AGU International Journal of Sciences – 2024, Vol. 11 (5), 81 – 87

AGU INTERNATIONAL JOURNAL OF SCIENCES
ISSN 0866 - 8086

WATER LEVEL PREDICTION AT CHAU DOC MONITORING STATION, AN GIANG PROVINCE

Tran Ngoc Chau[1], Nguyen Thi My Truyen[1]
[1]*An Giang University, VNU-HCM*

Information:
Received: 03/07/2024
Accepted: 25/10/2024
Published: 12/2024

Keywords:
Water Level, Water Level Prediction, Machine Learning, LSTM, Chau Doc Monitoring Station

ABSTRACT

In the Mekong Delta, flooding is an annual natural phenomenon that occurs on a large scale, affecting human life, agriculture, infrastructure, the economy, and society. On the other hand, during the dry season, water levels drop, leading to water shortages and drought. This study investigates the application of machine learning to predict water levels at Chau Doc monitoring station in An Giang province. Specifically, this study uses LSTM algorithms combined with random forests to predict water levels. The model was trained using water level data collected from 2019 to 2022 and is designed to predict levels until 2031. The observed data was used to implement the model and assess its accuracy, efficiency, and reliability. The results show that the random models achieved a Nash-Sutcliffe Efficiency (NSE) greater than 0.7 and a Root Mean Square Error (RMSE) between approximately 5 and 10 cm, showing that the model is both reliable and effective in simulating water levels. Local water quality management authorities have applied various measures, and early predictions of water level changes in the river help management agencies respond proactively to upcoming situations related to the water environment in the area. Moreover, this research serves as a useful reference for further studies and in management of water quality.

1. INTRODUCTION

Currently, to cope with floods, high tides and droughts, many solutions, policies and investments have been made for management, warning systems and water level monitoring in rivers and canals. One of the tools frequently used today for water level prediction is machine learning. However, in An Giang province, machine learning has not yet been applied to predict water levels, nor has research been conducted on using such models at Chau Doc monitoring station.

Water levels prediction using machine learning methods allows predictions without requiring detailed knowledge of complex hydrological processes, unlike traditional models. Additionally, the results produced by machine learning model are highly accurate with low error rates. This approach is also cost-effective and has great potential for water levels prediction based on data collected at monitoring stations. In particular, An Giang province possesses a large source of digital data from monitoring stations, with high reliability, which is a valuable asset for experts.

In recent years, various studies have ultilized physical simulation and calculation models such as HEC-RAS, MIKE-11, MIKE-21, SOBEK, EFDC in the field of environment and hydrometeorology. For example, research in the lower Sai Gon - Dong Nai River used the MIKE 11 model to calculate hydraulics, combined with the Delft3D oceanographic model to calculate tidal levels and meteorological models to forecast rainfall, temperature and evaporation (Bao Thanh, Tran Tuan Hoang & Ngo Nam Thinh, 2013).

Water level prediction research at the Quang Phuc and Cua Cam stations in Hai Phong also used the LSTM model to simulate hourly water levels at hydrological stations, predicting 1 to 5 hours excluding gas data. The model exdudes climate and terrain factors, resulting in highly accurate predictions. The 3 hour water level prediction in advance achieved an NSE > 0.987 and an RMSE value < 0.10 m at both stations (Le Xuan Hien & Ho Viet Hung, 2018). In addition, machine learning models were also studied for water levels on the Mekong River to compare the effectiveness of different machine learning models. The research focused on the machine learning models LASSO, RF (Random Forest), and SVR models at Thakhek station on the Mekong River (Tung Nguyen Thanh, Quynh Nguyen Huu & Mark Junjie Li, 2015).

Around the world, many studies have applied machine learning models for water level prediction at various scales. Typically, the LSTM network model has been used to predict flood water levels on the Bangladesh River based on spatial and temporal factors. The study used 5 machine learning models including ANN, LSTM, spatial attention LSTM, temporal attention LSTM and spatiotemporal attention LSTM (STAALSTM) (Fahima Noor et al., 2022). In predicting lake water levels using machine learning, researchers compared two models VARMAX and ARIMA (AutoRegressive Integrated Moving Average). At Lake Timah Tasoh, Northern Malaysia, deep learning with artificial neural networks was used to compare 2 models using error metrics such as MAE and RMSE and they tested 12 algorithms. The results showed that VARVAX had the highest RMSE value using a time-series data set that includes seasonal factors (Mohammad Amimul Ihsan Aquil1 & Wan Hussain Wan Ishak, 2023).

Another study applied the LSTM model to predict urban river water levels in real time in Fuzhou city, China, using a data set of water level observations from April to October 2020, comparing and evaluating data. The results showed that LSTM can effectively predict short-term river water levels in the study area (Yu Liu, Hao Wang, Wenwen Feng & Haocheng Huang, 2021).

In general, machine learning tools are widely researched and applied in predicting water levels of rivers, lakes and canals globally. However, when applying these tools for water level predictions at monitoring stations, it is important to pay attention to the following key characteristics. First, the data used for the training process must exhibit a strong correlation between the prediction goals and the characteristics of the collected data set. A lack of sufficient data can impact the training process and the final results. Additionally, algorithm performance including algorithms will produce different results with different prediction parameters. There are many machine learning algorithms such as artificial neural network (ANN), recurrent neural network (RNN), Support Vector Machine (SVM), Support Vector Regression (SVR), linear regression (LR), Random Forest (RF), XGBoost, LSTM (a special extension of RNN, known as long short-term memory neural network).

2. METHODOLOGY

2.1 Machine learning

The water level data set collected from Chau Doc monitoring station in the study has the characteristics of a time series data set, with the values exhibiting linear variability. Therefore, the research uses three algorithms including LSTM and Random forest. The implementation was done using Python 3.9, along with libraries such as scikit-learn, Tensorflow of LSTM and RandomForestRegressor of Random Forest.

2.2 Research sample

The water level data set, recorded hourly was collected from 2019 to 2022 at the Chau Doc monitoring station in An Giang province.

2.3 Research process

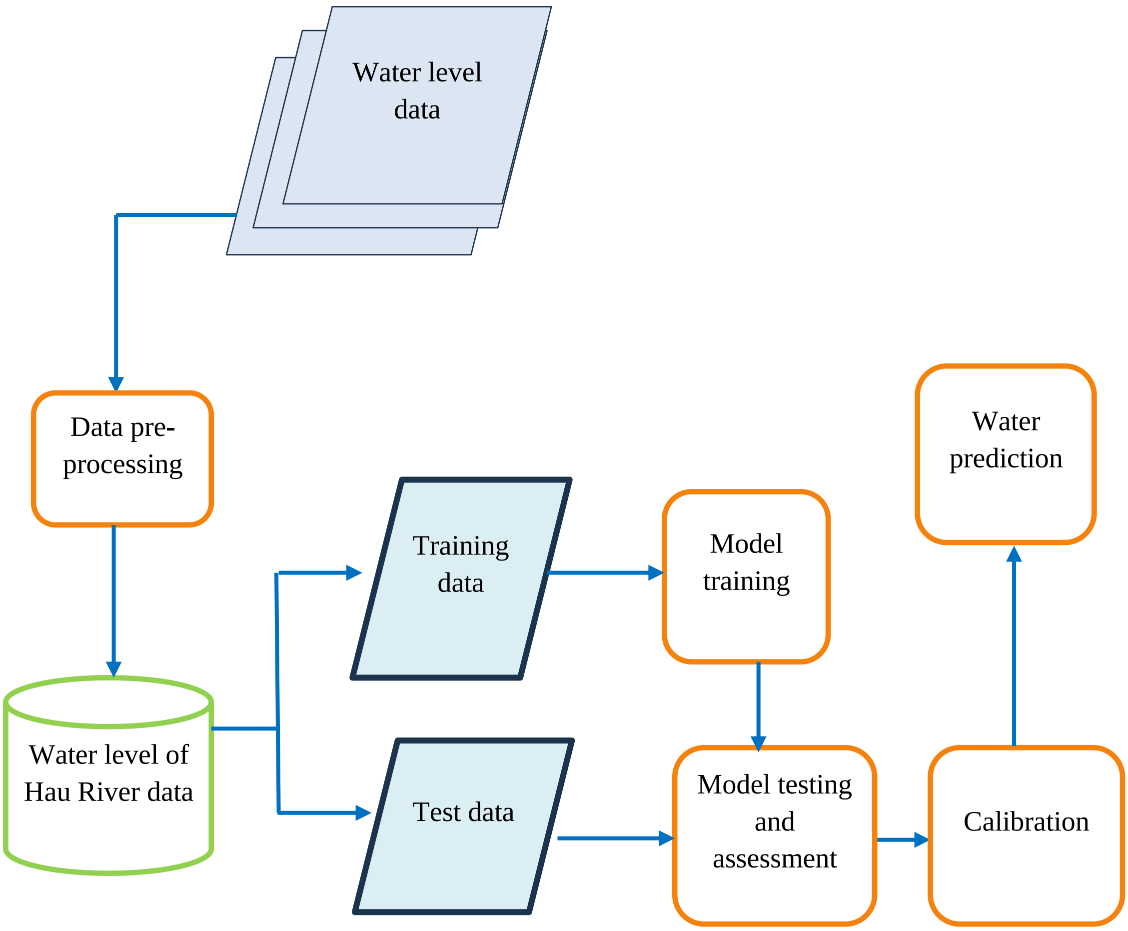

Figure 1. Research process

The process of researching and building machine learning models to predict water levels includes 6 main steps as follows.

Step 1. Data collection

The data set of water level at Hau River is recorded hourly from monitoring stations in chronological order covering each day over 12 months, and was collected from 2019 to 2022 at the Chau Doc monitoring station. The algorithm is trained across three stages including the dry season, the rising phase of the flood season and the receding phase of the flood season.

Step 2. Processing the training data set

Group the data into training and test sets with a ratio of 70 - 30. For both data sets, create the structure: "Output (window_size = 24, number_feature = 1 => input (number_output = 1)". Then, use the MinMaxScaler function from the scikit-learn library to normalize the data to the interval [0;1]. Finally, refactor the training and testing data to match the LSTM model input.

Step 3. Building a machine learning program and training the model

Table 1. Model design of LSTM and Random forest

<table>
<tr><td rowspan="2">Models</td><td>LSTM</td><td><pre>### Config lstm
model = Sequential()

model.add(InputLayer((window_size, n_features)))
model.add(LSTM(units=n_neurons))
model.add(Dense(n_predict, 'relu'))
model.add(Dense(n_predict, 'linear'))</pre></td></tr>
<tr><td>Random forest</td><td><pre># Config and compile model
params = {
        'n_estimators': 230,
        # 'max_depth': 4,
        # 'subsample': 0.7,
        # 'learning_rate': 0.08,
        'random_state': 0,
        # 'objective': 'reg:squarederror'
        'bootstrap': True,
        'oob_score': True
        }
model_rf_dry = RandomForestRegressor(**params)
model_rf_flood_inc = RandomForestRegressor(**params)
model_rf_flood_dec = RandomForestRegressor(**params)</pre></td></tr>
</table>

Step 4. Model checking and assessment

Using the test data set to check the accuracy of the model after training. Then use the NSE and RMSE index to perform validation testing.

Step 5. Calibration

Step 6. Water level prediction

The prediction results are tested using NSE and RMSE. The results are then compared between the different models.

3. RESULT AND DISCUSSION

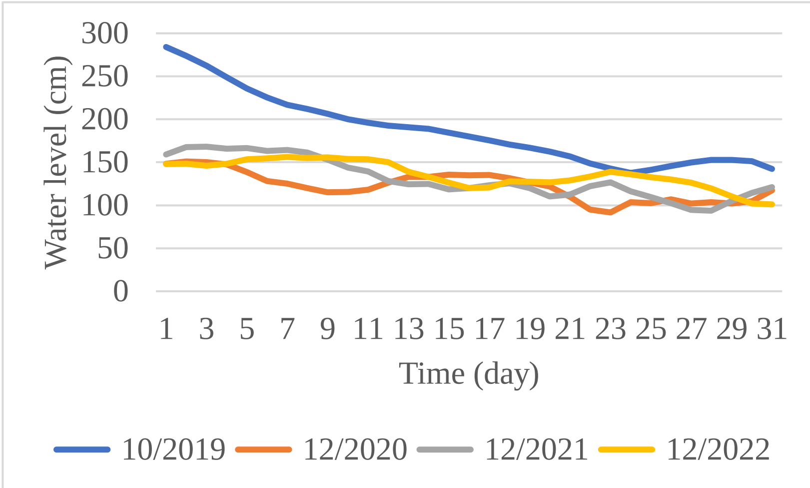

Figure 2. Average daily water levels at the Chau Doc monitoring station during the dry season of 2019, 2020, 2021 and 2022

According to the results in Figure 2, the water level in October 2019 was much higher than in other years, with a maximum of 284 cm on the 1st and a minimum of 138 cm on the 23rd. In December 2020, the water level fluctuated but remained the lowest compared to other years, reaching a low of 92 cm on the 23rd and a high of 148 cm. For December 2021, the average high water level reached 168 cm on the 2nd and 3rd, while the lowest water level was 94 cm on the 28th. Finally, in December 2022, the average high water level reached 156 cm on the 7th and 9th, with a notably low level of 101 cm on the 31st. On average, the water levels in 2020, 2021 and 2022 fluctuated only slightly, while the water level in 2019 remained relatively high throughout the monitoring period.

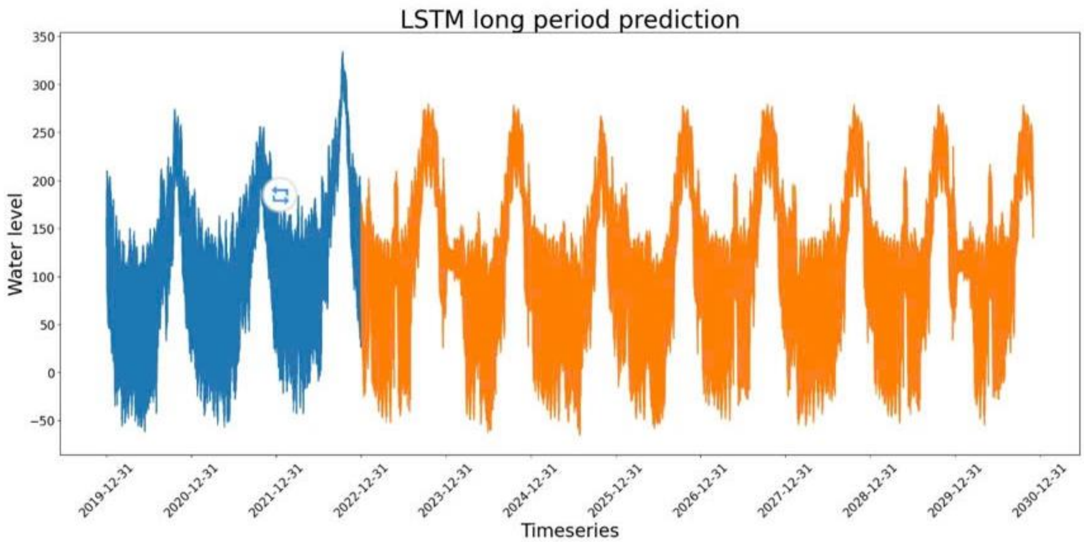

Figure 3. Water level prediction results in 2023 – 2031 period of LSMT model

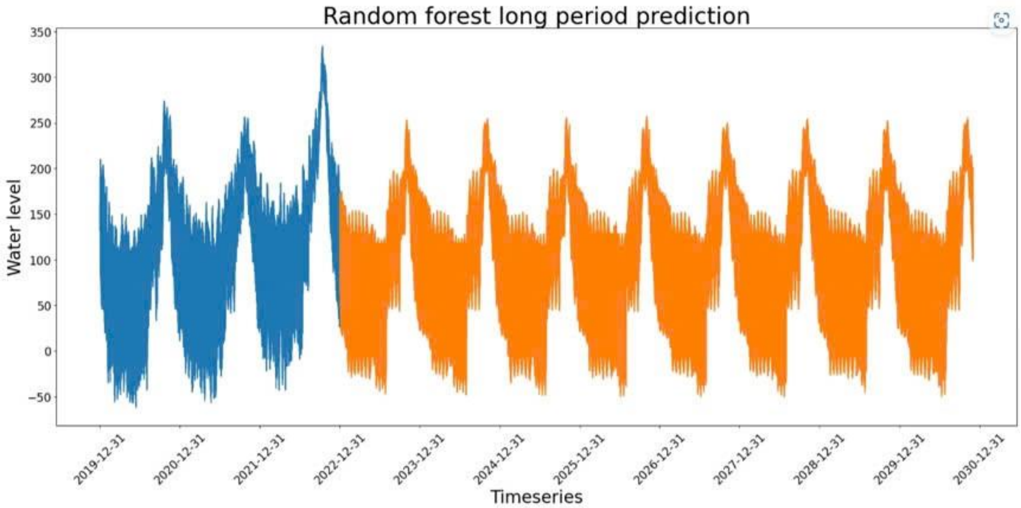

Figure 4. Water level prediction results in 2023 – 2031 period of Radom Forest model

According to Morias (2015), a value of NSE > 0.75 means the model has excellent performance, NSE > 0.6 shows the model is effective and NSE > 0.4 represents acceptable model results. For RMSE, the lower the RMSE value, the better the model performance, which indicates that the model is closer to the real value. The model results (Figure 3 and Figure 4), show it can be seen that the water level fluctuates between -50 to 350 cm, with the highest recorded level reaching about 350 cm during the period from 2019 to 2022, as represented by the blue line indicating the actual water level in both models. Long-term prediction for water level from 2023 to 2031 is shown by the red line on both models, showing that the LSTM and Random Forest models have a slightly high RMSE during the appraisal phase ranging from about 5 to 10. Therefore, these predictions may not accurately reflect reality. However, the NSE of the model > 0.7, showing that the model is reliable and effective in simulating water levels.

4. CONCLUSION

The water level predictions at Chau Doc Monitoring Station from 2019 to 2031 are based on actual data collected at the monitoring station during the 2019-2022 period. The recorded water levels from 2019 to 2021 ranged from 92 to 284 cm. Notably, the year 2019 had significantly higher water levels than the other years, with the highest recorded level reaching 284 cm. In general, fluctuations in water levels during 2020, 2021 and 2022 were not markedly different, but in 2020, the lowest recorded water level was 92 cm. According to the model results, both LSMT and Random Forest models achieved NSE > 0.7, proving their reliability, sustainability and effectiveness. However, the RMSE value, reaching from about 5 to 10, shows that the models may not closely reflect reality. This study demonstrates that using machine learning to forecast water level the Hau River at the Chau Doc monitoring station is completely appropriate; however, further research is needed to improve the capabilities of the models.

REFERENCES

Bao Thanh, Tran Tuan Hoang and Ngo Nam Thinh (2013). Testing a water level prediction model at some stations on the Saigon - Dong Nai River. *Hydrometeorological Journal*, December 2023.

Le Xuan Hien and Ho Viet Hung (2018). Applying LSTM network to forecast water levels at Quang Phuc and Cua Cam stations, Hai Phong City, Vietnam. *Water engineering science and environment*, 62(September 2018).

Tung Nguyen Thanh, Quynh Nguyen Huu and Mark Junjie Li (2015) . Forecasting Time series water levels on Mekong river using machine learning models. *Seventh International Conference on knowledge and Systems Engineering.*

Noor, F., Haq, S., Rakib, M., Ahmed, T., Jamal, Z., Siam, Z.S. andRahman, R.M. (2022). Water level forecasting using Spatiotemporal attention-based long short-term memory network. *Water 2022*, 14, 612. https://doi.org/10.3390/w14040612.

Mohammad Amimul Ihsan Aquil1 and Wan Hussain Wan Ishak (2023). Comparison of Machine learning models in forecasting reservoir water level. *Journal of Advanced Research in Applied Sciences and Engineering Technology*.

Yu Liu, Hao Wang, Wenwen Feng and Haocheng Huang (2021). Short term real-time rolling forecast of Urban River water levels based on LSTM: A case study in Fuzhou city, China. *Environmental Research and Public Health*.

D. N. Moriasi, M. W. Gitau, N. Pai and P. Daggupati (2015). Hydrologic and Water quality models: Performance measuresand Evaluation criteria. *Transactions of the ASABE (American Society of Agricultural and Biological Engineers)*, 58(6), 1763-1785. http://dx.doi.org/10.13031/trans.58.10715.