Forecasting Water Inflow for Bang Lang Dam

Akmal Benhawan
Management of Information Technology
Faculty of Engineering,
Prince of Songkla University
Songkla, Thailand
6010121025@psu.ac.th

Sureena Matayong
Management of Information Technology
Faculty of Engineering,
Prince of Songkla University
Songkla, Thailand
sureena.m@psu.ac.th

Chidchanok Choksuchat
ICT, Division of Computational Science
Faculty of Science,
Prince of Songkla University
Songkla, Thailand
chidchanok.ch@psu.ac.th

2022 IEEE 12th Annual Computing and Communication Workshop and Conference (CCWC) | 978-1-6654-8303-2/22/$31.00 ©2022 IEEE | DOI: 10.1109/CCWC54503.2022.9720865

***Abstract*— The information about water inflow for Bang Lang Dam Thailand is very important to the Pattani river management in order to manage the built multipurpose water storage dam for power generation and agriculture in the area. The management of water can be more efficient if the accurate forecasting of the water inflow into the dam is acknowledged. This study applied the H2O deep learning model, which is a multi-layer feed-forward Artificial Neural Network (ANN) to create the model for water inflow forecasting. The data was daily from January 1, 2012 to December 31, 2020. The model performed accuracy with MAE: 1.300, RMSE: 3.111, R^2: 0.767, R: 0.876 which provided a good result.**

Keywords—deep learning, forecasting, water inflow, Bang Lang Dam

I. Introduction

Water is a natural resource that is essential to life, economic and social development. The natural water used by people of all areas, including atmospheric water (rain), surface water and groundwater. These are a natural product exists in nature by itself, which we cannot produce more or reduce the amount as needed. As a result, some years there may be dearth, causing less water in rivers and streams, which unable to share thoroughly. However, some other years, it rains continuously until it is damaged the property and community areas due to flooding. Also, the presence of waste water or water pollution occurring in many areas can cause the crisis of water.

Water management is a process and method used to manage water resources, including procuring, developing, allocating, and using them for various purposes. It is also including the conservation and restoration of water resources for sustainable use such as solving problems caused by water resources in both quantity and quality aspects. Water management must be coordinated and integrated into many ways in order to make the most of water resources. At present, the Pattani river management of Thailand facing problems of water management from various factors, including global warming. The dearth during El Niño La Niña has resulted in drought problems as well as repeated floods over the years. This causes damage to life and property especially the agricultural areas are affected, which results in loss of productivity and other economic growth opportunities. In spite of the development of reservoirs, basins, dams, and numerous large, medium and small irrigation systems. In fact, the water storage and irrigation systems lack effective water management. This causes too little water in the storage that aggravates the dearth and finally causes a shortage of water for consumption and agriculture generally impacts on ecosystems downstream and vice versa. On the other hand, if there is too much water in the event of heavy and long rainfall, it will cause flooding to aggravate in a wide area including the crisis of the dam break that the damage is huge.

Pattani River Basin is located in the south of Thailand. It has a total watershed area of 3,684.21 sq km. Most of the area covers Yala and Pattani provinces and some parts of Songkhla and Narathiwat. The watershed is a longitudinal line, positioned along the north-south line. Pattani river is the main river and has the Yaha river as a tributary. In the end, there is a canal Nong Chik separated from the Pattani river with many other small canals. The Pattani river originates from the Sanglakiri Mountain Range in Betong district, Yala province, which flows from the south to the north and flows into the Gulf of Thailand at Mueang district Pattani province. Most of the area is forested with a slightly flat area. The lower part of the basin is a plain. The length of the river is about 210 kilometers.

Bang Lang Dam is a multipurpose dam that blocks the Pattani river, located at Ban Bang Lang. Khuean Bang Lang subdistrict, Bannang Sata district, Yala province. It is considered the first multi-purpose dam project in the south. The Electricity Generating Authority of Thailand is the main culprit of the construction, which was completed in June 1981. Since then, the Bang Lang Dam reservoir can benefit in the field of irrigation for farmland in Yala and Pattani provinces. It is an area of more than 380,000 hectares, where the released water can be used to generate about 200 million units of electric power per year, helping to promote the power system in the south to be more stable. It is a by-product that has enormous benefits. It also helps alleviate flooding in the lower area of the Pattani River Basin that has always happened. This area is an important inland fishery in the southern region that helps to provide the job and income for the people living near the dam. Moreover, it is a tourist attraction as well.

From such information, it can be seen that Bang Lang Dam is very important for the Pattani river management. This is because it is an important water reservoir of the Pattani river that affects irrigation, electricity generation, water allocation for agriculture and including protection against floods and dearth. Thus, effective water management is required. The key factor in determining the amount of water inflow into a dam can come from many factors such as rainfall, overflow or other weather conditions, which will affect the total amount of water in the dam. This research focuses on studying and analyzing factors affecting the amount of water inflow into the dam by applying the H2O deep learning model, a multi-layer feed-forward Artificial Neural Network (ANN) to create the model for water inflow forecasting. The created model can be created a smart report of water inflow information for the maximum benefit of water management in the Bang Lang Dam.

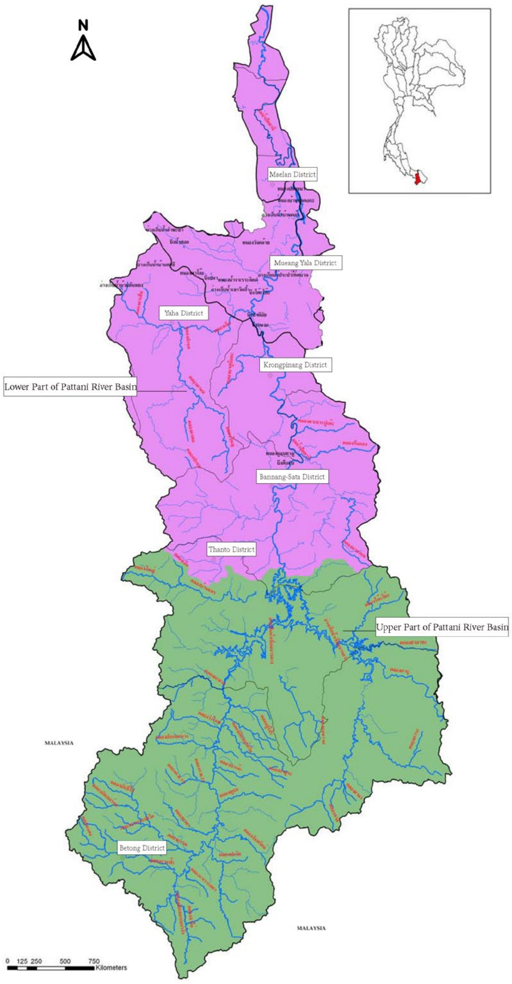

Fig. 1. Map of Pattani river basin

II. Related Works

A. Literature Review

Jehangir Ashraf Awan and Deg-Hyo Bae [1] studied system for forecasting the amount of water inflow into the dam by long-term forecasting by using Adaptive Neuro-fuzzy Inference System (ANFIS), In this short term forecasting of water inflow into dams is used for flood protection systems. This is a real-time data processing, but long-term forecasting of water inflow into the dam is used for planning years in advance. For example, in agricultural planning, water use, and hydroelectric power generation, This study applied ANFIS principles for long-term planning by studying rainfall data and the internal temperature of 60 data collection stations. The datasplit into two models: Model A uses normal rainfall, Model B uses average rainfall from forecasts, which the result shows Model B is more accurate.

Michele Pini et al [2] studied the evaluation of machine learning techniques for inflow prediction in lake Como, Italy. The study compared Linear Regression (LR), Random Forests (RF), Support Vector Regression (SVR) and Artificial Neural Networks (ANNs) techniques. The experimental results show that ANN is better estimate streamflow.

Li-Chiu Chang et al [3] developed the Intelligent Hydro informatics Integration Platform (IHIP) to create a flood warning system by showing the results as a flood map using machine learning techniques. The data was analyzed from flowing water, rainfall and current rainfall. The model was created a display of a flood forecast map, the current water volume in the form of web service on all platforms.

Fernando Salazar et al [4] reviewed the forecasting literature on dams based on past work. 7 types of dam forecasting and analysis techniques were organized: 1) Hydrostatic-Seasonal-Time (HST) Models 2) Model to Account for Delayed Effects 3) Auto Regression Models 4) Neural Networks 5) Adaptive Neuro- Fuzzy Inference (ANFIS), 6) Principal Component Analysis (PCA) and Dimensionality Reduction, and 7) Other Machine Learning (ML) techniques such as Support Vector Machine (SVM), K-Nearest Neibhbors (KNN), Genetic Algorithms (GA) and Ramdom Forest. A summary of 59 literature from 1985-2015 was summarized into two points: ML techniques are important in dam safety management and engineering-based judgments are necessary to make safe decisions.

Mahmood A. Khan, Md Zahidul Islam and Mohsin Hafeez [5] studied the accuracy of forecasting irrigation water demand from six different data mining models. To compare the performance of DTs, ANNs, Systematically Developed Forest of Multiple Trees (SysFor), SVM, Logistic Regression and traditional Evapotranspiration (ET) methods. The factors include maximum-minimum temperature, wind speed, humidity, rainfall and solar radiation. As a result, SysFor provides the most accurate performance at 97.5%.

T. Egawa, et al. [6] studied a dam inflow forecasting model using NNs and Regression to manage dam water safety in both normal and flooding seasons by forecasting the amount of water in advance within 1 day from the available information. Two models were obtained in two types of dams: regression model was used in dams without tributaries because it is not complicated, easy to understand. The NNs are used to forecast dams with tributaries because there are more factors involved.

Deg-Hyo Bae, Dae Myung Jeong and Gwangseob Kim [7] studied dam water inflow forecasting using ANFIS forecasting data to optimize the forecasting of water inflow into Soyanggang Dam, South Korea from the old system that uses only historical data. The researchers compared the models from the dataset that does not use weather forecast data with the weather forecast dataset. The results showed that the forecast weather data had higher correlation coefficient values and lower Root Mean Square Error (RMSE) values.

Noel B. Elizaga, Elmer A. Maravillas, and Bobby D. Gerardo [8] studied a model for forecasting water inflow into dams. The exponential smoothing time series method was compared with the NN backpropagation of Angat Dam, Philippines using the past 5 days water volume data to forecast daily water volume. It turns out that exponential smoothing model gives r=0.852 while backpropagation model gives r=0.959.

Ayesha Nayab and Muhammad Faisal [9] studied water management in the Tarbela Pakistan during torrential rain (15 June to 31 September) by building a water forecasting model using Bayesian's Stochastic Dynamic Programming statistical technique. The researchers studied and compared it with the model obtained from forecasting the inflow of water Autoregression and ANN. The study found that Bayesian had lower Mean Absolute Error (MAE) values, which the model used to determine water release policies during the flood season.

Arsalan Mahmoodzadeh, et al. [10] conducted a comparative best prediction model of water inflow into drill and blast tunnel by collecting water inflow data from 13 road tunnel construction sites in Iran. Data collected from the depth of drill and blast tunnels, groundwater content, rock quality. The study of prediction model using various machine learning techniques such as Long Short-Term Memory (LSTM), Deep Neural Networks (DNN), KNN, Gaussian Process Regression (GPR), Support Vector Regression (SVR) and Decision Tree (DT) to forecast water inflow data. Forecast results were measured using R^2, MAE, and RMSE values. The most accurate RMSE was 4.07486, of which DNN, GPR, SVR, KNN, and DT were the second most accurate, respectively, with RMSE values 4.66526, 5.77216, 12.95589, 16.63670, and 17.99058.

Di Zhang, et al. [11] studied modeling and simulating of reservoir operation using ANN, SVR and LSTM based on the water retention data of the Gezhouba dam across Yangtze river in China where the longest retention data exists. The study used monthly, daily and hourly data basis to compare the models. As a result, LSTM gave the most accurate water forecasting results in all cases of the time scale and water flow volume dimensions; monthly, daily and hourly scales for the time scale dimension, and low, intermediate and high for water flow volume dimensions.

Zachary C. Herbert, Zeeshan Asghar, and Carlos A. Oroza [12] studied long-term reservoir inflow forecasts: optimizing water use and water management using deep learning. The study applied Snow Water Equivalent (SWE) and reservoir time-series data inflow during April-July period between 1990 – 2020 of the Upper Stillwater Reservoir, Utah in the USA. The investigators improved the optimal parameters between LSTM and Convolutional Neural Network (CNN) techniques using Seasonal Autoregressive Integrated Moving Average (SARIMA) statistical techniques to optimize the efficiency of LSTM and CNN techniques. The accurate result disclose a simple LSTM-LSTM was the best performing model.

From the review, it was found that the forecast of the amount of water inflow into the dam bringing the surrounding weather information, will increase the accuracy of forecasting. From the review of literature, it was found that the forecast of the amount of water inflow into the dam will increase the accuracy performance when considering factors that are surrounding weather information. From the context of Bang Lang Dam, Yala province, which is a multipurpose dam located in the monsoon area therefore the relevant factors include rainfall, atmospheric pressure, humidity and temperature of the various rainfall and weather measurement stations to create a model using the ANN technique and measure the performance with the most efficient of accuracy values.

III. Data Preparation

This study collected water inflow volumes at Bang Lang dam, Yala province, include rainfall, atmospheric pressure, humidity and temperature data factors from 4 weather measurement stations in the Pattani River Basin above the Bang Lang dam from the Royal Irrigation Department, Ministry of Agriculture and Cooperatives and the Department of Water Resources, The Ministry of Water Resources and Environment used daily data from January 1, 2012 to December 31, 2020. The feature data are as follow:

- Daily Water inflow to Bang Lang Dam
- Daily rainfall
- Daily average temperature
- Daily average air pressure
- Daily average relative humidity

TABLE I. Summary of Features Considered in Analysis

Feature Name	Description
Inflow	Inflow water to Bang Lang Dam [10^6 m^3]
Inflow_1d	Inflow water to Bang Lang Dam in one day before [10^6 m^3]
Inflow_2d	Inflow water to Bang Lang Dam in two day before [10^6 m^3]
Rain	Rainfall in a day [mm]
Rain_1d	Rainfall in one day before [mm]
Rain_2d	Rainfall in two day before [mm]
Hum	Relative humidity [%RH]
Press	Barometric pressure [hPA]
Temp	Temperature [°C]

TABLE I presents the features of the dataset that is considered for analysis. All data was collected from selected stations by located in the upper part of Pattani River Basin which is the watershed of Bang Lang dam as shown in Fig. 1, which received data from 4 stations as shown in TABLE II.

TABLE II. Weather Stations In The Upper Pattani River Basin

Station list	Location	Latitude, longitude
BLD1	Banglang Dam, Banang-sata District, Yala Province	6.1608816, 101.2752384
BTGH	Tambon Yarom, Betong District, Yala Province	5.775768, 101.0924864
STH031	Ban Wang Sai, Than-To District, Yala Province	6.0092804, 101.2680512
VLGE35	Ban Muang, Betong District, Yala Province	5.9237996, 101.125984

All collected data total 5,112 records between January 1, 2012 to December 31, 2020 with 22 Attributes. Convert hourly data to find the average of the day, manage the missing value and fill by mean. Then analyzing with H2O deep learning and adjust epochs for stable values.

IV. Model Descriptions and Performance

The objectives of this study were to analyze the factors affecting the inflow of water in the dam in order to create and test a model for forecasting water inflow in Bang Lang Dam.

A. H2O Deep Learning (Feedforword Neural Network)

H2O Deep Learning developed base on multi-layer feed-forward artificial neural network for predictive modeling as Fig. 2. Model can contain many hidden layers. A feedforward neural network (ANN) is also known as a deep neural network (DNN) or multi-layer perceptron (MLP) Model [13].

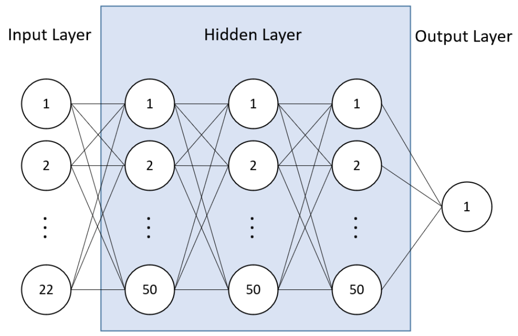

Fig. 2. Diagram of the H2O deep learning network.

B. Comparison of the Accuracy

After creating the model then run the forecast to find a suitable model and compare it with the actual value. Next, compare the accuracy of each individual model based on the value of Root Mean Square Error (RMSE), correlation coefficient (U) and The Mean Absolute Percentage Error (MAPE).

1) Root Mean Square Error : RMSE

RMSE is a value that shows the discrepancy between the estimated values from the model and the actual data values. If the RMSE is near zero, it means that the model has small discrepancies and can be used as a suitable representation of the real data. There is an equation as follows:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(\hat{Y}_t - Y_t)^2}$$

where $\hat{Y}_t$ is predicted data

Y_t is actual data

2) The correlation coefficient : CC : R

CC is the relationship of two variables, for example, what is the relationship between X and Y, with a range of -1 to 1, the meaning of the number is

-1.00 to -0.70 : Strong Negative (two variables are inverse)

-0.69 to -0.31 : Weak Negative

-0.30 to 0.30 : No correlation (two variables are not correlated)

0.31 to 0.69 : Weak Positive

0.70 to 1.00 : Strong Positive (two variables have values together).

C. Performance

Optimizing the model in the experiment is done by fetching the prepared data by splitting the training dataset and test dataset with a ratio of 70:30. Using a deep learning H2O set hidden layer model at L1 and L2 50 layers, we ran the Epochs setup test and measured the performance to determine the best Epochs with RMSE. The results obtained at Epochs 14 were valid. The best RMSE and correlation is shown in Fig. 3.

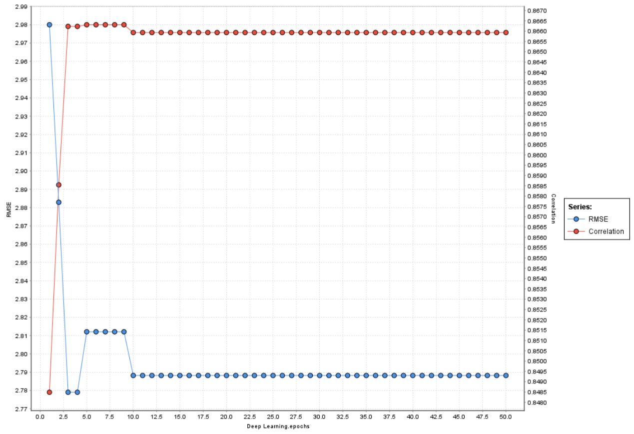

Fig. 3. RMSE and correlation of training model

V. Result and Discussion

The model was obtained from the previous steps. As shown in Fig. 3, RMSE and Correlation stabilized after epochs >10, epochs 20 was chosen for the H2O deep learning model. Forecasts of the test data were 5112 records, of which the experiments measured MAE: 1.300, RMSE: 3.111, R^2: 0.767, R: 0.876, and when plotting the result as a graph. will be as shown in Fig. 4-6.

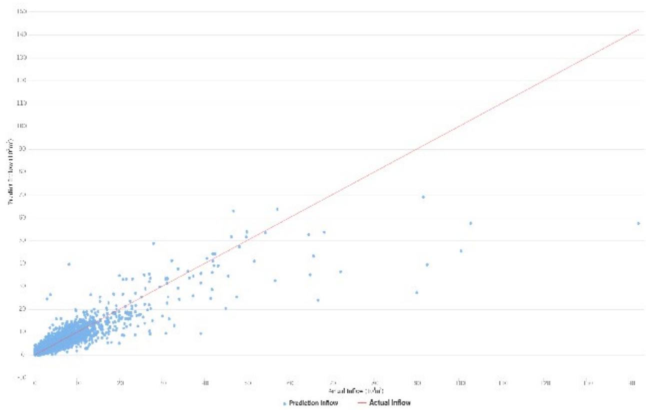

Fig. 4. Scatter plotting of predicted inflow vs actual inflow

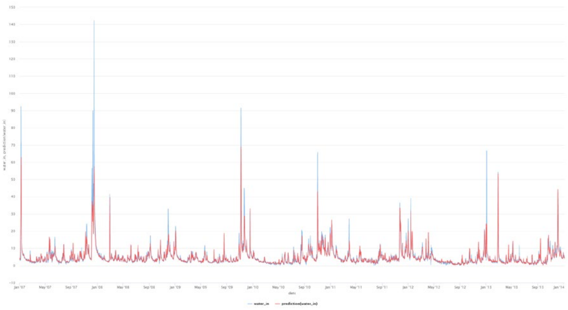

Fig. 5. Predicted inflow vs actual inflow (2007-2014)

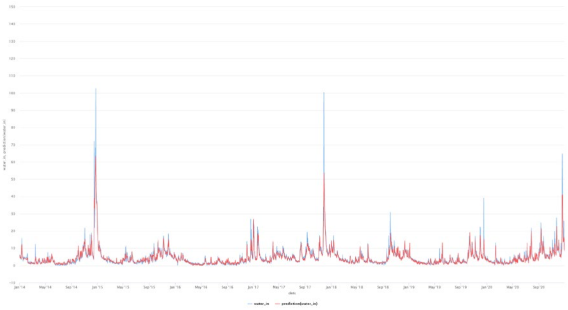

Fig. 6. Predicted inflow vs actual inflow (2014-2020)

VI. Conclusion

The forecasting of water inflow of Bang Lang Dam by applying the H2O Deep Learning model (ANN), which includes the dataset from many stations involved rainfall, atmospheric pressure, humidity, and temperature factors. The model can forecast the amount of water that will flow into the dam, leading to the planning of water efficiency management. The results of the experiment were satisfactory: RMSE was 3.111 and correlation (R) was 0.876, which was equal to square correlation (R^2) was 0.767. Lian et al. [14] indicated that if greater than 0.75 is a good prediction. Which after this the researcher will use the results of the experiment to create a dashboard system showing the forecast of water inflow in Bang Lang Dam.

References

[1] J. A. Awan and D. Bae, "Application of Adaptive Neuro-Fuzzy Inference System for dam inflow prediction using long-range weather forecast," Eighth International Conference on Digital Information Management (ICDIM 2013), 2013, pp. 247-251, DOI: 10.1109/ICDIM.2013.6693963.

[2] M. Pini, A. Scalvini, M. U. Liaqat, R. Ranzi, I. Serina, and T. Mehmood, "Evaluation of machine learning techniques for inflow prediction in Lake Como, Italy," in Procedia Computer Science, 2020, vol. 176, pp. 918–927. DOI: 10.1016/j.procs.2020.09.087.

[3] L. C. Chang et al., "Building an intelligent hydroinformatics integration platform for regional flood inundation warning systems," Water (Switzerland), vol. 11, no. 1. MDPI AG, Dec. 21, 2018. DOI: 10.3390/w11010009.

[4] F. Salazar, R. Morán, M. Toledo, and E. Oñate, "Data-Based Models for the Prediction of Dam Behaviour: A Review and Some Methodological Considerations," Archives of Computational Methods in Engineering, vol. 24, no. 1, Jan. 2017, DOI: 10.1007/s11831-015-9157-9.

[5] M. A. Khan, Z. Islam, and M. Hafeez, "Evaluating the Performance of Several Data Mining Methods for Predicting Irrigation Water Requirement," in Proceedings of the Tenth Australasian Data Mining Conference (AusDM 2012), Sydney, Australia. 2012. pp. 199-207

[6] T. Egawa, K. Suzuki, Y. Ichikawa, T. Iizaka, T. Matsui and Y. Shikagawa, "A water flow forecasting for dam using neural networks and regression models," 2011 IEEE Power and Energy Society General Meeting, 2011, pp. 1-6, DOI: 10.1109/PES.2011.6038925.

[7] D. H. Bae, D. M. Jeong, and G. Kim, "Monthly dam inflow forecasts using weather forecasting information and neuro-fuzzy technique," Hydrological Sciences Journal, vol. 52, no. 1, pp. 99–113, Feb. 2007, DOI: 10.1623/hysj.52.1.99.

[8] N. B. Elizaga, E. A. Maravillas and B. D. Gerardo, "Regression-based inflow forecasting model using exponential smoothing time series and backpropagation methods for Angat Dam," 2014 International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management (HNICEM), 2014, pp. 1-6, DOI: 10.1109/HNICEM.2014.7016185.

[9] A. Nayab and M. Faisal, "Water Management in Tarbela Dam By using Bayesian Stochastic Dynamic Programming in Extreme Inflow Season," Journal of Civil & Environmental Engineering, vol. 08, no. 02, 2018, DOI: 10.4172/2165-784x.1000299.

[10] A. Mahmoodzadeh et al., "Presenting the best prediction model of water inflow into drill and blast tunnels among several machine learning techniques," Automation in Construction, vol. 127, Jul. 2021, DOI: 10.1016/j.autcon.2021.103719.

[11] D. Zhang et al., "Modeling and simulating of reservoir operation using the artificial neural network, support vector regression, deep learning algorithm," Journal of Hydrology, vol. 565, pp. 720–736, Oct. 2018, DOI: 10.1016/j.jhydrol.2018.08.050.

[12] Z. C. Herbert, Z. Asghar, and C. Oroza, "Long-term Reservoir Inflow Forecasts: Enhanced Water Supply and Inflow Volume Accuracy Using Deep Learning," Journal of Hydrology, vol. 601, Oct. 2021, DOI: 10.1016/j.jhydrol.2021.126676.

[13] RapidMiner (2021) The core of RapidMiner is open source. https://docs.rapidminer.com/latest/studio/operators/modeling/predictive/neural_nets/deep_learning.html. Accessed 22 Oct 2021.

[14] Lian, Y., Chan, I-Chi., Singh, J., Demissie, M., Knapp, V. and H. Xie(2007), Coupling of hydrologic and hydraulic models for the Illinois River Basin, JournalofHydrology 344:210–222. referred to Motovilov, Y.G., Gottschalk, L., Engeland, K., and A.Rohde (1999), Validation of a distributed hydrological model against spatial observations. Agriculture and Forest Meteorology 98–99, 257–277