JOURNAL OF DATA SCIENCE | Vol.2024:53
eISSN:2805-5160

Water Level Prediction of Riam Kanan Dam Using ConvLSTM, BPNN, Gradient Boosting, and XGBoosting Stacking Framework (CLBGXGBoostS)

Usman Syapotro[1], Haldi Budiman[2], M. Rezqy Noor Ridha[3], Noor Azijah[4]

[1,2,3,4]Faculty of Information Technology, Islamic University of Kalimantan Muhammad Arsyad Al-Banjari, Indonesia

Email: 05usman.syapotro.study@gmail.com[1], haldibudiman@uniska-bjm.ac.id[2], rezqyridha@gmail.com[3], noorazijah444@gmail.com[4]

Abstract

Research focuses on developing a water level prediction framework for the Riam Kanan Dam using an innovative stacking approach called ConvLSTM-BPNN-Gradient Boosting and Stacking XGBoost (CLBGXGBoostS), which combines the strengths of Convolutional Long Short-Term Memory (ConvLSTM), Backpropagation Neural Network (BPNN), and Gradient Boosting. The study aims to evaluate the performance of the CLBGXGBoostS stacking framework in predicting the water level of the Riam Kanan Dam using 5 years of historical data. The results demonstrate that the CLBGXGBoostS framework provides more accurate predictions compared to single models, as evidenced by the Root Mean Squared Error (RMSE) values. CLBGXGBoostS achieves an RMSE of 0.0071, significantly lower than the RMSE of the individual models ConvLSTM (0.1006), BPNN (0.2618), and Gradient Boosting (0.6905). This research contributes to the development of a better water level prediction framework for the Riam Kanan Dam, supporting more effective water resource management and serving as a reference for future research in this field.

Keywords

Water level prediction, Riam Kanan Dam, ConvLSTM-BPNN-Gradient Boosting and XGBoost Stacking (CLBGXGBoostS)

Introduction

The Riam Kanan Dam, located in Banjar Regency, South Kalimantan, is one of the largest reservoirs in the region. Built by damming eight rivers originating from the Meratus Mountains, this dam plays a crucial role in providing water for hydroelectric power generation, irrigation, and flood control (Wikipedia, 2024). Given the importance of its functions, accurate water level prediction is essential for effective water resource management and disaster risk mitigation.

Previous research has explored various methods for predicting the water level of the Riam Kanan Dam. (Haldi Budiman, 2016) employed the Cascade Neural Network algorithm and reported good accuracy with a Root Mean Square Error (RMSE) of 0.0956. Meanwhile, (Ikhwani, 2018) implemented the Adaptive Neuro Fuzzy Inference System (ANFIS) by

Submission: 22 October 2024; **Acceptance:** 23 November 2024

comparing several membership functions and achieving the best RMSE of 0.010065. While these studies demonstrate the potential of artificial intelligence methods in water level prediction, there is still room for improvement in accuracy and the development of a more precise framework.

The novelty of this research is the development of a water level prediction framework for the Riam Kanan Dam by combining three base models: Convolutional Long Short-Term Memory (ConvLSTM), Backpropagation Neural Network (BPNN), and Gradient Boosting, within a stacking framework. This innovative stacking approach, called CLBGXGBoostS, leverages ConvLSTM's ability to extract spatial and temporal features, BPNN's flexibility in modeling, and Gradient Boosting's accuracy enhancement, with Extreme Gradient Boosting (XGBoost) used as a meta-model to combine the predictions from the three base models. This framework is expected to yield more accurate predictions compared to previous frameworks. Improved prediction accuracy will provide significant benefits in dam management, including optimization of power generation, more efficient irrigation scheduling, and enhanced preparedness for potential floods. Additionally, this research contributes to the development of water level prediction methods in general, which can be applied to other dams in Indonesia and globally.

Remainder of this paper consists of methodology, results and discussion, and conclusions and recommendations. The methodology section discusses the research methodology employed, including data collection methods, data processing, and framework development. The results and discussion section presents the experimental results, which are analyzed in-depth to gain a comprehensive understanding. The final section, conclusions and recommendations, presents the conclusions of this research along with suggestions for future research.

Methodology

The methodology employed is experimental and focuses on predicting water levels using a framework that combines ConvLSTM, BPNN, Gradient Boosting, and XGBoost Stacking (CLBGXGBoostS). The following is an explanation of the experimental stages:

Data Collection and Preprocessing:

The process begins with collecting 5 years of historical data related to the water level of the Riam Kanan Dam, which will be used for training and testing the prediction framework. The initial stage involves loading the historical water level data from an Excel file. The 'Month Year' column is set as the index and converted to the appropriate date format. The data is then normalized using MinMaxScaler to standardize the scale of the features. The final step in this stage is to divide the data into two parts: a training set (60%) to be used for training the framework and a testing set (40%) to evaluate the framework's performance.

Data Preparation:

During the data preparation stage, the dataset is constructed using the create_dataset function, accounting for timesteps. Timesteps specify how many past months should be utilized to predict the water level in the upcoming month. Subsequently, separate training and testing datasets are prepared for the ConvLSTM and BPNN models. Specifically for ConvLSTM, the data is reshaped to match the input format required by the model.

ConvLSTM Model Development and Training:

ConvLSTM model is constructed by combining ConvLSTM2D layers to capture spatial and temporal patterns, a Flatten layer to flatten the output, and a Dense layer to generate predictions. The 'mean_squared_error' loss function and the 'adam' optimizer are then used to construct the model. EarlyStopping is used in the training process to avoid overfitting. Once training is complete, the model is used to make predictions on the testing set. As the research conducted by (Nikunjkumar B. Nayak, 2024) highlights the superior performance of the CLSTM model in weather forecasting tasks, surpassing both LSTM and BiLSTM models, as evidenced by its lower RMSE value of 0.15836.

BPNN Model Development and Training:

BPNN (Backpropagation Neural Network) model is constructed with interconnected Dense layers. BPNN is a multi-layer feed forward network trained by the back propagation of error to optimize artificial neural networks (Hu et al., 2019). The compilation and training process of this model is similar to that of ConvLSTM, including the use of 'adam' as the optimizer, 'mean_squared_error' as the loss function, and the implementation of EarlyStopping to avoid overfitting. After training, the BPNN model is also used to generate predictions on the testing set.

Gradient Boosting and XGBoost Stacking:

XGBoost is one of the most influential and high-performance machine learning techniques. XGBoost implemented in Python uses the sci-kit-learn library and the Extreme Gradient Boosting class for machine learning regression tasks with the method showing superior performance compared to other methods (Wijayanti et al., 2024). Therefore, this combination is also done to predict the results generated by ConvLSTM and BPNN are returned to their original scale through denormalization to ensure the prediction results are within the appropriate range of the original data before further evaluation. Furthermore, Gradient Boosting, an ensemble technique, is applied to combine the predictions from ConvLSTM and BPNN.

Final Evaluation:

The three basic models are used as input features and the actual values are used as targets in the final step of training the XGBoost meta-model. The goal of this training procedure is to increase the final forecasts' accuracy even more. The final phase involves assessing the framework's performance by figuring out the Root Mean Squared Error (RMSE), which quantifies how accurately the framework forecasts water levels. Greater performance in predictions is shown by a lower RMSE score.

Results and Discussion

Experimental results of the ConvLSTM-BPNN-Gradient Boosting and XGBoost Stacking (CLBGXGBoostS) framework:

Table 1 Experimental Results

Model	Dataset Split	RMSE
ConvLSTM	60% Train, 40% Test	0.1006
BPNN	60% Train, 40% Test	0.2618
Gradient Boosting	60% Train, 40% Test	0.6905
Stacking (CLBGXGBoostS)	**60% Train, 40% Test**	**0.0071**

Based on the RMSE results of the ConvLSTM-BPNN-Gradient Boosting and XGBoost Stacking (CLBGXGBoostS) framework with an RMSE of 0.0071, it can be concluded that this framework demonstrates excellent performance in predicting water levels. There is minimal difference between the real and anticipated values, as evidenced by the incredibly low RMSE value. This shows that patterns and trends in the data can be accurately captured by the framework.

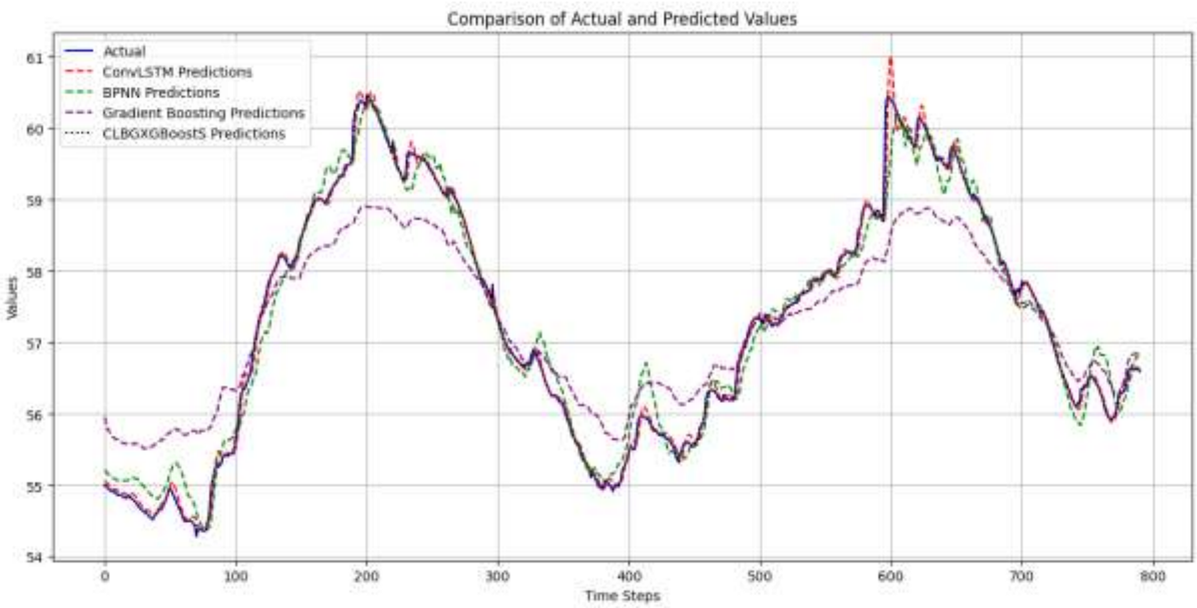

Figure 1 Graph of Test Results

Description:
Blue : Actual water level values
Red : Prediction from the ConvLSTM framework
Green : Prediction from the BPNN framework
Purple : Prediction from the Gradient Boosting framework (combined ConvLSTM & BPNN)
Black : ConvLSTM-BPNN-Gradient Boosting and XGBoost Stacking (CLBGXGBoostS)

The graph of the test results in Figure 1, "Comparison of Actual and Predicted Values", provides a clear visualization of the performance of the proposed framework. It can be seen that the prediction line from the ConvLSTM-BPNN-Gradient Boosting and XGBoost Stacking

(CLBGXGBoostS) framework closely follows the pattern of the actual value line. Fluctuations and trend changes in the data are accurately predicted by this framework.

Conclusions and Recommendations

This research has successfully developed an innovative water level prediction framework for the Riam Kanan Dam, namely ConvLSTM-BPNN-Gradient Boosting and XGBoost Stacking (CLBGXGBoostS), capable of providing high prediction accuracy by capturing complex patterns and trends in the data. The implementation of this framework contributes to water resource management and disaster risk mitigation at the Riam Kanan Dam by supporting decision-making related to dam operations such as water discharge regulation, power generation, and flood control.

Recommendations for further research are:
1. The developed CLBGXGBoostS framework should be implemented in the Riam Kanan Dam monitoring system with periodic performance monitoring.
2. Focus on testing the framework in other dams and developing a long-term prediction framework to support strategic water resource management planning.

References

Haldi Budiman, Y. I. dan M. R. (2016). PREDIKSI TINGGI MUKA AIR BENDUNGAN RIAM KANAN MENGGUNAKAN ALGORITMA CASCADE NEURAL NETWORK. *Prosiding Penelitian Dosen UNISKA MAB*, *66*(Hall 2009), 37–39. http://dx.doi.org/10.31602/ppdu.v0i1.8111

Hu, Y., Li, J., Hong, M., Ren, J., Lin, R., Liu, Y., Liu, M., & Man, Y. (2019). Short term electric load forecasting model and its verification for process industrial enterprises based on hybrid GA-PSO-BPNN algorithm—A case study of papermaking process. *Energy*, *170*, 1215–1227. https://doi.org/10.1016/j.energy.2018.12.208

Ikhwani, Y. (2018). Prediksi Tinggi Muka Air Bendungan Riam Kanan Menggunakan Adaptive Neuro Fuzzy Inference System. *Technologia: Jurnal Ilmiah*, *9*(1), 11. https://doi.org/10.31602/tji.v9i1.1097

Nikunjkumar B. Nayak. (2024). Comparative Analysis of Weather Forecasting Using LSTM, BiLSTM, and CLSTM Networks. *Journal of Electrical Systems*, *20*(3), 1511–1517. https://doi.org/10.52783/jes.3557

Wijayanti, E. B., Setiadi, D. R. I. M., & Setyoko, B. H. (2024). Dataset Analysis and Feature Characteristics to Predict Rice Production based on eXtreme Gradient Boosting. *Journal of Computing Theories and Applications*, *1*(3), 299–310. https://doi.org/10.62411/jcta.10057

Wikipedia. (2024). *Waduk Riam Kanan*. https://id.wikipedia.org/wiki/Waduk_Riam_Kanan