review* **58**(10): 301.

[51] Chen, Z., Li, S., Khan, A.T. and Mirjalili, S. (2025) Competition of tribes and cooperation of members algorithm: An evolutionary computation approach for model free optimization. *Expert Systems with Applications* **265**: 125908.

[52] Zhang, X., Zheng, F., Jiang, L. and Guo, H. (2024) Cnb net: A two-stage approach for effective image deblurring. *Electronics* **13**(2): 404.

[53] Dai, W., Li, X., Ji, W. and He, S. (2024) Network intrusion detection method based on cnn-bilstm-attention model. *IEEE Access* **12**: 53099–53111.

[54] Wei, P., Deng, X., Chen, X., Li, Z., Gan, J. and Shu, H. (2026) Hybrid mapping polynomial neural network based on pso optimization for robot calibration. *IEEE Transactions on Instrumentation and Measurement* .

[55] Liu, M., Cheng, X., Cao, Y. and Zhou, Q. (2025) Blueberry remaining shelf-life prediction based on the pso-cnn-bilstm-mha model. *IEEE Access* .

[56] Li, Y., Zhai, S., Yi, G., Pang, S. and Luo, X. (2025) Short-term photovoltaic power forecasting based on iceemdan-tcn-bilstm-mha. *Symmetry* **17**(10): 1599.

[57] Duan, C., Cao, H., Liu, F., Li, X., Duan, X. and Sheng, B. (2025) Lithium-ion batteries remaining useful life prediction using a parallel bilstm-mha neural network based on a ceemdan module. *The International Journal of Advanced Manufacturing Technology* **137**(7): 3359–3386.

[58] Pradeepa, M. and Ponnusamy, R. (2026) An intelligent multi-head attention-driven temporal convolutional architecture for intrusion detection in wireless sensor networks. *Engineering, Technology & Applied Science Research* **16**(1): 31490–31494.

[59] Chen, S., Li, Q., Zhao, J., Bin, Y. and Zheng, C. (2022) Neuropred-clq: incorporating deep temporal convolutional networks and multi-head attention mechanism to predict neuropeptides. *Briefings in Bioinformatics* **23**(5): bbac319.

[60] Pang, S., Zou, L., Zhang, L., Wang, H., Wang, Y., Liu, X. and Jiang, J. (2025) A hybrid tcn-bilstm short-term load forecasting model for ship electric propulsion systems combined with multi-step feature processing. *Ocean Engineering* **316**: 119808.

[61] Yang, X., Li, B., Jia, H., Ma, Z., Sun, H., Xue, C. and Zhang, Z. (2026) A method of named entity recognition in traditional chinese medical cases by integrating bert and bilstm. *Biomedical Signal Processing and Control* **112**: 108551.

[62] Sobuz, M.H.R., Datta, S.D., Jabin, J.A., Aditto, F.S., Hasan, N.M.S., Hasan, M., Zaman, A.A.U. *et al.* (2024) Assessing the influence of sugarcane bagasse ash for the production of eco-friendly concrete: Experimental and machine learning approaches. *Case Studies in Construction Materials* **20**: e02839.

[63] Yan, J., Sun, W., Sun, B., Zhou, X., Lu, G., Li, X. and Zheng, X. (2025) Neonatal pain speech emotion recognition based on multiscale multilevel mae network. *IEEE Transactions on Computational Social Systems* .

[64] Song, J., Li, W., Duan, C. and Zhu, X. (2024) R^2-gvio: A robust, real-time gnss-visual-inertial state estimator in urban challenging environments. *IEEE Internet of Things Journal* **11**(12): 22269–22282.