OPEN ACCESS

International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-13 Issue-8, July 2024

Transformer-Based Methods for Water Level Prediction: A Case Study of the Kien Giang River, Quang Binh Province

Bao Bui Quoc, Hung Nguyen Khanh, Hieu Nguyen Dac, Dat Tran-Anh, Quang Chieu Ta

Abstract. ***Accurate and timely water level prediction is of paramount importance in various applications, including flood forecasting, hydroelectric power management, and environmental monitoring. Traditional recurrent neural network (RNN)-based methods have been widely used for this task. However, recent advancements in long-term time-series forecasting have introduced transformer-based models that have significantly improved the performance in time-series prediction tasks. In this research, we investigate the application of transformer-based models to the task of water level prediction, specifically focusing on the Nhat Le River Basin. We conducted multiple experiments with different test cases and various model architectures, providing specific analyses of the model's prediction capabilities. The transformer-based models consistently outperformed conventional RNN-based methods across a range of evaluation metrics, including root mean square error (RMSE), mean absolute error (MAE), and coefficient of determination (R^2). Moreover, these models exhibited excellent flood peak prediction accuracy, with errors consistently below 0.02 meters. The robustness and scalability of transformer-based models make them promising for accurate water-level predictions in real-world applications.***

Keywords: Water level Prediction, Time Series Forecasting, Transformer-Based Models, Deep Learning

I. INTRODUCTION

The Central area of Vietnam is extremely susceptible to natural calamities, particularly cyclones and inundations. Between 1986 and 2006, data shows that storms caused the death of 76 people and injured 532 individuals. In addition, around 350,000 residences underwent modifications, and there were notable impairments to the infrastructure, as described 1. In early October 2020, there were concerning reports of flooding in Kien Giang (specifically in the Le Thuy area, as shown in Figure 1), which reached alarm level 3 and exceeded the previous record established in 1979. As a consequence, thousands of homes in Quang Binh 2 were flooded. These instances of extreme flooding frequently occur during the southwest monsoon season. Early detection of major flood events and accurate estimates of river levels are essential for implementing damage mitigation methods and responding immediately to assist in rescue operations. It is crucial to incorporate sophisticated modeling tools, such as artificial intelligence, into predictive models in order to achieve early enhancements in this area.

Conventional hydrological forecasting techniques frequently depend on historical data and basic models that may not fully capture the intricate and ever-changing characteristics of river systems. Recent breakthroughs in deep learning, specifically Transformer-based models, have demonstrated exceptional abilities in effectively managing long-sequence temporal data. As a result, these models are well-suited for enhancing the precision and dependability of water level predictions.

This work seeks to increase the accuracy and reliability of river water level forecasting systems by conducting a thorough investigation of historical water level data, utilizing state-of-the-art Transformer-based models, and employing rigorous validation methodologies. Our aim is to create a useful tool for disaster preparedness and sustainable management of water resources in the Quang Binh River basin and other comparable riverine regions globally.

The primary contributions of our work can be summarized as follows:

- We propose a comprehensive approach to anticipate water levels, which may be universally implemented across diverse geographical regions with distinct river systems.
- We employ Transformer-based models to effectively tackle the problem, consequently offering an assessment and analysis of prediction outcomes on a specific dataset in the Nhat Le River region in Quang Binh province.
- We present a collection of adaptable test cases that enable users to choose the most appropriate model for particular prediction tasks. Users have the ability to tailor the water level prediction model to meet their individual needs by adjusting various configurations, including input characteristics and hyperparameters. Users can utilize this flexibility to customize the Transformer-based method for various river systems and environmental situations.

Manuscript received on 30 June 2024 | Revised Manuscript received on 08 July 2024 | Manuscript Accepted on 15 July 2024 | Manuscript published on 30 July 2024.

**Correspondence Author*

Bao Bui Quoc, Hanoi University of Science and Technology, No. 1 Dai Co Viet, Hai Ba Trung, 100000, Hanoi, Vietnam. Email: Bao.BQ222002M@sis.hust.edu.vn, ORCID ID: 0000-0002-1158-9696

Hung Nguyen Khanh, Hanoi University of Science and Technology, No. 1 Dai Co Viet, Hai Ba Trung, 100000, Hanoi, Vietnam. Email: hung.nk194297@sis.hust.

Hieu Nguyen Dac, Department of Artificial Intelligence, Thuyloi University, 175 Tay Son, Dong Da, 100000, Hanoi, Vietnam. Email: dachieu@tlu.edu.vn.

Dat Tran Anh, Department of Artificial Intelligence, Thuyloi University, 175 Tay Son, Dong Da, 100000, Hanoi, Vietnam. Email: dat.trananh@tlu.edu.vn.

Quang Chieu Ta*, Department of Artificial Intelligence, Thuyloi University, 175 Tay Son, Dong Da, 100000, Hanoi, Vietnam. Email: quangchieu.ta@tlu.edu.vn, ORCID ID: 0009-0001-5079-7762

- We evaluate the performance of the proposed method through extensive experiments, comparing it against other existing approaches. The evaluation demonstrates that the Transformer-based method outperforms other methods in terms of prediction accuracy, as evidenced by the improved R^2 score and reduced RMSE and MAE metrics.

The remainder of this paper is organized as follows. In Section 2, we delve into the prior research related to water level prediction, encompassing various methodologies, including machine learning-based, data preprocessing, deep neural networks-based models, and Transformer-based approaches. Section 3 outlines our methodological approach, highlighting the utilization of the Informer and Autoformer models. In Section 4, we provide details regarding the dataset, experimental configuration, and metrics employed for evaluating model performance. We proceed to examine and discuss the outcomes of model comparisons. Lastly, in Section 5, we encapsulate our paper by presenting concluding remarks.

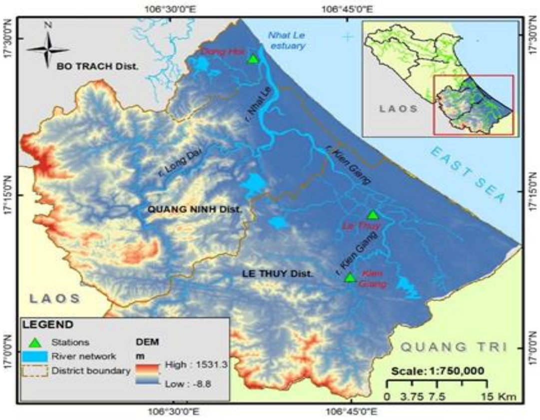

Fig. 1. Kien Giang River System and Location of Meteorological and Hydrological Stations [1]

II. RELATED WORD

Accurate prediction of river water levels is vital for effective water resource management and flood control. Over recent years, multiple approaches have been utilized to tackle the challenge of precise water level forecasting. In this section, we provide an overview of the existing body of literature related to machine learning-based models, deep neural network-based models, and Transformer-based models designed for water level prediction. Our focus is on understanding the strengths and limitations of these approaches.

A. Machine Learning-Based Models

In this part, we delve into the details of the different machine-learning approaches used in water level prediction.

Linear Regression [5] is a simple and widely used machine learning algorithm that assumes a linear relationship between the input variables (e.g., meteorological and hydrological factors) and the target variable (water level). It estimates the coefficients of the linear equation to minimize the difference between the predicted and actual values. Linear regression models are interpretable, allowing for the identification of the influence of individual features on water level prediction. However, they may struggle to capture complex nonlinear relationships.

Ensemble methods combine multiple individual models to improve prediction accuracy. Two commonly used ensemble methods are Random Forests (RF) [6] and Gradient Boosting Machines (GBM) [7]. Random Forests build a collection of decision trees and aggregate their predictions, while Gradient Boosting Machines iteratively train weak models and combine their predictions. Ensemble methods can handle nonlinear relationships, capture interactions between features, and reduce the risk of overfitting. They often provide robust and accurate predictions in water level forecasting tasks.

Time series analysis techniques, such as Autoregressive Integrated Moving Average (ARIMA) [8] and Seasonal Decomposition of Time Series (STL) [9], have been employed for water level prediction. ARIMA models capture the autocorrelation and seasonality in the data, making them suitable for short-term predictions. STL decomposes the time series into trend, seasonal, and residual components, allowing for separate modeling of each component. Time series analysis techniques are effective in capturing temporal dependencies and seasonality patterns in water level data.

While machine learning techniques possess the capability to learn from historical data and provide interpretable prediction results, they face challenges in capturing intricate nonlinear relationships within the data. Additionally, they demand substantial amounts of data and training time. Consequently, to enhance water level predictions, it is essential to explore synergies with other methods like modeling or time series analysis, aiming for greater prediction accuracy and reliability.

B. Deep Neural Network-Based Models

In recent years, deep neural network-based models have gained significant attention and recognition for their effectiveness in river water level prediction. These models leverage the power of artificial neural networks, particularly recurrent neural networks (RNNs) [16] and convolutional neural networks (CNNs) [13], to capture complex temporal patterns and nonlinear relationships within water-level data.

Recurrent Neural Networks (RNN) have proven to be formidable tools for time series forecasting, including river water level prediction. RNN architectures like Long Short-Term Memory (LSTM) [15,16] and Gated Recurrent Unit (GRU) [14] have become go-to choices due to their ability to model temporal dependencies effectively. LSTM, in particular, is well-suited for capturing long-range dependencies and has shown remarkable performance in modeling river water level fluctuations. The recurrent nature of these networks allows them to maintain internal states and remember past information, making them highly suitable for sequential data like time series.

Convolutional Neural Networks (CNN), originally designed for image analysis, have been adapted to handle time series data, including river water level records. These models are capable of extracting spatial and temporal features from input sequences, making them well-suited for capturing patterns in water level data.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

CNN-based models have demonstrated strong performance in capturing complex relationships and are particularly effective in situations where spatial dependencies exist within the data.

Hybrid Architectures: Researchers have also explored hybrid architectures that combine both RNNs and CNNs [12] to harness the strengths of both network types. These hybrid models aim to capture both short-term and long-term dependencies in the data. By utilizing the convolutional layers for feature extraction and the recurrent layers for sequential modeling, these architectures offer a comprehensive approach to river water level prediction.

While deep neural network-based models have shown promise in river water level prediction, they are not without challenges. Training these models can be computationally intensive, especially for large datasets, and may require significant computational resources. Additionally, addressing issues such as vanishing or exploding gradients is crucial to ensure model stability. The choice of architecture, hyperparameters, and regularization techniques also play a vital role in achieving optimal performance.

C. Transformer Based Models

Transformer-based models have recently gained significant attention for their ability to address complex challenges in time series prediction. They particularly excel in capturing intricate relationships among essential time series components.

The Informer model [4], which utilizes the ProbSparse self-attention mechanism, has significantly improved time complexity and memory utilization compared to the Transformer. However, despite these performance enhancements, there are still challenges related to time dependencies that result in disparities between predicted outcomes and actual results. On the other hand, the Autoformer model [3] expands and enhances the capabilities of the Informer model by introducing a self-correlation mechanism. This enhancement enables the model to better capture temporal dependencies compared to the traditional attention mechanism. Its primary goal is to accurately separate temporal trends and seasonal components within the time data. Furthermore, the FEDformer model [10] decomposes time series into multiple frequency-domain modes to extract features using the Fourier transform method. This approach has led to improved model performance, particularly on long sequences. Pyraformer [11], with its pyramidal attention module, consistently achieves superior prediction accuracy in both one-step and long-range forecasting tasks while maintaining minimal time and memory requirements, especially for lengthy sequences.

In our research, we undertake a comprehensive evaluation of these models to discern their individual con-tributions to the field of water level prediction.

III. METHODOLOGY

A. Problem Definition

In this paper, we tackle the challenge of predicting water levels at 03 meteorological and hydrological stations within a specific river region in Vietnam. We collect data from various stations, which we label as *1, 2, ..., S,* in the designated river area. These data encompass two key parameters: water levels and rainfall, denoted as H_t^s and R_t^s respectively $(H_t^s, R_t^s \in R)$. In this context, H represents the water level observed at station *s* $(s = \overline{1,S})$ at timestep *t*, measured in centimeters, while *R* signifies the precipitation recorded at station s at time *t*, expressed in millimeters. Our objective is to predict future water levels by constructing a network denoted as f. This network takes as input time series sequences spanning *m* time steps into the past and outputs forecasts for water levels *n* time steps into the future. Concretely, consider timestep *T*:

- **Input**. The input data utilizes time series sequences from specific water level and rainfall information, which are denoted as follows:

$$H_{in}^s = \left\{H_t^s\right\}_{t=T-m+1}^{T}, R_{in}^s = \left\{R_t^s\right\}_{t=T-m+1}^{T}$$

- **Output:**

$$H_{out}^s = \left\{H_t^s\right\}_{t=T+1}^{T+n}$$

B. Method

Based on the general pipeline in [17], we construct an overall flow for the water level prediction problem with four Transformer-based TSF solutions, as shown in Figure 2. In this context, details about the Data Preprocessing Processing Block will be presented specifically in Section IV.A.1. Regarding the two Encoder and Decoder blocks, the overall equation of the l-th layer can be summarized as follows:

$$\chi_{en}^l = Encoder(\chi_{en}^{l-1}) \qquad (1)$$

$$\chi_{de}^l = Decoder(\chi_{en}^{l-1}, \chi_{en}^N) \qquad (2)$$

1. Decomposed Transformer architecture

These approaches incorporate the concept of decomposition, which allows for the separation of time series into trend-cyclical and seasonal elements. Within every neural block, both the Decomp module and the EIU module (designed to enhance information utilization) form the core components in every encoder or decoder layer. To be more precise:

- Encoder. With χ_{en}^0 being the embedded historical series, serving as the input for the first layer; each layer $l \in \{1,\ldots,N\}$ has the general formula of the Encoder(·) as follows:

$$S_{en}^{l,1}, - = Decomp(EIU(\chi_{en}^{l-1}) + (\chi_{en}^{l-1}),$$
$$S_{en}^{l,2}, - = Decompt(FeedForward(S_{en}^{l,1}) + S_{en}^{l,1}),$$
$$\chi_{en}^l = S_{en}^{l,2},$$

where $S_{en}^{l,i}, i \in 1,2$ denotes the seasonal component obtained after the *i*-th decomposition block within the *l*-th layer, respectively.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

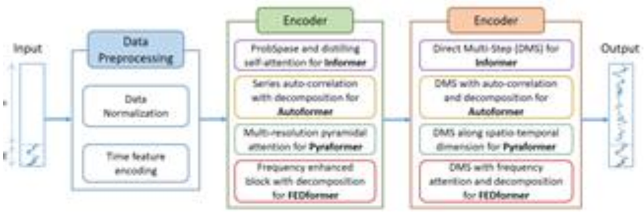

Fig. 2. Overall Process of the Transformer-Based LSTF Method for Water Level Prediction

- Decoder. The decoder in Equation 2 is formalized as:

$$S_{de}^{l,1}, T_{de}^{l,1} = Decomp(EIU(\chi_{de}^{l-1}) + (\chi_{de}^{l-1}),$$
$$S_{de}^{l,2}, T_{de}^{l,2} = Decomp(EIU(S_{de}^{l,1}, \chi_{en}^{N}) + S_{de}^{l,1}),$$
$$S_{de}^{l,3}, T_{de}^{l,3} = Decomp(FeedForward(S_{de}^{l,2}) + S_{de}^{l,2}),$$
$$\chi_{de}^{l} = S_{de}^{l,3},$$
$$T_{de}^{l} = T_{de}^{l-1} + W_{l,1}.T_{de}^{l,1} + W_{l,2}.T_{de}^{l,2} + W_{l,3}.T_{de}^{l,3}$$

Here, $S_{de}^{l,i}$ and $T_{de}^{l,i}$, where $i \in \{1,2,3\}$, denote the seasonal and trend components obtained after the i-th decomposition block in the l th layer, respectively. Additionally, $W_{l,i}$ with i ∈ {1, 2, 3}, represents the projector for the i-th extracted trend $T_{de}^{l,i}$.

2. Auto Former

Autoformer [3] is a Decomposition Based approach, utilizing the SeriesDecomp(.) module as the Decomp(.) component. This module progressively extracts the long term stationary trend from predicted intermediate hidden variables. Furthermore, it employs Auto-Correlation(.) as the EIU module, enabling the discovery of period-based dependencies by calculating series autocorrelation and aggregating similar subseries based on time delays.

3. FED Former

FEDformer [10] combines the Transformer architecture with signal processing techniques such as Fourier analysis and seasonal-trend decomposition. To be specific, FEDformer leverages the Encoder-Decoder architecture of the Transformer to depict intricate connections within the data. Significantly, the Frequency Enhanced Decomposed Transformer component fuses Fourier analysis and seasonal-trend decomposition to capture recurring cyclic patterns as well as the holistic distribution of time series.

The FED former utilizes the MOEDecomp (Mixture Of Experts Decomposition block) and serves as the Decomp module. Moreover, in the encoder, it employs FEB (Frequency Enhanced Block) as its EIU(.), while in the decoder, both FEA (Frequency Enhanced Attention) and FEB are employed as EIU mechanisms. The detailed descriptions of these modules are provided ex-plicitly in [10].

4. Informer

We employed the Informer [4] model to predict water levels, a model crafted on the foundation of the Transformer architecture [2]. The Transformer greatly improves predictive capabilities; however, it encounters hurdles such as quadratic memory consumption, temporal intricacies, and inherent constraints embedded within the encoder-decoder design when applied to Long Sequence Temporal Forecasting (LSTF). The Informer model has effectively overcome these challenges with three distinctive attributes.

Firstly, **the ProbSparse Self-Attention** method accomplishes memory utilization and temporal intricacy at $O(L \times \log L)$ simultaneously showcasing exceptional performance in aligning sequence dependencies. Instead of using the typical self-attention formula, we use the following formula by having each key only attend to the important query u:

$$A(Q,K,V) = Softmax(\frac{\bar{Q}K^T}{\sqrt{d}})V$$

In which Q is a $L_0 \times d$ sized matrix, K and V are key and value matrices and $\bar{Q}$ are sparse matrices of the same size dimension of Q, which only contains the Top-u queries.

Secondly, Self-attention distillation manages extensive input sequences by condensing information from subsequent layers to illuminate a dominant focal point of attention.

$$X_{j+1}^{k} = MaxPool\left(ELU\left(Conv1d\left(\left[X_j^k\right]_{AB}\right)\right)\right)$$

Where $[.]_{AB}$ signifies the attention block and X^k denotes the input sequence at time step k.

Lastly, a generative-style decoding technique is employed to anticipate lengthy time series sequences in a single forward direction, substantially boosting the speed of inference for prolonged time-series prediction.

5. Pyra Former

Pyraformer [11] emerges as an exceptional approach, boasting several salient advantages.

Firstly, it seamlessly amalgamates the capacity to efficiently embrace both short-term and long-term models through its ingenious pyramidal attention mechanism. In contrast to the conventional attention module**, the Pyramidal Attention Module** (PAM) utilizes a pyramidal graph to depict temporal relationships within the observed time series. To elaborate, assume $N_\ell^{(s)}$ represents the ℓ-th node at scale s, with s = 1, . . . , S denoting the hierarchy from the lowest scale to the highest scale in a sequential manner. In general, each node in the graph can focus on a restricted set of neighboring nodes $N_\ell^{(s)}$ at three different scales. Given this context, the attention computation at node n(s) can be simplified as follows:

$$y_i = \sum_{\ell \in N_\ell^{(s)}} \frac{\exp(q_i k_\ell^T / \sqrt{d_K}) \upsilon_\ell}{\sum_{\ell \in N_\ell^{(s)}} \exp(q_i k_\ell^T / \sqrt{d_K})}$$

Where q_i is the i-th row in the matrix Q, k_ℓ^T represents the transpose of row l in the matrix K. And N is the number of attention layers.

Secondly, Pyraformer deftly manages long-range temporal connections, effectively curtailing the length of the signal path. Lastly, it attains an ideal equilibrium between model capacity and complexity, concurrently economizing both time and computational space.

By virtue of these amalgamated strengths, Pyraformer holds the promise of making a substantial impact in the domain of time series prediction. It facilitates the adept modeling of intricate and remote relationships while mitigating the computational and memory burdens.

Retrieval Number: 100.1/ijitee.H993613080724
DOI: 10.35940/ijitee.H9936.13080724
Journal Website: www.ijitee.org

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-13 Issue-8, July 2024

IV. EXPERIMENTAL EVALUATION

A. Data Inputs and Experimental Settings

1. Data inputs

Dataset Information. The dataset for this task comprises rainfall and water level data collected from three hydro-meteorological stations (Le Thuy, Kien Giang and Dong Hoi) situated within the Nhat Le river basin (in this experiment, we denote these three stations as s = 1, 2, 3, respectively). Figure 3 presents a Box and Whisker plot illustrating a pronounced increase in water levels along the Kien Giang River during the period from September to December, aligning with the flood season in the research area.

The dataset used for training and validation was continuously collected, with hourly measurements recorded from January 1, 2020, at 00:00, to December 31, 2020, at 23:00, resulting in a total of 8780 data points. To facilitate model development and evaluation, these records were split into training, validation, and testing sets, with a distribution of 70% for training, 10% for validation, and 20% for testing.

Data Preprocessing. Initially, we systematically inspect the data for missing values and address any such instances. We employ forward-filling and backward-filling techniques, wherein empty elements are replaced with the values of the preceding or succeeding elements, respectively. Once we confirm the data's integrity, we standardize it before commencing the training phase. To be precise, the data is adjusted to have a mean of 0 and a standard deviation of 1, as an example of normalizing the water level data at station s -

$$H^{s}_{\tau} = \frac{H^{s}_{\tau} - \mu}{\sigma}$$

Here, $\tilde{H}^s_\tau$ represents the normalized value, H^s_τ stands for the initial feature value, μ denotes the feature's mean value across the data sequence, and σ indicates the feature's standard deviation within the data sequence.

2. Implementation details

Evaluation Test Cases. We carried out experiments using the following six test cases:

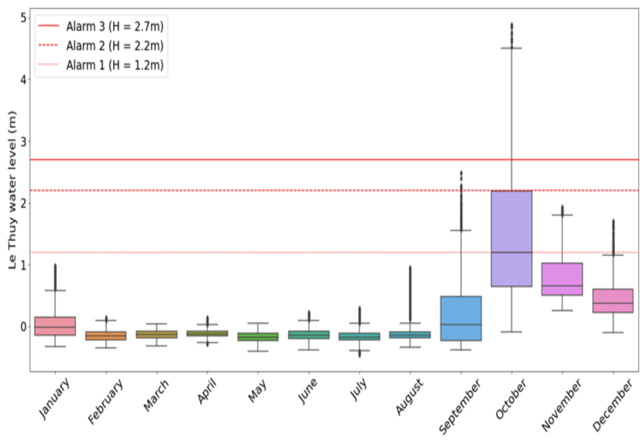

Fig. 3. Average Water Levels Recorded Hourly at Le Thuy's Station in the Years 2010, 2012, 2016, and 2020, along with three alarm Thresholds

Table-I: Statistics on the Mean and Standard Deviation of the Fields in the Dataset

	Water Level (m), Rainfall (mm)					
	WL_KG	**RF_KG**	**WL_LT**	**RF_LT**	**WL_DH**	**RF_DH**
μ	5.8827	0.0590	-0.1141	0.0395	-0.0109	0.0704
σ	0.1596	0.6654	0.1309	0.4383	0.3061	0.8083

- TC1: Utilize hourly water level at 01 station from t m time lags to t.
- TC2: Utilize hourly water level at 03 stations from t m time lags to t;
- TC3: Utilize hourly water level at 01 station, rainfall at 01 stations from t m time lags to t.
- TC4: Utilize hourly water level at 03 stations, rainfall at 03 stations from t m time lags to t.
- TC5: Utilize hourly water levels at 03 stations, and rainfall at 01 station from the time interval t m to t.
- TC6: Utilize hourly water level at 01 station, rainfall at 03 stations from t − m time lags to t.

Here, m and n represent the count of time lags (1hr, 4hr, 9hr) and time leads (1hr, 3hr, 6hr, 12hr).

Implementation Configurations. The model was trained on an NVIDIA Tesla T4 GPU with 15GB of RAM memory. Throughout the training process, the utilization of patience and batch-size management has been demonstrated to be crucial in enhancing performance and maintaining training stability. We opted for a batch size of 256 data samples for each weight update iteration in our model. Additionally, we configured the patience parameter to a value of 5, and the maximum epoch is 50, signifying that, following an assessment of the model's performance on the validation dataset, if the model demonstrates no improvement for three consecutive epochs, the training process will be terminated. This application of patience ensures that the model doesn't overtrain post-convergence without significantly boosting accuracy on the validation dataset. This approach contributes to time savings in training and guards against overfitting. We implemented a learning rate of 1e-4, guided by observing smooth loss curves during the model training phase. In all experiment evaluations, we chose the station for water level prediction to be Le Thuy station (due to its significant role in terms of geography). This selection guarantees a smooth and efficient progression in the model's training process.

Loss Function. We utilize the Mean Squared Error (MSE) as our loss function. MSE computes the average of the squared deviations between the predictions made by our model and the actual observed values.

$$MSE = \frac{1}{k}\sum_{\tau=1}^{k}(H^{s}_{\tau} - H^{s}_{\tau})$$

In this context, $\hat{H}_t$ denotes the real water level at a specific time τ, and H_t stands for the corresponding predicted value, while k is the number of data points. By employing MSE as our loss function, our objective is to fine-tune our models to minimize the average squared deviations between predictions and actual values, ultimately enhancing our capacity to predict river water levels with precision.

Retrieval Number: 100.1/ijitee.H993613080724
DOI: 10.35940/ijitee.H9936.13080724
Journal Website: www.ijitee.org

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

Table-II: Test Cases for the Model

Test case	Inputs		Output
	Water Level at 03 Stations	Rainfall at 03 Stations	
TC 1	H^s_{in}, $s \in \{1,2\ 3\}$	-	
TC 2	H^s_{in}, $s = \{1,2,3\}$	-	
TC 3	H^s_{in} , $s \in \{1,2,3\}$	R^s_{in} , $s \in \{1,2,3\}$	H^s_{out}
TC 4	H^s_{in} , $s = \{1,2,3\}$	R^s_{in} , $s \in \{1,2,3\}$	$s \in \{1,2,3\}$
TC 5	H^s_{in} , $s \in \{1,2,3\}$	R^s_{in} , $s = \{1,2,3\}$	
TC 6	H^s_{in} , $s = \{1,2,3\}$	R^s_{in} , $s = \{1,2,3\}$	

3. Evaluation Metric

To evaluate the performance of our model, we utilize three evaluation criteria: R-squared, Root Mean Square Error(RMSE), and Mean Absolute Error(MAE).

R-squared (R^2): Quantifies the goodness of fit be-

tween a predictive model and observed data. A high R2 value (close to 1) signifies that the model effectively captures the data's variability. Conversely, when R2 equals 0, the model performs no better than a simplistic model that merely predicts the data's mean.

$$R^2 = 1 - \frac{\sum_{\tau=1}^{k}\left(H^s_\tau - \overline{H}^s_\tau\right)^2}{\left(H^s_\tau - \overline{H}^s_\tau\right)^2}$$

where $\overline{H}^s_\tau$ is the mean of the dependent variable across all data points.

RMSE (Root Mean Square Error): RMSE measures the average magnitude of the errors between predicted and actual values. Lower RMSE values indicate better model performance.

$$RMSE = \sqrt{\frac{\sum_{\tau=1}^{k}(H^s_\tau - \overset{s}{H_\tau})^2}{k}}$$

MAE (Mean Absolute Error): MAE is another measure of the average magnitude of errors, but it doesn't square the errors as RMSE does. Instead, it calculates the mean of the absolute differences between predicted and actual values. Like RMSE, lower MAE values indicate better model performance.

$$MAE = \frac{\sum_{\tau=1}^{k}|H^s_\tau - \overset{s}{H_\tau}|}{k}$$

B. Experimental Results

Our practical experiments in this subsection are geared towards tackling the following crucial research questions.

- RQ1: What factors influence the forecasted water level at the Le Thuy station, including the input data and the consideration of various data fields?
- RQ2: What's the comparison between the experimental results of transformer-based models and certain baseline prediction models?
- RQ3: How does the model's prediction capability vary over time lead and time lag intervals?
- RQ4: How accurately does the Autoformer model's prediction align with the real value?

1. The impact of test cases on model performance. (RQ1)

The primary objective of this experiment is to observe the dependence or correlation between different data fields and their impact on the model's predictive capabilities for the Le Thuy station. Geographically, the three measurement stations are located on three main river branches (including Kien Giang, Long Dai, and Nhat Le) within the larger Nhat Le river basin, with a total area of 2,612 km2. Therefore, water level and rainfall data at the Kien Giang and Dong Hoi (abbreviated as KG and DH) stations may be correlated and have an influence on the data at the Le Thuy station (abbreviated as LT). In the various test cases, TC1 entails the model learning past water level data at LT for future predictions. TC2 incorporates the additional learning of water level data from KG and DH, while TC3 utilizes both water level and rainfall data from LT. Notably, TC6 integrates all six mentioned data fields. Furthermore, TC4 employs four data fields, encompassing water level data from the three stations and rainfall data at LT, whereas TC5 is designed to assess the influence of rainfall at all three stations. The results in Table 3 demonstrate that the R2 scores for all 6 test cases are consistently above 0.99, indicating that the model has learned well and fits the training data effectively. Notably, in TC1, the results for all three evaluation metrics are superior compared to the other cases, while TC2 to TC6 exhibit relatively minor differences. From this result table, we can speculate that, for this specific dataset, adding additional data fields may introduce noise to the model during training and prediction. One contributing factor to this phenomenon could be the large standard deviations observed in certain fields, as described in Table 1.

However, it's important to note that each problem and dataset possesses unique characteristics. Therefore, to configure the model for optimal performance, it is necessary to carry out such test cases, evaluate and compare the results specifically for each dataset.

2. Comparision with Baseline Models (RQ2)

In this study, we compare the outcomes of transformer based approaches against three different RNN-based deep learning models: Gated Recurrent Unit (GRU), Long Short-Term Memory (LSTM), and Bidirectional Long-Short Term Memory (Bi-LSTM). It's worth noting that these three techniques have also demonstrated favorable results in the realm of Long-Term Series Forecasting, particularly when dealing with hourly data.

All models listed in the table demonstrate exceptionally high R2 values, ranging from 0.986 to 0.999. This indicates that the models excel in explaining the variability in the observed data, with Autoformer achieving the highest R2 score, suggesting an almost perfect fit to the data. Furthermore, when it comes to RMSE and MAE metrics, Informer and Autoformer surpass RNN-based models. Specifically, the RMSE metric experiences a 4-fold reduction ($0.061 \rightarrow 0.015$), and the MAE metric witnesses a 5-fold decrease ($0.040 \rightarrow 0.008$) when comparing Autoformer to GRU (the top-performing RNN-based model).

In conclusion, the table offers a clear and informative comparison of different models' performance, highlighting Autoformer and Informer as top-performing choices for water level prediction in the context of our experiments.

Retrieval Number: 100.1/ijitee.H993613080724
DOI: 10.35940/ijitee.H9936.13080724
Journal Website: www.ijitee.org

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

These results are highly promising and suggest the potential for improved water level forecasting using advanced transformer-based models.

3. (RQ3)

One of the advantages of transformer-based models lies in their capacity to make long-term forecasts.

The table illustrates this capability by presenting the prediction outcomes of the Autoformer model with different time lags (m = 4, 6, 9) and time leads (n = 1, 3, 6). Furthermore, the table exhibits specific prediction results for three stations: Kien Giang, Le Thuy, and Dong Hoi (it's important to note that RNN-based models' long-term forecasts yielded unsatisfactory results, with Autoformer surpassing them; as a result, detailed results for RNN-based models are omitted from this table to simplify it).

- We can observe that when n = 1, the results for all three stations are favorable.
- For the same n value, as m increases, the model's performance at KG and LT tends to decline, unlike DH. This indicates that the data for KG and LT rely more on nearby data points, whereas DH requires a longer historical time frame.
- With the same m value, as the prediction horizon n increases, the model's predictive ability gradually diminishes.
- Among the stations, LT consistently delivers the best and most consistent results when varying the m and n values. In contrast, KG and DH exhibit significant drops in performance as these values increase. This can be partially attributed to the greater standard deviation across the entire dataset for these two stations, coupled with increased data volatility. These factors directly impact the model's forecasting capability.

4. Qualitative Study (RQ4)

The primary purpose of addressing the water level forecasting issue is to offer assistance and notifications to authorized personnel, including alerting them when the water level surpasses predefined alarm thresholds. As evident from the data presented in Chart 1, there has been a notable and sudden surge in water levels during the last quarter of the year. An effective model should not only exhibit favorable evaluation metrics but also make accurate predictions, even when faced with atypical data points like flood peaks. Chart 4 illustrates the predictive lines generated by the autoformer model for three corresponding test cases, alongside the actual observed values (here, we use 3 test cases for easier tracking and visualization). The data displayed spans from October 18, 2020, at 06:00:00 to October 20, 2020, at 08:00:00, with the peak water level reaching 4.88 meters. The model's forecasts closely match the actual observed line and accurately anticipate the timing of peak levels. Notably, TC1 produces exceptional results both in terms of evaluation metrics and in its visual resemblance to the actual observed line. This chart underscores the model's remarkable adaptability, even when dealing with uncommon data points.

Table-III: Test Case Comparison (lags m = 4, leads n = 1)

Model	TC1	TC2	TC3	TC4	TC5	TC6
R^2	0.9996	0.9978	0.9972	0.9976	0.9973	0.9979
RMSE	0.0152	0.0421	0.0476	0.0447	0.0467	0.0417
MAE	0.0079	0.0237	0.0304	0.0266	0.0211	0.0224

Table-IV: Model Performance Comparison (lags m = 4, leads n = 1)

Model	R^2	RMSE	MAE
LSTM [18][19][20][21]	0.993	0.083	0.050
GRU	0.996	0.061	0.040
Bi-LSTM	0.986	0.121	0.074
Informer	0.998	0.030	0.011
Autoformer	0.999	0.015	0.008
FEDformer	0.999	0.019	0.010
Pyraformer	0.998	0.025	0.014

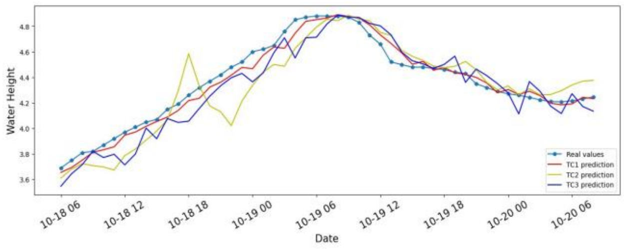

Fig. 4. Actual and Predicted Water Level

V. CONCLUSION

In this study, we conducted an extensive exploration of the application of transformer-based models for water level prediction in the Nhat Le River Basin. Through a rigorous series of comparative experiments between the transformer and RNN architectures, we have unambiguously validated the superior performance of transformers for this task. Their accuracy, adaptability, and robustness firmly establish them as a promising tool for advancing real-world flood forecasting and water resource management systems. While our focus was on the Nhat Le River Basin, the tremendous potential shown by transformers suggests they can achieve even better results on other river networks. Our future work will expand these experiments to include new river basins, diverse types of supplementary input data, and different time series datasets. In conclusion, transformers open up an exciting new avenue for modeling and predicting time series data.

DECLARATION STATEMENT

Funding	No, we did not receive.
Conflicts of Interest	No conflicts of interest to the best of our knowledge.
Ethical Approval and Consent to Participate	No, the article does not require ethical approval and consent to participate with evidence.
Availability of Data and Material	Not relevant.
Authors Contributions	All authors having equal contribution for this article.

REFERENCES

1. Ly N, Duong N, and Dai N, "Khi hau va thuy van tinh Quang Binh", Science and Technics Publishing House Science and Technics Publishing House, 2013
2. Vaswani, Ashish, et al. "Attention is all you need." Advances in neural information processing systems 30 (2017).

3. H. Wu, J. Xu, J. Wang, and M. Long, "Autoformer: Decomposition Transformers with Auto-Correlation for Long-Term Series Forecasting," Adv. Neural Inf. Process. Syst., vol. 27, no. NeurIPS, pp. 22419-22430, 2021.
4. Haoyi Zhou and Shanghang Zhang and Jieqi Peng and Shuai Zhang and Jianxin Li and Hui Xiong and Wancai Zhang, "Informer: Beyond Efficient Transformer for Long Sequence TimeSeries Forecasting," The Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2021, Virtual Conference, vol. 35, 11106-11115, 2021. https://doi.org/10.1609/aaai.v35i12.17325
5. Montgomery, Douglas C., Elizabeth A. Peck, and G. Geoffrey Vining. Introduction to linear regression analysis. John Wiley & Sons, 2021.
6. Breiman, Leo. "Random forests." Machine learning 45 (2001): 5-32. https://doi.org/10.1023/A:1010933404324
7. Friedman, Jerome H. "Greedy function approximation: a gradient boosting machine." Annals of statistics (2001): 1189-1232. https://doi.org/10.1214/aos/1013203451
8. Zhang, G. Peter. "Time series forecasting using a hybrid ARIMA and neural network model." Neurocomputing 50 (2003): 159-175. https://doi.org/10.1016/S0925-2312(01)00702-0
9. Ouyang, Zuokun, Philippe Ravier, and Meryem Jabloun. "STL decomposition of time series can benefit forecasting done by statistical methods but not by machine learning ones." Engineering Proceedings 5.1 (2021): 42. https://doi.org/10.3390/engproc2021005042
10. Zhou, Tian, et al. "Fedformer: Frequency enhanced decomposed transformer for long-term series forecasting." International Conference on Machine Learning. PMLR, 2022.
11. Liu, Shizhan, et al. "Pyraformer: Low-complexity pyramidal attention for long-range time series modeling and forecasting." International conference on learning representations. 2021.
12. Athiyarath, Srihari, Mousumi Paul, and Srivatsa Krishnaswamy. "A comparative study and analysis of time series forecasting techniques." SN Computer Science 1.3 (2020): 175. https://doi.org/10.1007/s42979-020-00180-5
13. Mehtab, Sidra, and Jaydip Sen. "Analysis and forecasting of financial time series using CNN and LSTM-based deep learning models." Advances in Distributed Computing and Machine Learning: Proceedings of ICADCML 2021. Springer Singapore, 2022. https://doi.org/10.1007/978-981-16-4807-6_39
14. Yamak, Peter T., Li Yujian, and Pius K. Gadosey. "A comparison between arima, lstm, and gru for time series forecasting." Proceedings of the 2019 2nd international conference on algorithms, computing and artificial intelligence. 2019. https://doi.org/10.1145/3377713.3377722
15. Elsworth, Steven, and Stefan Güttel. "Time series forecasting using LSTM networks: A symbolic approach." arXiv preprint arXiv:2003.05672 (2020).
16. Salinas, David, et al. "DeepAR: Probabilistic forecasting with autoregressive recurrent networks." International Journal of Forecasting 36.3 (2020): 1181-1191. https://doi.org/10.1016/j.ijforecast.2019.07.001
17. Zeng, Ailing, et al. "Are transformers effective for time series forecasting?." Proceedings of the AAAI conference on artificial intelligence. Vol. 37. No. 9. 2023.
18. Hieu, T. T., Chieu, T. Q., Quang, D. N., & Hieu, N. D. (2023). Water level prediction using deep learning models: A case study of the Kien Giang River, Quang Binh province. River, 1–12. https://doi.org/10.1002/rvr2.63
19. Dakshin, D., Rupesh, V. R., & Kumar, S. P. (2019). Water Hazard Prediction using Machine Learning. In International Journal of Innovative Technology and Exploring Engineering (Vol. 9, Issue 1, pp. 1451–1457). https://doi.org/10.35940/ijitee.a4245.119119
20. Singh, N., & Panda, S. P. (2020). Stimulating Deep Learning Network on Graphical Processing Unit To Predict Water Level. In International Journal of Engineering and Advanced Technology (Vol. 9, Issue 4, pp. 1222–1229). https://doi.org/10.35940/ijeat.d8452.049420
21. Perunkolam, A., Baskaran, M. P., Tripathi, S., & Sahu, O. P. (2020). Smart Water Distribution System using Machine Learning and IoT. In International Journal of Recent Technology and Engineering (IJRTE) (Vol. 9, Issue 2, pp. 1189–1194). https://doi.org/10.35940/ijrte.b4115.079220

AUTHORS PROFILE

Bui Bao Quoc received his engineering degree in Applied Mathematics and Informatics from Hanoi University of Science and Technology in 2021. Currently, he is pursuing a Master's degree at Hanoi University of Science and Technology. Since 2023, he has been a research fellow at Vin University. His research interests include image recognition, deep learning, and reinforcement learning.

Hung Nguyen Khanh, a final-year Engineering student at Hanoi University of Science and Technology (HUST), is specializing in Computer Vision and Artificial Intelligence. He is a Research and Development Intern at Mobi Fone Information Technology Center. At Mobi Fone, Hung is actively involved in developing and researching Computer Vision solutions, focusing on image recognition, deep learning, reinforcement learning, and edge device processing.

Hieu Nguyen Dac is a lecturer and researcher in the Department of Artificial Intelligence at Thuyloi University. He received his Master's degree in Information Technology at Ha Noi University of Science and Technology. His research focuses on natural language processing, predictive analytics, and big data.

Dat Tran-Anh received his engineering degree in Information Security from the Posts and Telecommunications Institute of Technology in 2020 and his Master of Science in Computer Science from the Posts and Telecommunications Institute of Technology, Vietnam in 2022. Currently, he is a PhD student in Computer Science at the Vietnam Academy of Science and Technology. His research interests include image recognition and deep learning.

Quang Chieu Ta is a lecturer and researcher in the Department of Artificial Intelligence at Thuyloi University. He received his Ph.D. degrees in Computer Science from the University of Tours, France in 2015. His current research focuses on machine learning, deep learning, data analytics and big data.