

RESEARCH ARTICLE

Chinese Water Demand Forecast Based on iTransformer Model

ZHI-WEI TIAN AND RU-LIANG QIAN^{ID}

School of Mathematics and Statistics, Qinghai Minzu University, Xining, Qinghai 810007, China

Corresponding author: Ru-Liang Qian (nauri1@163.com)

ABSTRACT This paper presents a novel deep learning-based model for forecasting water demand. Specifically, a transformer network architecture-based iTransformer model is introduced to forecast total water demand at both country and province levels over the medium term. Comparative evaluations with Transformer, PatchTST, and LSTM models are conducted across various forecasting lengths, with hyperparameter optimization performed through grid search. The optimal model and parameters are then applied to historical water demand data from 2000 to 2023, yielding forecasts for subsequent years. Results demonstrate that the iTransformer model achieves the lowest RMSE (92.72/1.39/22.71/21.69/9.16), MAE (68.65/1.11/17.42/13.38/5.85), and MAPE (0.01/0.28/0.03/0.08/0.01) in forecasting water demand for China, Beijing, Jiangsu, Zhejiang, and Guangdong respectively. The study emphasizes the importance of considering population size and economic activity in managing socio-economic water demand in China, advocating for a balanced approach to water resource utilization. While the research offers valuable insights for water management authorities, challenges remain in quantifying future water allocations and refining prediction methodologies for enhanced accuracy. Nonetheless, the study paves the way for future research in advancing water demand forecasting methodologies.

INDEX TERMS Deep learning, water demand forecasting, iTransformer.

I. INTRODUCTION

China is notably grappling with severe water scarcity issues on a global scale [1], [2], [3], [4], [5]. Concurrently, the mismanagement of water resources persists as a pressing concern. Essentially, a lack of public awareness regarding water conservation, substantial wastage, inefficient agricultural and industrial water practices, environmental pollution, and excessive groundwater extraction have compounded China's water scarcity crisis, posing a significant impediment to the country's economic development [1], [2], [3], [4], [6]. Consequently, effective water resource management holds paramount importance for China. Acknowledged as a pivotal cornerstone for both water resource governance and economic progress, accurate water demand forecasting is imperative [7], [8], [9]. Therefore, precise prediction of water demand and the provision of recommendations based

on forecasted outcomes are crucial steps toward improving the water resource landscape.

Amidst the backdrop of water scarcity, an increasing array of studies has pivoted their attention in water resources management from supply to demand. Rezaee et al. [10] underscored the importance of judiciously allocating water resources among diverse water use sectors. He et al. [11] found that a majority of cities facing water scarcity can alleviate their water shortage by investing in infrastructure. However, it is crucial to be cautious about the potential environmental consequences that may arise from implementing large-scale solutions to address water scarcity. For instance, Holland et al. [3] revealed that while the electric and gas sectors primarily affect freshwater consumption domestically, the petroleum sector has a significant international footprint, underscoring the need for comprehensive resource-management strategies to ensure energy and freshwater security amidst broader environmental and societal concerns. Scanlon et al. [12] evaluates the current and historical evolution of water resources, highlighting the challenges

The associate editor coordinating the review of this manuscript and approving it for publication was Alba Amato^{ID}.

faced in ensuring its availability due to climate extremes and human intervention, and proposes diversifying management strategies as a solution to increase water-resource resilience. Assessing water resource sustainability depends on various factors including social and economic elements.

Population size and economic status play a critical role in shaping socio-economic scenarios and significantly impact socio-economic water demand, as demonstrated by extant research. For example, Huggins et al. [13] examined the interconnectedness of humans and ecosystems through the hydrological cycle, evaluates the combined impacts of hydrological changes on social and ecological systems globally, and proposes strategies to reduce vulnerability, such as hydro-diplomacy and integrated water resources management practices. Rezaee et al. [10] focused on addressing concerns about water resource sustainability amidst population growth, urbanization, and industrial development, proposing a system dynamics modeling framework integrating economic, social, and environmental dimensions for water resources allocation decisions, with the TOPSIS method utilized to identify optimal allocation strategies, exemplified through a case study in East Azerbaijan province, Iran, emphasizing the need for improved consideration of social, economic, and environmental factors in water allocation.

Currently, water demand forecasting methods primarily fall into three categories: statistical models, fuzzy logic models and machine learning models.

A. STATISTICAL MODELS

In a study conducted in Fortaleza, Brazil, researchers examined the residential water demand forecasting and investigated the potential impact of integrating spatial effects into the modeling process [14]. Their findings indicated that neglecting spatial effects led to an underestimation of the influence of income and toilet numbers on residential water demand marginal prices. Initially, they employed an econometric water demand method without spatial effects, using average price (AP), marginal price with difference (MP), and marginal price with difference using the McFadden method (McFadden model). Subsequently, they explored three models to incorporate spatial effects: Spatial Error Model (SEM), Spatial Autoregressive Model (SAR), and Spatial Autoregressive Moving Average Model (SARMA), considering various explanatory variables. The results favored SARMA, which contradicted the findings of [15], [16] who advocated for spatial approaches. By incorporating spatial effects, the accuracy of demand forecasts improved, increasing price elasticity by 24.66% in the AP model and 13.32% in the McFadden model. Additionally, the seasonal ARIMA model was found to be more effective in forecasting water consumption compared to other methods, yielding mean absolute percentage errors ranging from 1.19% to 15.74% [17].

A short-term water demand forecasting approach grounded in Markov Chain (MC) statistical principles has been

proposed [18]. This method estimates future demands and the associated probabilities of demand falling within expected variability. They presented two techniques, Homogeneous Markov Chains (HMC) and Non-Homogeneous Markov Chains (NHMC), which were applied to forecast water demands in three District Metered Areas (DMA) in Yorkshire, UK, spanning from 1 to 24 hours ahead. Comparative analysis with benchmark methods (such as ANN and Naive Bayes) revealed that HMC outperformed NHMC in providing more accurate short-term predictions. Both HMC and NHMC methodologies offered probabilistic insights into stochastic demand forecasting while exhibiting reduced computational complexity compared to existing techniques. However, they did not match the computational intensity of benchmarks like ANN or Naive Bayes, which could be achieved through Monte Carlo simulations.

B. MACHINE LEARNING MODELS

In recent decades, there has been significant research into the utilization of machine learning or deep learning models as substitutes for statistical models in the estimation and prediction of water demand [7], [19], [20], [21], [22], [23], [24], [25], [26]. Numerous studies have evaluated and compared various machine learning forecasting models based on their accuracy, performance, and practical usability [26], [27], [28], [29].

Long Short-Term Memory (LSTM) [30] tackles the diminishing gradient predicament encountered in conventional Recursive Neural Networks (RNNs). The diminishing gradient issue emerges when the slopes utilized to adjust the weights during instruction decrease exponentially over time, resulting in the inept capture of long-range dependencies. LSTM overcomes this hurdle by introducing a more intricate memory cell framework that permits the retention of information over prolonged time intervals. By managing the information flow through these gateways, LSTM can sustain long-term dependencies more efficiently compared to typical RNNs. Furthermore, the introduction of a cell condition that spans the entire sequence aids in upholding and transmitting information across diverse time intervals without substantial degradation. Fundamentally, LSTM's framework, with its capability to comprehend long-term dependencies and avert the diminishing gradient complication, has markedly enhanced the effectiveness of recursive neural networks, rendering them more appropriate for tasks concerning sequential data and time series forecasting [30]. Wang et al. [31] proposed a novel framework of short-term water demand forecast using the clouded leopard algorithm-based multiple adaptive mechanisms-long short-term memory networks to improve the accuracy of water demand predictions in urban water supply systems.

Transformer [32] embodies a notable progression beyond LSTM networks by surmounting difficulties associated with capturing extensive dependencies in sequences. Via the application of self-attention mechanisms, Substitutes excel in

grasping contextual associations among remote elements within the input sequence, a undertaking that LSTM networks encounter challenges with due to the vanishing gradient dilemma. This capacity enables Substitutes to proficiently represent intricate dependencies and interactions within the information, resulting in enhanced effectiveness for tasks involving sequential data processing. Moreover, Substitutes provide upgraded parallelization during instruction in contrast to LSTM networks. By handling the complete sequence concurrently, Substitutes demonstrate swifter convergence rates and diminished training durations, augmenting their efficiency in managing copious sequential data. This concurrent processing proficiency not only expedites training but also guarantees adaptability with elongated sequences, empowering Substitutes to preserve functionality and efficacy even with extended input sequences, an accomplishment that LSTM networks could struggle to attain. Transformer based water demand forecasting methods have been proved to be more and more popular with high accuracy and efficiency [33], [34], [35].

The Transformer architecture, celebrated for its achievements in natural language processing and computer vision [36], [37], [38], [39], [40], has emerged as a fundamental model in accordance with the scaling principle [37], [38], [39], [41], [42], [43], [44], [45], [46], [47], [48], [49], [50], [51], [52], [53]. With its capacity to capture pairwise connections and derive multi-level representations within sequences, Transformers are currently leaving their imprint on time series prediction. Nonetheless, uncertainties have surfaced among scholars regarding the efficacy of Transformer-driven prediction models. These models typically combine various variables from the same time step into indistinguishable channels and employ attention mechanisms to seize temporal correlations. Recent research indicates that straightforward linear layers, reminiscent of conventional statistical predictors, outperform intricate Transformers concerning both effectiveness and efficiency due to the numerical rather than semantic links among time points [41], [54].

Additionally, current studies highlight the significance of guaranteeing variable autonomy and utilizing shared knowledge. Clearly representing multivariate connections has been recognized as essential for attaining precise prediction. Nonetheless, this goal is difficult to achieve without straying from the traditional Transformer method [55].

Concerning the dangers related to integrating multiple variables of a timestamp as a single temporal token, we suggest an alternate method by treating each variable's entire time series independently and embedding them into distinct variable tokens. This technique, similar to Patching as delineated by [54], showcases an extreme scenario that boosts the local receptive field. By embracing this inverted viewpoint, the embedded token consolidates global representations of the series, enabling a more variable-focused approach that can be efficiently harnessed by advanced attention mechanisms for correlating multiple variables. Simultaneously, the feed-forward network can acquire generalized

representations for various variables derived from arbitrary historical series, aiding in forecasting future series.

This investigation concentrates on constructing a Chinese water demand prediction model based on the deep learning framework. By incorporating temporal information effectively, the model strives to improve the precision of water demand predictions in China. The distinctive attributes of the Chinese water demand sector, characterized by agriculture, industry, household and environment, and evolving water regulations, pose challenges necessitating sophisticated forecasting techniques capable of capturing subtle consumption behavior patterns.

The inspiration behind this inquiry arises from the growing significance of dependable water demand predictions for sustainable water planning and resource distribution in China. Through harnessing the iTransformer deep learning model, this study aims to enrich water demand forecasting methods tailored to the unique characteristics of the Chinese water resources. The integration of spatial features permits a more thorough examination of consumption trends across diverse regions, empowering policymakers and water stakeholders to make informed choices concerning water resource conservations and reallocations. The objective is to introduce and validate a novel deep learning-based model, the iTransformer, which addresses these gaps by outperforming existing models (Transformer, PatchTST, and LSTM) in forecasting accuracy and incorporating key socio-economic variables.

In the subsequent segments, we will explore the methodology employed in constructing the Chinese water demand prediction model grounded on the iTransformer deep learning structure. We will delve into the data sources, model design, training procedures, and evaluation metrics utilized to gauge the model's performance. Moreover, we will present our experiment outcomes and offer insights into the ramifications of precise water.

II. MATERIALS AND METHODS

A. LONG SHORT TERM MEMORY NETWORK(LSTM)

The Long Short-Term Memory (LSTM) model is a form of recurrent neural network (RNN) structure crafted to tackle the fading gradient dilemma in conventional RNNs. Introduced by Hochreiter and Schmidhuber [30], LSTMs are proficient in grasping prolonged dependencies in sequential data by preserving and updating information across lengthy time steps.

Critical elements of an LSTM unit encompass the cell state, input gate, neglect gate, and production gate. These components collaborate to control the transfer of information within the network, enabling LSTMs to retain pertinent data for an extensive duration and selectively discard superfluous specifics. This process empowers LSTMs to capture sequential patterns effectively, rendering them suitable for duties such as voice recognition, linguistic modeling, and temporal series prediction. The depiction of the extensive design of LSTM is exhibited in Figure 1. Its core equations include the following components:

1. Forget Gate:

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

This gate decides what information to forget from the previous cell state.

2. Input Gate:

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

$$\tilde{C}_t = \tanh(WC \cdot [h_{t-1}, x_t] + b_C)$$

The input gate determines what new information to store in the cell, while \tilde{C}_t represents the candidate values for the cell state.

3. Cell State Update:

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t$$

This equation updates the cell state by combining the old cell state and the new candidate values.

4. Output Gate:

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

$$h_t = o_t \odot \tanh(C_t)h_t = o_t \odot \tanh(C_t)$$

The output gate determines what the next hidden state (output) should be based on the current cell state.

LSTMs have garnered broad acceptance in diverse domains due to their effectiveness in modeling sequential data and managing far-reaching dependencies. Their durable framework and capability to recollect information over prolonged sequences have established LSTMs as a pivotal instrument in the domain of profound learning, particularly for assignments necessitating the manipulation of sequential data.

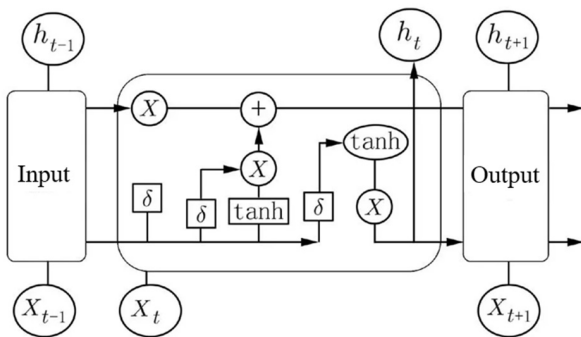


FIGURE 1. Illustration of the comprehensive design of LSTM.

B. TRANSFORMER

The Transformer model, a popular deep learning architecture, was first introduced in the groundbreaking paper “Attention is All You Need” by [32]. Unlike traditional sequence models like LSTMs, Transformers rely on self-attention mechanisms to seize dependencies between input and output elements simultaneously. This framework empowers Transformers to

shine in various natural language processing tasks such as machine translation, text generation, and sentiment analysis.

The fundamental constituents of a Transformer model encompass self-attention coatings and feed-forward neural networks (FFN). Self-attention permits the model to evaluate the significance of different input elements when forecasting, whereas the feedforward networks analyze this data to produce the ultimate outcome. Transformers have soared in popularity owing to their ability to model extensive dependencies proficiently, manage sequential data effectively, and adapt to massive datasets.

The Transformer framework has been broadly embraced and acts as a foundation for numerous state-of-the-art models in natural language processing and other sectors. Its adaptability, parallelization capacities, and exceptional performance have positioned it as a pivotal innovation in deep learning exploration and applications. The depiction of the comprehensive design of Transformer was revealed in Figure 2. Here are the key components and formulas:

1. Input Embeddings

The input tokens are first embedded into continuous vectors:

$$E = \text{Embedding}(X)$$

where XX is the input sequence of token indices.

2. Positional Encoding

Since Transformers do not have a built-in notion of order, positional encodings are added to the embeddings:

$$PE(pos, 2i) = \sin(pos10000^{2i/dmodel})$$

$$PE(pos, 2i + 1) = \cos(pos10000^{2i/dmodel})$$

where pos is the position, ii is the dimension index, and $dmodel$ is the model dimension.

3. Self-Attention Mechanism

For an input sequence of length nn , the self-attention mechanism computes the attention scores as follows:

4. Query, Key, and Value Matrices:

$$Q = EW_Q, K = EW_K, V = EW_V$$

where W_Q, W_K, W_V are learned weight matrices.

5. Scaled Dot-Product Attention:

The attention scores are computed, scaled, and passed through a softmax function:

$$\text{Attention}(Q, K, V) = \text{softmax}(QK^Tdk)V$$

where dk is the dimension of the keys.

6. Multi-Head Attention

To allow the model to focus on different parts of the input, multiple attention heads are used:

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

where each head is defined as:

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

7. Feed-Forward Neural Network

Each position's output from the attention layer is passed through a feed-forward network:

$$\text{FFN}(x) = \text{ReLU}(xW_1 + b_1)W_2 + b_2$$

8. Layer Normalization and Residual Connections

To stabilize training, layer normalization and residual connections are applied:

$$\text{LayerNorm}(x + \text{Sublayer}(x))$$

9. Final Output

The output of the final decoder is passed through a linear layer followed by softmax to produce probabilities for each token in the vocabulary:

$$P(y|x) = \text{softmax}(HW + b)$$

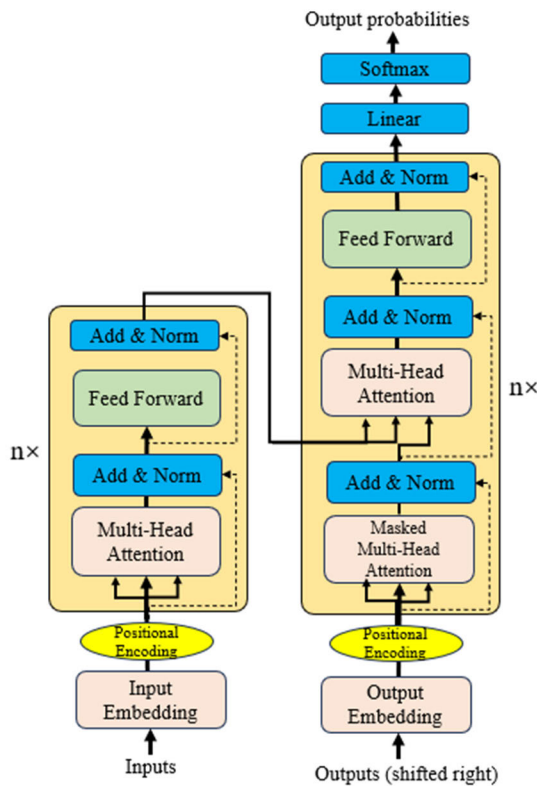


FIGURE 2. Illustration of the comprehensive design of transformer.

C. PatchTST

The PatchTST model embodies the Transformer architecture, tailored for time series forecasting tasks [54]. This model merges the potent sequential modeling prowess of Transformer networks with the demands of time series forecasting, adeptly managing intricate patterns and prolonged relationships within temporal data. Here are key traits and functionalities of the PatchTST model: PatchTST roots itself in the Transformer architecture, renowned for its triumphs in domains like natural language processing. Leveraging

self-attention mechanisms, the Transformer model adeptly captures global dependencies in sequential data, encompassing distant correlations. Time Series Prediction: The PatchTST model is purpose-built for forecasting temporal sequences, proficient in handling continuous data streams and predicting forthcoming intervals. Capitalizing on the Transformer architecture, the model proficiently discerns intricate trends and patterns within temporal data. The Transformer model ingests sequences, processing them through multi-head self-attention mechanisms and feed-forward neural network layers. This enables the PatchTST model to unveil crucial temporal dependencies and patterns in time series data. Chunk Processing: The PatchTST model partitions time series data into multiple segments (patches), each processed autonomously. This chunk-based approach aids in managing lengthy time series data, elevating both the efficiency and accuracy of the model. Data Synchronization: Aligned with other models in the Darts library, the PatchTST model underscores the significance of data synchronization, ensuring accurate data usage during prediction. This bolsters the stability and precision of the model. By harnessing the PatchTST model grounded in the Transformer architecture, one can exploit sophisticated sequence modeling techniques to navigate time series data and attain more precise predictions. This model amalgamates Transformer's strengths with the requisites of time series forecasting, furnishing a robust instrument for addressing intricate temporal data. The comprehensive design of PatchTST is delineated in Figure 3.

D. iTransformer

The inverted Transformer, commonly known as iTransformer, presents a variation of the Transformer framework tailored to counter the challenge of autoregressive models generating sequences in reverse order [56]. This model is specifically crafted to manage tasks necessitating the generation of output sequences in the opposite order of the input sequence. Salient features and functionalities of the inverted Transformer (iTransformer) encompass:

(1) Reverse Sequence Generation: iTransformer is optimized for producing output sequences in reverse sequence compared to the input. This capability proves particularly valuable for tasks demanding output in a reversed sequence pattern.

(2) Modified Attention Mechanism: iTransformer integrates adjustments into the standard attention mechanism, ensuring the model adeptly captures dependencies and patterns during the reverse sequence generation process.

(3) Contextual Encoding: Employing contextual encoding techniques, the model encodes input sequences to facilitate generating output sequences in reverse order while preserving coherence and accuracy.

(4) Bidirectional Processing: The iTransformer architecture facilitates bidirectional processing, enabling efficient utilization of information from both past and future contexts to generate the desired output sequence.

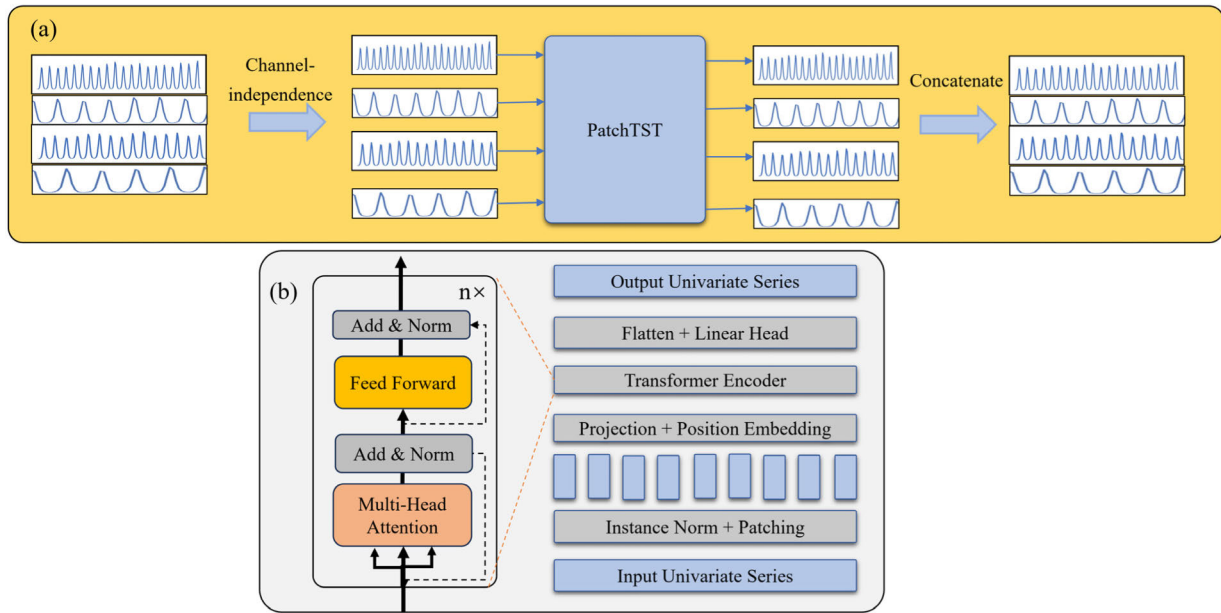


FIGURE 3. Visual depiction of PatchTST's holistic design. (a) Multivariate time series data is segregated into separate channels, all employing a shared Transformer framework, while functioning autonomously during processing. (b) Within each channel, individual univariate series undergo instance normalization and are divided into patches, which subsequently act as input tokens for the transformer.

By harnessing the inverted Transformer (iTransformer) architecture, researchers and practitioners can effectively tackle tasks requiring the generation of output sequences in reverse order, offering a specialized solution for specific sequence generation needs. The architecture of iTransformer, depicted in Figure 4, adopts the encoder-only structure of Transformer, encompassing embedding, projection, and Transformer blocks.

Many forecasting models built on Transformer architectures typically integrate multiple variables as temporal tokens and employ a generative approach for forecasting tasks. However, we've observed that focusing solely on numerical aspects may not optimize learning attention maps effectively. This is evident in the growing adoption of Patching techniques, which have expanded the scope of this domain. Furthermore, the success of linear forecasting models raises questions about the necessity of employing a complex encoder-decoder Transformer for token generation.

In contrast, our proposed encoder-only iTransformer emphasizes representation learning and adaptive correlation within multivariate time series data. Each time series, influenced by intricate underlying processes, undergoes initial tokenization to capture its unique characteristics. It then proceeds through self-attention mechanisms to facilitate inter-variable interactions and individual processing via feed-forward networks to establish series representations. Notably, the task of generating predicted series primarily relies on linear layers, a strategy supported by previous research, with a comprehensive analysis provided in the subsequent section.

E. FORECASTING WATER DEMANDS USING DEEP LEARNING MODELS

Analyzing consumption trends serves the primary goal of forecasting future electricity demand based on timestamps. This process involves determining various parameters, such as input and output sequence lengths, followed by constructing deep learning models to predict water demand using socio-economic data. Commonly employed machine models like LSTM excel in predicting time series due to their ability to capture intricate relationships. In this study, four deep learning models—iTransformer, Transformer, PatchTST, and LSTM—are established to forecast future electricity consumption.

To construct these models, the water demand dataset is divided into training, validation, and testing sets. During the training phase, all training patterns are fed into the regression models to establish an effective mapping from input to output. Validation data is then utilized to impartially evaluate model fit on the training data while fine-tuning hyperparameters. Subsequently, the performance of the trained models is evaluated on the testing set.

Several considerations are crucial when constructing prediction models:

Hyperparameters Selection: Parameters like the number of input neurons, output neurons, batch size, etc., must be accurately determined for effective prediction models. Sequence length parameters are determined based on input and output data, while other parameters like number of epochs and batch size are chosen through iterative training and validation runs.

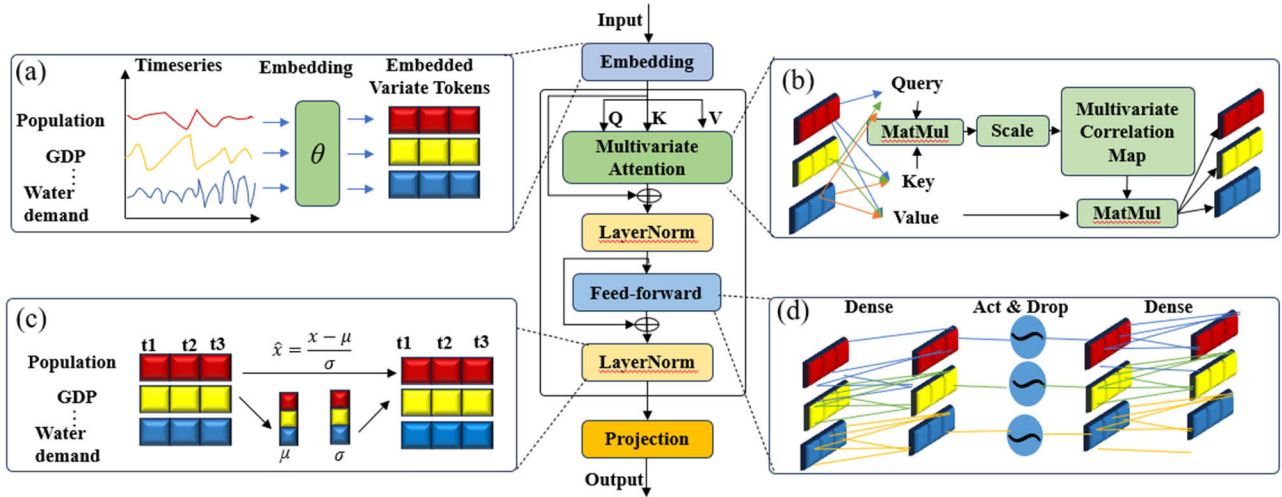


FIGURE 4. Visual representation of the comprehensive design of iTransformer, exhibiting a modular structure akin to the Transformer encoder. (a) Different raw data series are encoded independently into tokens. (b) Self-attention mechanisms amplify interpretability, revealing complex multivariate correlations within the embedded tokens. (c) A common feed-forward network extracts series representations for each token. (d) Layer normalization is applied to mitigate discrepancies among the variates.

Optimization Technique and Loss Function: In this study, the Root Mean Squared Error (RMSE) loss function is utilized to assess network performance on training and validation datasets.

Three model performance metrics are employed to evaluate the accuracy of the machine learning models:

RMSE: It quantifies the standard deviation of differences between predicted and actual values.

$$RMSE = \sqrt{\frac{1}{n} \sum_{j=1}^n (y_j - o_j)^2} \quad (1)$$

Mean Absolute Error (MAE): It calculates the average absolute difference between the observed and forecasted values in a dataset.

$$MAE = \frac{1}{n} \sum_{j=1}^n |o_j - y_j| \quad (2)$$

Mean Absolute Percentage Error (MAPE): It calculates the average absolute difference between the observed and forecasted values in a dataset.

$$MAPE = \frac{1}{n} \sum_{j=1}^n \left| \frac{o_j - y_j}{o_j} \right| \quad (3)$$

where o is observed value and y is forecasted value, j indicates time step and n indicates the total time steps, σ_y and σ_o are the variance of y and o .

III. RESULTS

In this segment, deep learning models are utilized to predict the annual water demand in China and four key regions: Beijing, Jiangsu, Zhejiang, and Guangdong. Initially, the annual water demand data for both China and the aforementioned regions is gathered from the China Statistical Yearbook, accessible at <http://www.stats.gov.cn/english/>, as detailed in Table 1 and Table 2. This data is divided into

two sets, with information spanning from 1951 to 2021 utilized for constructing prediction models, while the remaining samples are reserved for evaluating the prediction accuracies of these models. Drawing inspiration from existing literature, a structural diagram depicting the forecast of annual water demand in China is presented in Figure 5.

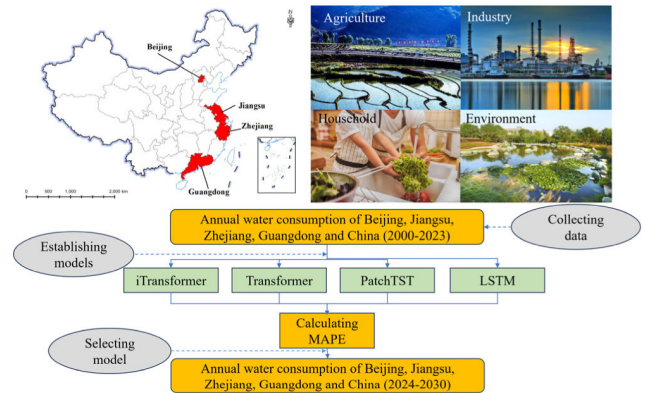


FIGURE 5. Illustration of the forecast of water demand structure.

A. WATER DEMAND OF CHINA

In this section, China's water demand is analyzed using four machine learning models. Table 3 presents the performance of these models in forecasting China's water demand, while Figure 6 visually illustrates the trajectory of estimated and observed water demand in China. Initially, the four deep learning models (iTransformer, PatchTST, Transformer, and LSTM) were trained and tested using China's water demand data from 2000 to 2023, considering various forecast lengths. Subsequently, the best-performing model was selected based on the minimum Mean Absolute Percentage Error (MAPE).

TABLE 1. Population, GDP (ten thousand Yuan), cropland area (ha) and water demand (100 million tons) of China during 2000-2023.

Year	Population	GDP (Ten thousand Yuan)	Crop land (ha)	water demand (100 million tons)
2000	126743	100280.1	128243000	5497.6
2001	127627	110863.1	127616000	5567.4
2002	128453	121717.4	125930000	5497.3
2003	129227	137422	123392000	5320.4
2004	129988	161840.2	122444000	5547.8
2005	130756	187319	122083000	5633
2006	131448	219439	121776000	5795
2007	132129	270232.3	121735000	5818.7
2008	132802	319515.5	121716000	5910
2009	133450	349081.4	135385000	5965.2
2010	134091	413030.3	135268000	6022
2011	134916	489300.6	135239000	6107.2
2012	135922	540367.4	135158000	6131.2
2013	136726	595244.4	135163400	6183.4
2014	137646	643974	135057300	6094.9
2015	138326	689052.1	134998700	6103.2
2016	139232	746395.1	134920900	6040.2
2017	140011	832036	134881200	6043.4
2018	140541	919281.1	134881200	6015.5
2019	141008	990865	127861900	6021.2
2020	141212	1013567	127436700	5812.9
2021	141260	1149237	127516800	5920.2
2022	141175	1197250.4	127579900	5998.2
2023	140967	1260582	127580000	5907

From Table 3, it is evident that iTransformer outperformed PatchTST, Transformer, and LSTM, exhibiting the lowest RMSE (92.72), MAE (68.65), and MAPE (0.01) in water demand forecasting with 1 year forecast length. Moreover, iTransformer also demonstrated superior performance in long-term forecasting with forecast lengths ranging from 2 to 7 years compared to PatchTST, Transformer, and LSTM. Finally, the best model (iTransformer) was employed to forecast China's water demand for the period 2022-2030 (Table 4 and Figure 6). The water demand in China is projected to experience a slight increase and then stabilize at a steady level.

B. WATER DEMAND OF BEIJING CITY

In this section, the pre-trained models are utilized to forecast Beijing's water demand. The performance of these models in predicting Beijing's water demand is outlined in Table 5, while the trajectory of estimated and observed water demand for Beijing is visually represented in Table 6 and Figure 7.

TABLE 2. Water demand of Beijing, Jiangsu, Zhejiang and Guangdong (100 million tons) during 2000-2023.

Year	Beijing	Jiangsu	Zhejiang	Guangdong
2000	39.5	353.36	201.15	525.61
2001	38.9	332.15	205.36	511.55
2002	34.6	380.99	208.01	548.03
2003	35	433.5	205.98	457.5
2004	34.6	525.6	207.78	464.82
2005	34.5	519.72	209.91	459
2006	34.3	546.4	208.26	459.4
2007	34.81	558.34	210.98	462.51
2008	35.1	558.32	216.62	461.53
2009	35.5	549.2	217.07	463.41
2010	35.2	552.2	220.08	469
2011	35.19	552.19	222.24	469.01
2012	35.9	552.2	222.31	451
2013	36.4	576.7	224.75	443.2
2014	37.5	591.3	220.24	442.5
2015	38.2	574.5	186.06	443.1
2016	38.8	577.4	181.15	435
2017	39.5	591.3	179.5	433.5
2018	39.3	592	173.81	421
2019	41.7	619.1	165.79	412.3
2020	40.6	572	163.94	405.1
2021	40.8	567.5	166.4	407
2022	40.0	611.8	167.8	401.7
2023	41.8	600.9	174.43	393.2

The water demand in Beijing is expected to increase over the years.

C. WATER DEMAND OF JIANGSU

In this section, the pre-trained models are utilized to forecast the water demand in Jiangsu. The performance of these models in predicting Jiangsu's water demand is outlined in Table 7, while the trajectory of estimated and observed water demand for Jiangsu is visually depicted in Table 8 and Figure 8. The water demand in Jiangsu is anticipated to fluctuate over the years without experiencing significant increases or decreases.

D. WATER DEMAND OF ZHEJIANG

In this section, Zhejiang's water demand is forecasted using pre-trained models. The performance of these models in predicting Zhejiang's water demand is outlined in Table 9, while the trajectory of estimated and observed water demand for Zhejiang is visually depicted in Table 10 and Figure 9. The water demand in Zhejiang is projected to experience a slight increase over the years and then stabilize at a steady level.

E. WATER DEMAND OF GUANGDONG

In this section, the pre-trained models are employed to forecast Guangdong's water demand. The performance of these models in predicting Guangdong's water demand is detailed in Table 11, while the trajectory of estimated and observed water demand for Guangdong is visually illustrated in Table 12 and Figure 10. The water demand in Guangdong is expected to decrease over the years.

The proposed approach operates in the following environment:

TABLE 3. The metrics of models in estimating water demand of China.

Forecast length	Metrics	iTransformer		PatchTST		Transformer		LSTM	
		train	test	Train	test	train	test	train	test
1	RMSE (100 million tons)	92.72	97.54	98.50	105.73	104.29	110.55	111.52	118.75
	MAE (100 million tons)	68.65	68.92	69.20	70.02	70.29	71.00	71.11	71.93
	MAPE	0.01	0.05	0.10	0.23	0.27	0.38	0.40	0.53
2	RMSE (100 million tons)	94.88	94.95	95.01	95.21	95.28	95.45	95.48	95.68
	MAE (100 million tons)	70.01	70.05	70.09	70.22	70.26	70.37	70.38	70.51
	MAPE	0.01	0.04	0.06	0.13	0.16	0.22	0.23	0.30
3	RMSE (100 million tons)	97.07	97.11	97.15	97.28	97.32	97.43	97.44	97.57
	MAE (100 million tons)	71.10	71.24	71.39	71.82	71.96	72.33	72.39	72.82
	MAPE	0.01	0.02	0.03	0.05	0.05	0.07	0.08	0.10
4	RMSE (100 million tons)	95.55	95.98	96.41	97.69	98.12	99.23	99.40	100.69
	MAE (100 million tons)	72.46	72.92	73.38	74.76	75.23	76.42	76.61	77.99
	MAPE	0.01	0.05	0.09	0.20	0.24	0.34	0.35	0.47
5	RMSE (100 million tons)	105.22	105.58	105.95	107.04	107.40	108.35	108.49	109.59
	MAE (100 million tons)	83.95	84.11	84.28	84.77	84.93	85.36	85.43	85.92
	MAPE	0.01	0.06	0.11	0.25	0.30	0.43	0.44	0.59
6	RMSE (100 million tons)	144.00	144.21	144.42	145.06	145.27	145.82	145.90	146.54
	MAE (100 million tons)	121.46	121.80	122.14	123.16	123.50	124.38	124.52	125.54
	MAPE	0.02	0.06	0.09	0.20	0.24	0.33	0.34	0.45
7	RMSE (100 million tons)	419.60	427.56	429.15	441.09	438.71	449.05	450.65	462.59
	MAE (100 million tons)	221.46	221.78	222.09	223.04	223.35	224.17	224.30	225.24
	MAPE	0.03	0.08	0.13	0.27	0.32	0.45	0.47	0.61

Hardware Configuration: 64 GB RAM, Xeon Processor with 20 cores, 1 TB Hard Disk.
Software Configuration: Operating System – Windows 10.

F. ABLATION EXPERIMENT
To assess the effectiveness of iTransformer components, we present detailed ablation studies involving both

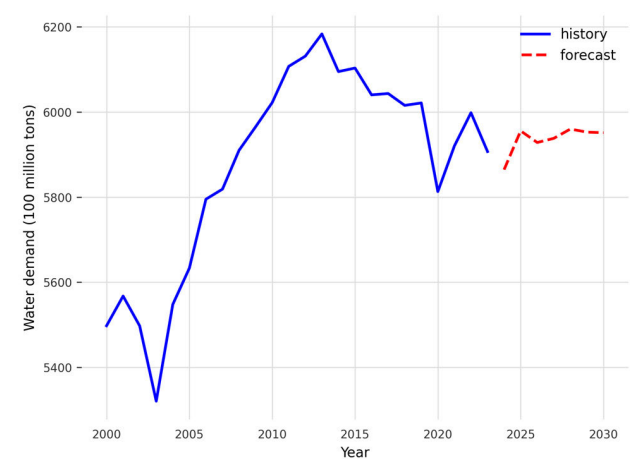


FIGURE 6. Historical and forecasted water demand (100 million tons) of China during 2000-2023 and 2023-2030 with iTransformer.

TABLE 4. The forecasted water demand (100 million tons) of China during 2024-2030.

Year	Forecast
2024	5864.53
2025	5955.58
2026	5928.26
2027	5938.07
2028	5959.90
2029	5952.64
2030	5951.34

TABLE 5. The metrics of models in forecasting water demand of Beijing.

Metrics	iTransformer		PatchTST		Transformer		LSTM	
	train	test	Train	test	train	test	train	test
RMSE (100 million tons)	1.39	1.67	1.72	2.14	2.06	2.42	2.48	2.89
MAE (100 million tons)	1.11	1.11	1.11	1.12	1.12	1.13	1.13	1.14
MAPE	0.28	0.30	0.32	0.39	0.41	0.47	0.48	0.55

TABLE 6. The forecasted water demand (100 million tons) of Beijing during 2024-2030.

Year	Forecast
2024	41.34
2025	42.27
2026	42.26
2027	42.83
2028	43.03
2029	43.46
2030	43.74

component replacement (Replace) and component removal (w/o) experiments. The findings are summarized in Table 13. The iTransformer, which employs attention mechanisms on the variate dimension and feed-forward networks on the temporal dimension, consistently delivers the best performance. In contrast, the vanilla Transformer (shown in the

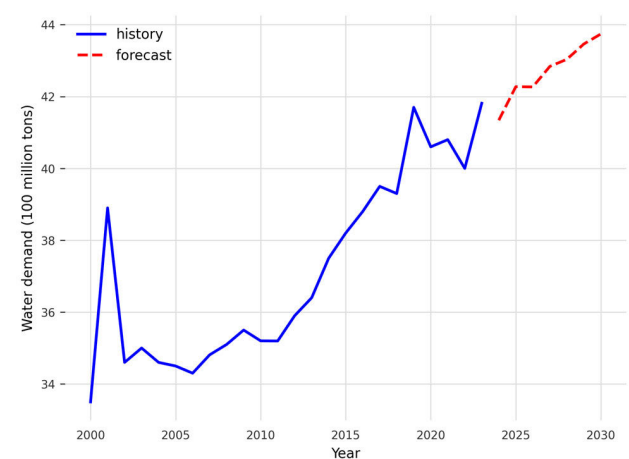


FIGURE 7. Historical and forecasted water demand (100 million tons) of Beijing during 2000-2023 and 2024-2030 with iTransformer.

TABLE 7. The metrics of models in estimating water demand of Jiangsu.

Metrics	iTransformer		PatchTST		Transformer		LSTM	
	train	test	Train	test	train	test	train	test
RMSE (100 million tons)	22.71	23.70	23.90	25.39	25.09	26.38	26.58	28.07
MAE (100 million tons)	17.42	17.44	17.46	17.52	17.54	17.59	17.59	17.65
MAPE	0.03	0.06	0.08	0.16	0.19	0.25	0.26	0.34

TABLE 8. The forecasted water demand (100 million tons) of Jiangsu during 2024-2030.

Year	Forecast
2024	581.94
2025	595.99
2026	592.48
2027	580.02
2028	588.32
2029	592.07
2030	585.04

third row) demonstrates the poorest performance among the designs, highlighting potential drawbacks of the traditional architecture.

G. EFFICIENCY ANALYSIS

In this section, we evaluate the efficiency of iTransformer in comparison to other models (PatchTST, Transformer, LSTM) using the China dataset. To ensure a fair comparison, we examine the average training time per epoch for each model with a batch size of 32. As illustrated in Table 14, iTransformer exhibits a relatively least training time. Unlike other models, iTransformer can enjoy a comparable speed and memory footprint with linear models.

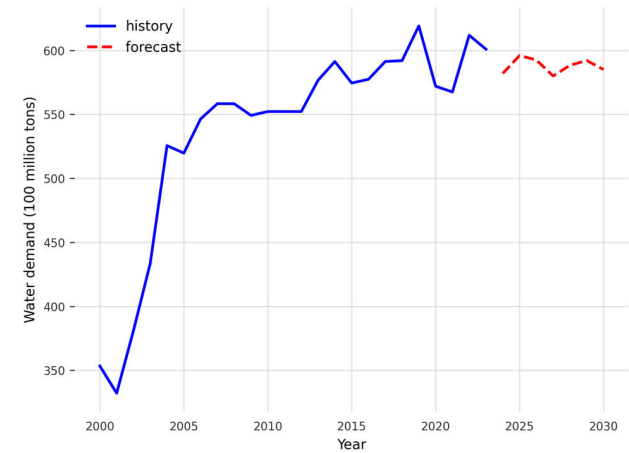


FIGURE 8. Historical and forecasted water demand (100 million tons) of Jiangsu during 2000-2023 and 2024-2030 with iTransformer.

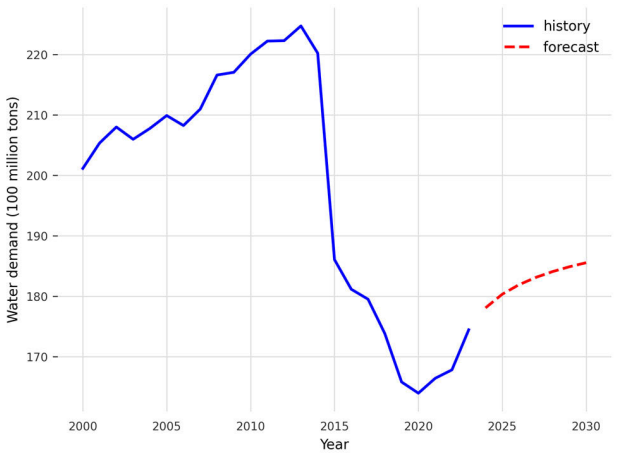


FIGURE 9. Historical and forecasted water demand (100 million tons) of Zhejiang during 2000-2023 and 2024-2030 with iTransformer.

TABLE 9. The metrics of models in forecasting water demand of Zhejiang.

Metrics	iTransformer		PatchTST		Transformer		LSTM	
	train	test	Train	test	train	test	train	test
RMSE (100 million tons)	21.69	22.07	22.15	22.73	22.61	23.11	23.19	23.77
MAE (100 million tons)	13.38	13.42	13.47	13.60	13.65	13.76	13.78	13.92
MAPE	0.08	0.12	0.16	0.29	0.33	0.44	0.45	0.58

TABLE 10. The forecasted water demand (100 million tons) of Zhejiang during 2022-2030.

Year	Forecast
2024	178.07
2025	180.32
2026	181.90
2027	183.11
2028	184.08
2029	184.87
2030	185.54

IV. DISCUSSION

A. WATER DEMAND OF CHINA

It is forecasted that the water demand in China will exhibit a slight rise but eventually stabilize at a consistent level. Over an extended period, extensive irrigation projects and the expansion of water-intensive industries have led to a significant imbalance between water demand and available surface water resources in the China. This has resulted in prolonged overexploitation of groundwater for agricultural, industrial and household water demand.

B. WATER DEMAND OF BEIJING

The water demand in Beijing is expected to increase over the years. As the capital of China, Beijing assumes increasingly

TABLE 11. The metrics of models in estimating water demand of Guangdong.

Metrics	iTransformer		PatchTST		Transformer		LSTM	
	train	test	Train	test	train	test	train	test
RMSE (100 million tons)	9.16	9.17	9.17	9.19	9.18	9.20	9.20	9.21
MAE (100 million tons)	5.85	5.90	5.95	6.10	6.15	6.27	6.29	6.44
MAPE	0.01	0.05	0.08	0.18	0.21	0.30	0.31	0.41

TABLE 12. The forecasted water demand (100 million tons) of Guangdong during 2024-2030.

Year	Forecast
2024	386.28
2025	379.13
2026	371.66
2027	363.62
2028	355.11
2029	346.08
2030	336.52

significant roles and attracts a growing influx of people and investment. Consequently, this trend is expected to lead to a rise in water demand for the city.

C. WATER DEMAND OF JIANGSU

The water demand in Jiangsu is anticipated to fluctuate over the years without experiencing significant increases or decreases. The anticipated fluctuations in water demand in Jiangsu without significant increases or decreases can be attributed to various factors. These may include stable population growth, consistent economic activity, and efficient

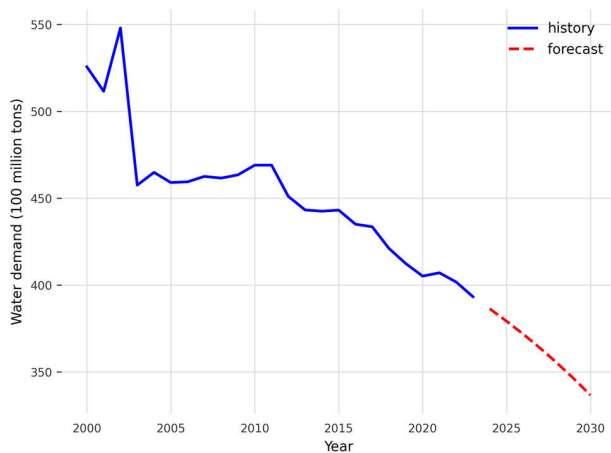


FIGURE 10. Historical and forecasted water demand (100 million tons) of Guangdong during 2000-2023 and 2022-2030 with iTransformer.

TABLE 13. Ablations on iTransformer. We replace different components on the respective dimension to learn multivariate correlations (Variate) and series representations (Temporal), in addition to component removal. The average results of all predicted lengths are listed here.

Design	Variate	Temporal	RMSE	MAE
iTransformer	Attention	FFN	92.2	68.65
	Attention	Attention	100.36	152.36
Replace	FFN	Attention	105.04	156
	FFN	FFN	94.64	149.24
w/o	Attention	w/o	98.28	144.56
	w/o	FFN	100.36	143.52

TABLE 14. Training time for each epoch of different models.

Model	Speed(s/epoch)
iTransformer	42.3
PatchTST	43.4
Transformer	45.2
LSTM	46.2

water management practices that help optimize water usage within the region.

D. WATER DEMAND OF ZHEJIANG

The water demand in Zhejiang is projected to experience a slight increase over the years and then stabilize at a steady level. In the case of Zhejiang, the projection of a slight increase in water demand over the years followed by stabilization can be influenced by factors such as population growth, economic development, and infrastructure expansion. As the population and economy grow, there is likely to be a corresponding increase in water demand. However, with effective water management strategies and resource allocation, the water demand can eventually stabilize at a steady level.

E. WATER DEMAND OF GUANGDONG

The water demand in Guangdong is expected to decrease over the years. The expected decrease in water demand in Guangdong over the years might be attributed to several

factors. These could include advancements in water-saving technologies, implementation of conservation measures, and improvements in water management practices. Additionally, shifts in industrial activities, changes in economic structure, and potential population trends can also contribute to the decrease in water demand in the region.

Some of our results are consistent with previous studies. Xiangmei et al. [57] proposed a grey multivariate convolution model with adjacent accumulation (AGMC(1,N)) to forecast the annual water consumption across 31 regions (including provinces, municipalities, and autonomous regions) in China, considering varying growth rates of regional GDP and population. Their findings demonstrate that the AGMC(1,N) model achieves superior prediction accuracy. Moreover, over 50% of the regions exhibit a decline in water consumption attributed to the growth of regional GDP. Similarly, the influence of population leads to a decrease in water consumption in over 50% of the areas. The trend in Guangdong are both decreasing in our results and their results. However, the trend in Beijing is increase first and decrease [57], the trend in Jiangsu is increasing and the trend in Zhejiang is decreasing are different from ours. Their results are based on data from 2012 to 2018, which is a little shorter than ours (2000-2023) and their method is statistics model, which may not fully capture the non-linear relationships among water demand and social-economic data.

The importance of accurately predicting water demand has been somewhat overlooked compared to other areas, despite its crucial significance. Recent advancements have shown notable progress in developing precise models for forecasting water demand. Machine Learning (ML) methods have emerged as effective tools for addressing non-linear challenges, gaining prominence in this domain.

In this study, we employed a deep learning model called iTransformer for forecasting water demand. This methodology builds upon the Transformer network technique, allowing for the estimation of water demand across multiple time series. Through comparative performance analysis against established models like PatchTST, Transformer, and LSTM, our results indicate that the iTransformer framework surpasses traditional methods like LSTM, demonstrating its efficacy in accurately predicting electricity consumption.

Our proposed approach offers several advantages over existing methods in demand forecasting. The framework adeptly handles non-linear complexities and captures both short-term and long-term dependencies within water demand time series data. Simulation results showcase minimal prediction errors with the iTransformer framework, ensuring precise consumption estimates for various forecasting lengths. Moreover, the developed framework can be readily applied to estimate demand in different geographic locations, relying solely on historical data.

Future research directions could explore incorporating nonlinear exogenous features such as climate conditions and economic variables to analyze trends in water demand patterns. Additionally, optimization techniques could be devised

to further enhance the prediction accuracy of learning models, thereby advancing the field of water demand forecasting.

Our results help in determining future water needs, allowing for better planning in the allocation of water resources among various sectors (e.g., agriculture, industry, domestic use). Insights from forecasts guide decisions on investing in infrastructure, such as reservoirs, pipelines, and treatment plants, to ensure they meet future demands. Accurate forecasts inform policy-makers about future water scarcity or surpluses, influencing regulations on water usage, conservation practices, and sustainable development. Forecasting assists in estimating future water costs and planning budget allocations, which is vital for maintaining economic stability. By predicting future demand, businesses and governments can make informed decisions about investing in water-efficient technologies and practices. Water demand forecast results can also be used to raise awareness about water scarcity issues and encourage community involvement in water conservation efforts.

V. CONCLUSION

This paper introduces a deep learning-based model for forecasting water consumption, representing a significant contribution to the application of deep learning in water demand forecasting. The primary innovation lies in developing a medium-term forecasting model trained on historic social-economic data, enabling accurate predictions for water demand. A transformer network architecture based iTransformer model is proposed to forecast total water demand at country and province level over the medium term. Furthermore, iTransformer model was compared with Transformer, PatchTST and LSTM models at different forecasting length. Furthermore, a grid search is conducted to optimize hyperparameters such as the number of layers, neurons per layer, and learning rate. Once optimal model and parameters are identified, the best-performing model is applied to historical water demand data from 2000 to 2023 at a yearly frequency to generate forecasts for the subsequent years. Results, based on historical water demand data and social-economic data, demonstrate a lowest RMSE (92.72/ 1.39/ 22.71/ 21.69/ 9.16), MAE (68.65/ 1.11/ 17.42/ 13.38/ 5.85), and MAPE (0.01/0.28/0.03/0.08/0.01). Comparisons with Transformer, PatchTST and LSTM models reveal that the proposed iTransformer network achieves the lowest errors in forecasting water demand of China, Beijing, Jiangsu, Zhejiang and Guangdong. The paper recommends applying this forecasting model when historical water demand and social-economic data (population and GDP) are available.

Essentially, the demand for water resources arises from human livelihood necessities and economic progress, making the total socio-economic water demand of China intricately linked to population size and economic activity. Hence, it is vital to strike a balance between socio-economic advancement and water resource utilization, avoiding both stifling population and economic growth solely for water conservation purposes and recklessly pursuing economic expansion

at the expense of water resource limitations. Rational water conservation involves enhancing water resource efficiency, optimizing resource allocation, and fostering sustainable social and economic growth.

This research represents an innovative approach to water demand forecasting through the integration of deep learning method, offering significant value for water management authorities in crafting long-term water demand management strategies and mitigating water conflicts. However, a limitation persists. The challenge lies in quantifying future water allocations across various trajectories, leading to the utilization of anticipated water allocations based on current trends for water demand projections. Nonetheless, a detailed analysis is conducted for each trajectory to bolster accuracy. Alternative methodologies can be explored to acquire more precise parameters, and further investigation can be conducted to enhance water demand prediction methodologies.

In the future, advanced machine learning techniques like deep learning and ensemble methods are needed to improve the accuracy of demand forecasting. Adaptive algorithms that can also adjust to changes in water consumption patterns and other dynamic factors. Real-time data from smart meters and IoT devices can be used to enhance forecasting models. Diverse data sources such as weather forecasts, land use changes, and socio-economic factors can be used to create more robust models.

Our models can be used to balance water supply and demand, optimizing reservoir management and distribution systems. Infrastructure investments and upgrades can be based on predicted water demand trends. Early warning systems for droughts and water shortages can also be based on demand forecasts.

CONFLICTS OF INTEREST

The authors declare no conflicts of interest.

REFERENCES

- [1] C. Dalin, H. Qiu, N. Hanasaki, D. L. Mauzerall, and I. Rodriguez-Iturbe, "Balancing water resource conservation and food security in China," *Proc. Nat. Acad. Sci. USA*, vol. 112, pp. 4588–4593, Mar. 2015, doi: [10.1073/pnas.1504345112](https://doi.org/10.1073/pnas.1504345112).
- [2] A. Y. Hoekstra and M. M. Mekonnen, "The water footprint of humanity," *Proc. Nat. Acad. Sci. USA*, vol. 109, no. 9, pp. 3232–3237, Feb. 2012, doi: [10.1073/pnas.1109936109](https://doi.org/10.1073/pnas.1109936109).
- [3] R. A. Holland, K. A. Scott, M. Flörke, G. Brown, R. M. Ewers, E. Farmer, V. Kapos, A. Muggeridge, J. P. W. Scharlemann, G. Taylor, J. Barrett, and F. Eigenbrod, "Global impacts of energy demand on the freshwater resources of nations," *Proc. Nat. Acad. Sci. USA*, vol. 112, no. 48, pp. E6707–E6716, Dec. 2015, doi: [10.1073/pnas.1507701112](https://doi.org/10.1073/pnas.1507701112).
- [4] S. Li, Y. Zhuang, H. Liu, Z. Wang, F. Zhang, M. Lv, L. Zhai, X. Fan, S. Niu, J. Chen, C. Xu, N. Wang, S. Ruan, W. Shen, M. Mi, S. Wu, Y. Du, and L. Zhang, "Enhancing rice production sustainability and resilience via reactivating small water bodies for irrigation and drainage," *Nature Commun.*, vol. 14, no. 1, p. 3794, Jun. 2023, doi: [10.1038/s41467-023-39454-w](https://doi.org/10.1038/s41467-023-39454-w).
- [5] X. Zhao, J. Liu, Q. Liu, M. R. Tillotson, D. Guan, and K. Hubacek, "Physical and virtual water transfers for regional water stress alleviation in China," *Proc. Nat. Acad. Sci. USA*, vol. 112, no. 4, pp. 1031–1035, Jan. 2015, doi: [10.1073/pnas.1404130112](https://doi.org/10.1073/pnas.1404130112).

- [6] E. R. Jones, M. F. P. Bierkens, and M. T. H. van Vliet, "Current and future global water scarcity intensifies when accounting for surface water quality," *Nature Climate Change*, vol. 14, no. 6, pp. 629–635, Jun. 2024, doi: [10.1038/s41558-024-02007-0](https://doi.org/10.1038/s41558-024-02007-0).
- [7] G. Fu, Y. Jin, S. Sun, Z. Yuan, and D. Butler, "The role of deep learning in urban water management: A critical review," *Water Res.*, vol. 223, Sep. 2022, Art. no. 118973, doi: [10.1016/j.watres.2022.118973](https://doi.org/10.1016/j.watres.2022.118973).
- [8] W. Guo, T. Liu, F. Dai, and P. Xu, "An improved whale optimization algorithm for forecasting water resources demand," *Appl. Soft Comput.*, vol. 86, Jan. 2020, Art. no. 105925, doi: [10.1016/j.asoc.2019.105925](https://doi.org/10.1016/j.asoc.2019.105925).
- [9] S. S. Virnoddar, V. K. Pachghare, V. C. Patil, and S. K. Jha, "Remote sensing and machine learning for crop water stress determination in various crops: A critical review," *Precis. Agricult.*, vol. 21, no. 5, pp. 1121–1155, Oct. 2020, doi: [10.1007/s11119-020-09711-9](https://doi.org/10.1007/s11119-020-09711-9).
- [10] A. Rezaee, O. Bozorg-Haddad, and X. Chu, "Reallocation of water resources according to social, economic, and environmental parameters," *Sci. Rep.*, vol. 11, no. 1, Sep. 2021, Art. no. 17514, doi: [10.1038/s41598-021-96680-2](https://doi.org/10.1038/s41598-021-96680-2).
- [11] C. He, Z. Liu, J. Wu, X. Pan, Z. Fang, J. Li, and B. A. Bryan, "Future global urban water scarcity and potential solutions," *Nature Commun.*, vol. 12, no. 1, p. 12, Aug. 2021, doi: [10.1038/s41467-021-25026-3](https://doi.org/10.1038/s41467-021-25026-3).
- [12] B. R. Scanlon et al., "Global water resources and the role of groundwater in a resilient water future," *Nature Rev. Earth Environ.*, vol. 4, no. 2, pp. 87–101, Jan. 2023, doi: [10.1038/s43017-022-00378-6](https://doi.org/10.1038/s43017-022-00378-6).
- [13] X. Huggins, T. Gleeson, M. Kumm, S. C. Zipper, Y. Wada, T. J. Troy, and J. S. Famiglietti, "Hotspots for social and ecological impacts from freshwater stress and storage loss," *Nature Commun.*, vol. 13, no. 1, p. 439, Jan. 2022, doi: [10.1038/s41467-022-28029-w](https://doi.org/10.1038/s41467-022-28029-w).
- [14] D. de Maria André and J. R. Carvalho, "Spatial determinants of urban residential water demand in Fortaleza, Brazil," *Water Resour. Manage.*, vol. 28, no. 9, pp. 2401–2414, Jul. 2014, doi: [10.1007/s11269-014-0551-0](https://doi.org/10.1007/s11269-014-0551-0).
- [15] H. Chang, G. H. Parandvash, and V. Shandas, "Spatial variations of single-family residential water consumption in Portland, Oregon," *Urban Geography*, vol. 31, no. 7, pp. 953–972, Oct. 2010.
- [16] L. A. House-Peters and H. Chang, "Urban water demand modeling: Review of concepts, methods, and organizing principles," *Water Resour. Res.*, vol. 47, no. 5, May 2011, Art. no. W05401, doi: [10.1029/2010wr009624](https://doi.org/10.1029/2010wr009624).
- [17] D. C. M. Ristow, E. Henning, A. Kalbusch, and C. E. Petersen, "Models for forecasting water demand using time series analysis: A case study in Southern Brazil," *J. Water, Sanit. Hygiene Develop.*, vol. 11, no. 2, pp. 231–240, Mar. 2021, doi: [10.2166/washdev.2021.208](https://doi.org/10.2166/washdev.2021.208).
- [18] F. Gagliardi, S. Alvisi, Z. Kapelan, and M. Franchini, "A probabilistic short-term water demand forecasting model based on the Markov chain," *Water*, vol. 9, no. 7, p. 507, Jul. 2017, doi: [10.3390/w9070507](https://doi.org/10.3390/w9070507).
- [19] A. A. Alhendi, A. S. Al-Sumaiti, F. K. Elmay, J. Wescaot, A. Kavousi-Fard, E. Heydarian-Forushani, and H. H. Alhelou, "Artificial intelligence for water–energy Nexus demand forecasting: A review," *Int. J. Low-Carbon Technol.*, vol. 17, pp. 730–744, Feb. 2022, doi: [10.1093/ijlct/ctac043](https://doi.org/10.1093/ijlct/ctac043).
- [20] H. Namdari, S. M. Ashrafi, and A. Haghighi, "Deep learning-based short-term water demand forecasting in urban areas: A hybrid multichannel model," *AQUA-Water Infrastructure, Ecosystems Soc.*, vol. 73, no. 3, pp. 380–395, Mar. 2024, doi: [10.2166/aqua.2024.200](https://doi.org/10.2166/aqua.2024.200).
- [21] H. Namdari, A. Haghighi, and S. M. Ashrafi, "Short-term urban water demand forecasting; application of 1D convolutional neural network (1D CNN) in comparison with different deep learning schemes," *Stochastic Environ. Res. Risk Assessment*, pp. 1–16, Sep. 2023, doi: [10.1007/s00477-023-02565-3](https://doi.org/10.1007/s00477-023-02565-3).
- [22] T. M. N. Carvalho, F. D. A. de Souza Filho, and V. C. Porto, "Urban water demand modeling using machine learning techniques: Case study of Fortaleza, Brazil," *J. Water Resour. Planning Manage.*, vol. 147, no. 1, Jan. 2021, Art. no. 05020026, doi: [10.1061/\(asce\)wr.1943-5452.0001310](https://doi.org/10.1061/(asce)wr.1943-5452.0001310).
- [23] G. D. S. Groppo, M. A. Costa, and M. Libânio, "Predicting time-series for water demand in the big data environment using statistical methods, machine learning and the novel analog methodology dynamic time scan forecasting," *Water Supply*, vol. 23, no. 2, pp. 624–644, Feb. 2023, doi: [10.2166/ws.2023.008](https://doi.org/10.2166/ws.2023.008).
- [24] O. Iwakiri and F. Moazeni, "Improving urban water demand forecast using conformal prediction-based hybrid machine learning models," *J. Water Process Eng.*, vol. 58, Feb. 2024, Art. no. 104721, doi: [10.1016/j.jwpe.2023.104721](https://doi.org/10.1016/j.jwpe.2023.104721).
- [25] M. Rezaali, J. Quilty, and A. Karimi, "Probabilistic urban water demand forecasting using wavelet-based machine learning models," *J. Hydrol.*, vol. 600, Sep. 2021, Art. no. 126358, doi: [10.1016/j.jhydrol.2021.126358](https://doi.org/10.1016/j.jhydrol.2021.126358).
- [26] K. Smolak, B. Kasieczka, W. Fiałkiewicz, W. Rohm, K. Siła-Nowicka, and K. Kopańczyk, "Applying human mobility and water consumption data for short-term water demand forecasting using classical and machine learning models," *Urban Water J.*, vol. 17, no. 1, pp. 32–42, Jan. 2020, doi: [10.1080/1573062x.2020.1734947](https://doi.org/10.1080/1573062x.2020.1734947).
- [27] X. Liu, Y. Zhang, and Q. Zhang, "Comparison of EEMD-ARIMA, EEMD-BP and EEMD-SVM algorithms for predicting the hourly urban water consumption," *J. Hydroinformatics*, vol. 24, no. 3, pp. 535–558, May 2022, doi: [10.2166/hydro.2022.146](https://doi.org/10.2166/hydro.2022.146).
- [28] E. Pacchin, F. Gagliardi, S. Alvisi, and M. Franchini, "A comparison of short-term water demand forecasting models," *Water Resour. Manage.*, vol. 33, no. 4, pp. 1481–1497, Mar. 2019, doi: [10.1007/s11269-019-02213-y](https://doi.org/10.1007/s11269-019-02213-y).
- [29] Y. Gong, Z. Wang, G. Xu, and Z. Zhang, "A comparative study of groundwater level forecasting using data-driven models based on ensemble empirical mode decomposition," *Water*, vol. 10, no. 6, p. 730, Jun. 2018, doi: [10.3390/w10060730](https://doi.org/10.3390/w10060730).
- [30] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997.
- [31] K. Wang, Z. Ye, Z. Wang, B. Liu, and T. Feng, "MACLA-LSTM: A novel approach for forecasting water demand," *Sustainability*, vol. 15, no. 4, p. 3628, Feb. 2023, doi: [10.3390/su15043628](https://doi.org/10.3390/su15043628).
- [32] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," in *Proc. 31st Annu. Conf. Neural Inf. Process. Syst. (NIPS)*, Long Beach, CA, USA, Dec. 2017, pp. 1–11.
- [33] D. Li and Y. Fu, "Deep learning model-based demand forecasting for secondary water supply in residential communities: A case study of Shanghai city, China," *IEEE Access*, vol. 12, pp. 38745–38757, 2023, doi: [10.1109/ACCESS.2023.3288817](https://doi.org/10.1109/ACCESS.2023.3288817).
- [34] R. G. Perea, E. C. Poyato, and J. A. R. Díaz, "Attention is all water need: Multistep time series irrigation water demand forecasting in irrigation districts," *Comput. Electron. Agricult.*, vol. 218, Mar. 2024, Art. no. 108723, doi: [10.1016/j.compag.2024.108723](https://doi.org/10.1016/j.compag.2024.108723).
- [35] W. Wu and Y. Kang, "Ensemble empirical mode decomposition Granger causality test dynamic graph attention transformer network: Integrating transformer and graph neural network models for multi-sensor cross-temporal granularity water demand forecasting," *Appl. Sci.*, vol. 14, no. 8, p. 3428, Apr. 2024, doi: [10.3390/app14083428](https://doi.org/10.3390/app14083428).
- [36] T. B. Brown et al., "Language models are few-shot learners," 2020, *arXiv:2005.14165*.
- [37] S. Khan, M. Naseer, M. Hayat, S. W. Zamir, F. S. Khan, and M. Shah, "Transformers in vision: A survey," *ACM Comput. Surv.*, vol. 54, no. 10s, pp. 1–41, Jan. 2022, doi: [10.1145/3505244](https://doi.org/10.1145/3505244).
- [38] O. Dalmaz, M. Yurt, and T. Çukur, "ResViT: Residual vision transformers for multimodal medical image synthesis," *IEEE Trans. Med. Imag.*, vol. 41, no. 10, pp. 2598–2614, Oct. 2022, doi: [10.1109/TMI.2022.3167808](https://doi.org/10.1109/TMI.2022.3167808).
- [39] W. Wang, E. Xie, X. Li, D. P. Fan, K. Song, D. Liang, T. Lu, P. Luo, and L. Shao, "PVT v2: Improved baselines with pyramid vision transformer," *Comput. Vis. Media*, vol. 8, pp. 415–424, Mar. 2022, doi: [10.1007/s41095-022-0274-8](https://doi.org/10.1007/s41095-022-0274-8).
- [40] M.-H. Guo, T.-X. Xu, J.-J. Liu, Z.-N. Liu, P.-T. Jiang, T.-J. Mu, S.-H. Zhang, R. R. Martin, M.-M. Cheng, and S.-M. Hu, "Attention mechanisms in computer vision: A survey," *Comput. Vis. Media*, vol. 8, no. 3, pp. 331–368, Sep. 2022, doi: [10.1007/s41095-022-02711-y](https://doi.org/10.1007/s41095-022-02711-y).
- [41] K. Han, Y. Wang, H. Chen, X. Chen, J. Guo, Z. Liu, Y. Tang, A. Xiao, C. Xu, Y. Xu, Z. Yang, Y. Zhang, and D. Tao, "A survey on vision transformer," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 1, pp. 87–110, Jan. 2023, doi: [10.1109/TPAMI.2022.3152247](https://doi.org/10.1109/TPAMI.2022.3152247).
- [42] C. Zhang, L. Wang, S. Cheng, and Y. Li, "SwinSUNet: Pure transformer network for remote sensing image change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5224713, doi: [10.1109/TGRS.2022.3160007](https://doi.org/10.1109/TGRS.2022.3160007).
- [43] C. Zhang, W. Jiang, Y. Zhang, W. Wang, Q. Zhao, and C. Wang, "Transformer and CNN hybrid deep neural network for semantic segmentation of very-high-resolution remote sensing imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4408820, doi: [10.1109/TGRS.2022.3144894](https://doi.org/10.1109/TGRS.2022.3144894).

- [44] L. Wang, R. Li, C. Zhang, S. Fang, C. Duan, X. Meng, and P. M. Atkinson, "UNetFormer: A UNet-like transformer for efficient semantic segmentation of remote sensing urban scene imagery," *ISPRS J. Photogramm. Remote Sens.*, vol. 190, pp. 196–214, Aug. 2022, doi: [10.1016/j.isprsjprs.2022.06.008](https://doi.org/10.1016/j.isprsjprs.2022.06.008).
- [45] L. Sun, G. Zhao, Y. Zheng, and Z. Wu, "Spectral-spatial feature tokenization transformer for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5522214, doi: [10.1109/TGRS.2022.3144158](https://doi.org/10.1109/TGRS.2022.3144158).
- [46] J. Ma, L. Tang, F. Fan, J. Huang, X. Mei, and Y. Ma, "SwinFusion: Cross-domain long-range learning for general image fusion via Swin Transformer," *IEEE/CAA J. Autom. Sinica*, vol. 9, no. 7, pp. 1200–1217, Jul. 2022, doi: [10.1109/JAS.2022.105686](https://doi.org/10.1109/JAS.2022.105686).
- [47] Z. Liu, Y. Tan, Q. He, and Y. Xiao, "SwinNet: Swin Transformer drives edge-aware RGB-D and RGB-T salient object detection," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 7, pp. 4486–4497, Jul. 2022, doi: [10.1109/TCSVT.2021.3127149](https://doi.org/10.1109/TCSVT.2021.3127149).
- [48] A. Lin, B. Chen, J. Xu, Z. Zhang, G. Lu, and D. Zhang, "DS-TransUNet: Dual Swin Transformer U-Net for medical image segmentation," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–15, 2022, doi: [10.1109/TIM.2022.3178991](https://doi.org/10.1109/TIM.2022.3178991).
- [49] D. Hong, Z. Han, J. Yao, B. Gao, B. Zhang, A. Plaza, and J. Chanussot, "SpectralFormer: Rethinking hyperspectral image classification with transformers," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5518615, doi: [10.1109/TGRS.2021.3130716](https://doi.org/10.1109/TGRS.2021.3130716).
- [50] X. He, Y. Zhou, J. Zhao, D. Zhang, R. Yao, and Y. Xue, "Swin transformer embedding UNet for remote sensing image semantic segmentation," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4408715, doi: [10.1109/TGRS.2022.3144165](https://doi.org/10.1109/TGRS.2022.3144165).
- [51] M. Elsis, M. Tran, K. Mahmoud, D.-E.-A. Mansour, M. Lehtonen, and M. M. F. Darwish, "Effective IoT-based deep learning platform for online fault diagnosis of power transformers against cyberattacks and data uncertainties," *Measurement*, vol. 190, Feb. 2022, Art. no. 110686, doi: [10.1016/j.measurement.2021.110686](https://doi.org/10.1016/j.measurement.2021.110686).
- [52] Z. Chen, D. Chen, X. Zhang, Z. Yuan, and X. Cheng, "Learning graph structures with transformer for multivariate time-series anomaly detection in IoT," *IEEE Internet Things J.*, vol. 9, no. 12, pp. 9179–9189, Jun. 2022, doi: [10.1109/JIOT.2021.3100509](https://doi.org/10.1109/JIOT.2021.3100509).
- [53] V. Bagal, R. Aggarwal, P. K. Vinod, and U. D. Priyakumar, "Mol-GPT: Molecular generation using a transformer-decoder model," *J. Chem. Inf. Model.*, vol. 62, no. 9, pp. 2064–2076, May 2022, doi: [10.1021/acs.jcim.1c00600](https://doi.org/10.1021/acs.jcim.1c00600).
- [54] Y. Nie, N. H. Nguyen, P. Sinthong, and J. Kalagnanam, "A time series is worth 64 words: Long-term forecasting with transformers," 2022, *arXiv:2211.14730*.
- [55] A. Das, W. Kong, A. Leach, S. Mathur, R. Sen, and R. Yu, "Long-term forecasting with TiDE: Time-series dense encoder," 2023, *arXiv:2304.08424*.
- [56] Y. Liu, T. Hu, H. Zhang, H. Wu, S. Wang, L. Ma, and M. Long, "ITransformer: Inverted transformers are effective for time series forecasting," 2023, *arXiv:2310.06625*.
- [57] M. Xiangmei, T. Leping, Y. Chen, and W. Lifeng, "Forecast of annual water consumption in 31 regions of China considering GDP and population," *Sustain. Prod. Consumption*, vol. 27, pp. 713–736, Jul. 2021, doi: [10.1016/j.spc.2021.01.036](https://doi.org/10.1016/j.spc.2021.01.036).



ZHI-WEI TIAN is currently pursuing the bachelor's degree with the School of Mathematics and Statistics, Qinghai Minzu University, Xining, China. His research interests include machine learning and data mining.



RU-LIANG QIAN is currently pursuing the bachelor's degree with the School of Mathematics and Statistics, Qinghai Minzu University, Xining, China. His research interests include machine learning and data mining.

...