



River flow prediction based on improved machine learning method: Cuckoo Search-Artificial Neural Network

Wan Norsyuhada Che Wan Zania¹ · Marlinda Binti Abdul Malek² · Mohd Nadzri Md Reba³ · Nuratiah Zaini¹ · Ali Najah Ahmed¹ · Mohsen Sherif^{4,5} · Ahmed Elshafie⁶

Received: 17 October 2022 / Accepted: 16 November 2022 / Published online: 5 December 2022
© The Author(s) 2022

Abstract

One of the largest hydropower facilities currently in operation in Malaysia is the Terengganu hydroelectric facility. As a result, for hydropower generation to be sustainable, future water availability in hydropower plants must be known. Therefore, it is necessary to precisely estimate how the river flow will alter as a result of changing rainfall patterns. Finding the best value for the hyper-parameters is one of the problems with machine learning algorithms, which have lately been adopted by many academics. In this research, Artificial Neural Network (ANN) is integrated with a nature-inspired optimizer, namely Cuckoo search algorithm (CS-ANN). The performance of the proposed algorithm then will be examined based on statistical indices namely Root-Mean-Square Error (RSME) and Determination Coefficient (R^2). Then, the accuracy of the proposed model will be then examined with the stand-alone Artificial Neural Network (ANN). The statistical indices results indicate that the proposed Hybrid CS-ANN model showed an improvement based on R^2 value as compared to ANN model with R^2 of 0.900 at training stage and R^2 of 0.935 at testing stage. RMSE value, for ANN model, is 127.79 m³/s for training stage and 12.7 m³/s at testing stage. While for the proposed Hybrid CS-ANN model, RMSE value is equal to 121.7 m³/s for training stage and 10.95 m³/s for testing stage. The results revealed that the proposed model outperformed the stand-alone model in predicting the river flow with high level of accuracy. Although the proposed model could be applied in different case study, there is a need to tune the model internal parameters when applied in different case study.

Keywords River flow · Hydropower Plant · ANN · Hybrid CS-ANN

✉ Ali Najah Ahmed
Mahfoodh@uniten.edu.my

Wan Norsyuhada Che Wan Zania
syuhadazania@gmail.com

Marlinda Binti Abdul Malek
marlinda@iiu.edu.my

Mohd Nadzri Md Reba
nadzri@utm.my

Nuratiah Zaini
Nur_Atiah@uniten.edu.my

Mohsen Sherif
msherif@uaeu.ac.ae

Ahmed Elshafie
elshafie@um.edu.my

² Cataclysmic Management and Sustainable Development Research Group (CAMSDE), Department of Civil Engineering, Kulliyah of Engineering, International Islamic University Malaysia, 53100 Selangor, Malaysia

³ Geoscience and Digital Earth Centre (INSTeG), Research Institute for Sustainable Environment (RISE), Faculty of Built Environment & Surveying, Universiti Teknologi Malaysia, 81310 Johor Bahru, Johor, Malaysia

⁴ Civil and Environmental Engineering Department, College of Engineering, United Arab Emirates University, P.O. Box 15551 Al Ain, United Arab Emirates

⁵ National Water and Energy Center, United Arab Emirates University, P.O. Box 15551 Al Ain, United Arab Emirates

⁶ Department of Civil Engineering, Faculty of Engineering, University of Malaya (UM), 50603 Kuala Lumpur, Malaysia

¹ Department of Civil Engineering, College of Engineering, Universiti Tenaga Nasional (UNITEN), 4300 Kajang, Selangor, Malaysia

Introduction

Hydropower plant is one of the primary sustainable power sources accessible. Hydropower is considered environmental friendly due to its clean nature and renewable energy source (Owusu and Asumadu-sarkodie 2016). The term renewable refers to the hydrologic cycle that recycles water back into rivers, streams, and lakes each year.

River is a nature flowing watercourse that flowing towards an ocean, lake or another river. River flow is continually changing from day to day, the main fundamental impact on river flow is rainfall runoff in the watershed. Predicting and managing river flows is a need for flood control, water supply, agriculture and energy creation (Lallement 2019). Nevertheless, in hydrology, prediction of river flow is the method of evaluation the data has not yet been recorded. Reliable and precise river flow prediction is crucial for water resources supervision, optimization and planning (Kassem et al. 2019).

Recently, one of the techniques in Artificial Intelligence known as Artificial Neural Network has been widely used to solve engineering problems which including hydrology and water resources management (Azamathulla et al. 2010, 2016; Najah Ahmed et al. 2019). This technique assisted with reducing computational time required to limit the ideal solution. The utilization of ANNs in financial matters and account is a promising field of exploration particularly given the prepared accessibility of huge mass of data collections and the capacity of ANNs to recognize connections between an enormous number of variables (Safi 2016). Furthermore, ANN models have the ability to cater for both linear and nonlinear data without the need to make any presumptions as assumed in most traditional statistics and stochastic computable practices (Essam et al. 2022).

The artificial neural network (ANN) model, which served as the study's basic model, was initially introduced as a method of modelling with numerous inputs and a single output to forecast river flow. However, a number of issues with the corresponding mathematical steps in the ANN model may limit the model's capacity to identify and simulate the highly stochastic patterns in the input-target dataset, particularly for river flow. It is imperative to address the issue of establishing consistent river flow prediction accuracy because the ideal architecture of the ANN model, which includes the number of internal variables, is essential for achieving a high level of accuracy. In brief, despite ANN models' advantages, effective training methods are required to speed up the models' convergence during training. Furthermore, ANN models with traditional training algorithms may get stuck in local optimums, making the findings obtained less dependable and effective.

Therefore, ANN models that integrate optimization algorithms are required. The accuracy of the ANN model might be increased owing to the most recent optimization approach, which was introduced in the current study and is based on specialised operators. Even the avoidance of trapped local optimums will be aided by these operators.

Nevertheless, the main problem of Artificial Neural Networks (ANN) is the problems of local minima and slow convergence speeds (Ibrahim et al. 2021). Many solutions have been proposed by researcher to overcome this issue, where they are trying to hybrid Artificial Neural Network with innovative advance algorithms and have an excellent searching abilities to attain the global optima and avoid the local optima such as Genetic Algorithm (GA), Particle Swarm Optimization (PSO), Harmony Search (HS) and Ante Colony Optimization (ACO) (Tikhmarine et al. 2020). The performance of advance algorithm and artificial neural network are found not comparable with ANN. Therefore, another advancement calculation called as Cuckoo Search Optimization calculation was proposed to hybrid with Artificial Neural Networks to form Hybrid Cuckoo Search–Artificial Neural Network (CS-ANN).

For this proposed, this study used Artificial Neural Network and Hybrid Cuckoo Search–Artificial Neural Network (CS-ANN) technique to predict future river waters flowing into hydropower plant. Prediction of river flow into a hydro-power plant by using Artificial Neural Network and Hybrid Cuckoo Search–Artificial Neural Network (CS-ANN) technique has yet to be established in Malaysia. This study therefore seeks to fill this gap. The objective of this paper is to build up a reliable prediction model to predict future river water flowing into hydropower plant by using Artificial Neural Network and Hybrid Cuckoo Search–Artificial Neural Network (CS-ANN) technique.

Methodology

Study area

Hydropower plant in Terengganu is one of the major hydro-power plants in Malaysia. It is also serving as a multipurpose hydropower plant. This hydropower station is located in the district of Hulu Terengganu in Terengganu, East of Malaysia. The total area of Hulu Terengganu District is 3,874.63 km². Its mean temperature ranges from a minimum of 21 °C to a maximum of 31 °C, and the average annual rainfall is estimated at 2300 mm/year. In this study, rainfall and river flow are the parameter used to develop the proposed model at the chosen study area. Based on its strategic location and proximity to hydropower plant, Sg. Gawi (5,128,001) rainfall station monitored by Department of Drainage and Irrigation (DID) Malaysia, situated at Latitude of 5.143 °N

and Longitude of 102.84 °E was chosen as the main rainfall station used in this study as shown in Fig. 1. This study proposed the application of Artificial Intelligence technique to predict future river flow into hydropower plant namely, Artificial Neural Network and Hybrid Cuckoo Search–ANN technique.

Data exploration

Rainfall and river flow data used in this study are from 1971 until 2017 obtained from Department of Irrigation and Drainage Malaysia (DID) were used in this study. In this study, prediction model utilized recorded data to predict future events. Typically, recorded data are used to build a numerical model that catches significant patterns and to predict what will occur straightaway, or to propose moves to make for ideal results. The recorded data used would enhance the results of prediction model develop in this study.

Data Pre-processing

In order to predict future river waters flowing into hydro-power plant, rainfall data are one of the utmost pertinent meteorological parameters to be used in this study. Rainfall

is one of the most difficult elements of hydrological cycle to forecast (Bates et al. 2008). In most locations, ground rain gauge stations are accessible as precipitation estimation tools. Although the results are generally acceptable, constraints such as no coverage at certain remote regions or with one or two ground rain gauges at a single enormous catchment area are common in developing countries such as Malaysia. Therefore, rainfall and river flow data in this study undergo for pre-processing stage prior to be applied in prediction model. The purpose of data pre-processing is to identify and resolve issues on inadequate or missing data (Kuhn and Johnson 2013). There are three types of data pre-processing can be used to produce high quality of data, namely, data transformation, information gathering and generation of new information (Famili et al. 1997).

There are numerous reasons why a user might apply the pre-processing on the collected data. The goal of data pre-processing is to fill in missing values, aggregate data, classify data (data binning), and smooth the obtained data. In fact, the majority of the rainfall dataset experienced these characteristics, hence it is necessary to use a pre-processing strategy to provide the model with accurate data.

In this study, data transformation was applied using data normalization method to transform all variables in

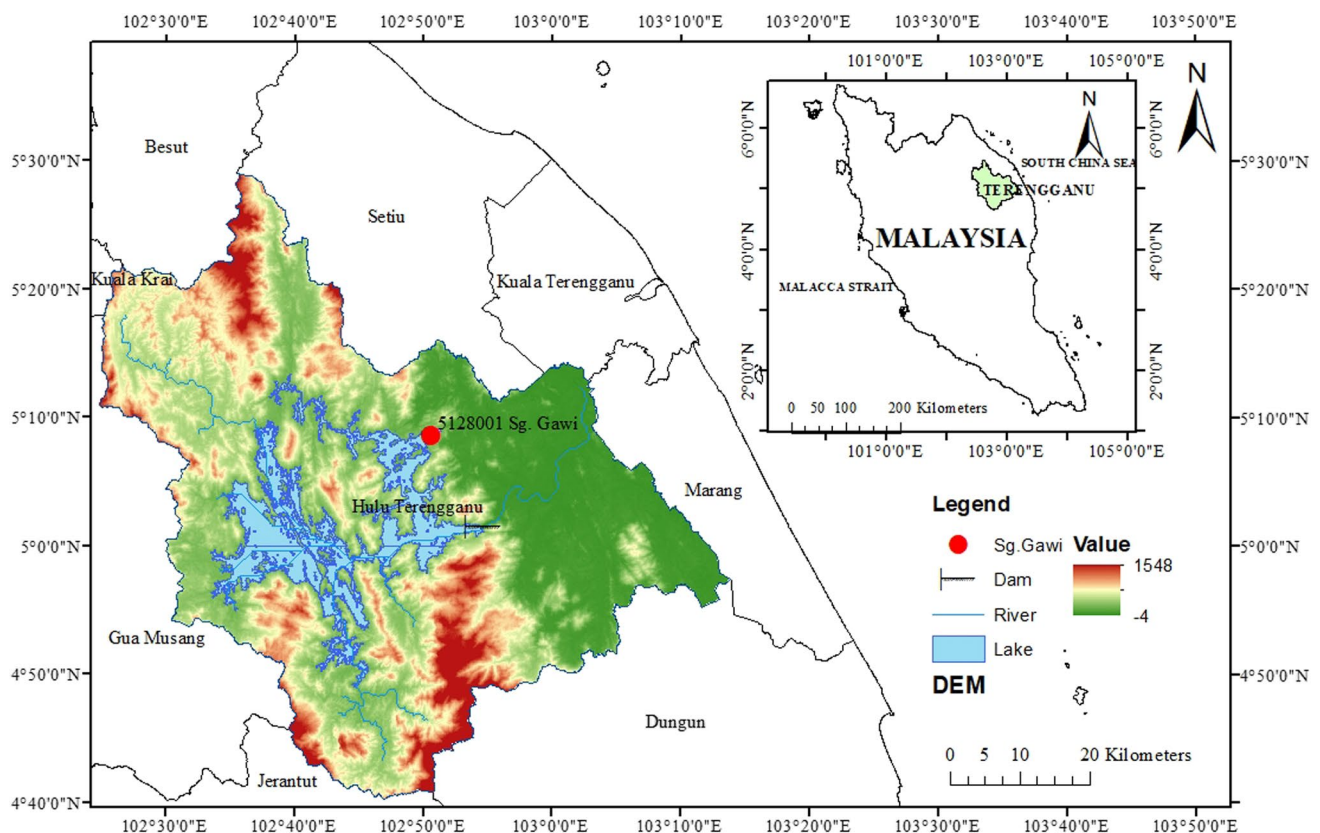


Fig. 1 Location of Study Area

the data to a specific range from 0 to 1. Data normalization can avoid overflows due to very large or very small weights in ANN during training, especially because ANN activation function is restricted and sensitive to predefined ranges (Shanmuganathan 2016). Therefore, ANN only can process the input data that range between 0 and 1 and the output of the data will produce the value range between 0 and 1. Equation 1 shows the equation of normalization used in this study.

$$\hat{Z} = \frac{Z_a - Z_{\min}}{Z_{\max} - Z_{\min}} \tag{1}$$

where \hat{Z} is the normalized value; and z_{\min} and z_{\max} are the minimum and maximum values of Z_a , respectively. The values of Z_a in this equation is the input of the prediction model, namely, rainfall (mm) and river flow (m³/s). The results obtained are then retransformed to identify the actual values using Eq. 2.

$$Z_b = Z_{\min} + \hat{Z} (Z_{\max} - Z_{\min}) \tag{2}$$

where \hat{Z} is the normalized value or the input value of river flow since the output of the prediction model is river flow; and z_{\min} and z_{\max} are the minimum and maximum values of Z_a , respectively. The values of Z_b is the output of the prediction model, namely river flow (m³/s).

Validation of the proposed model

In this study, statistical analyses were conducted to evaluate the performance of predicted River Flow compared to actual River Flow. Quantitative statistical tests using metrics, for example Root-Mean-Square Error (RMSE) and coefficient of determination (R^2), are widely accepted in many numerical metrics (Bennett et al. 2013; Amirabadizadeh et al. 2016).

The RMSE is used for determination of confidence intervals with range from 0 to ∞ with 0 corresponding to the ideal condition. The equation is given by:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (RF_{iactual} - RF_{ipredicted})^2} \tag{3}$$

where and $RF_{ipredicted}$ and $RF_{iactual}$ are the predicted and actual River Flow parameter measured for i^{th} test instance and n the number of the instances.

R^2 is coefficient determination that used statistical measure to assess the performance of proposed model that fitted regression line; the range of R^2 varies between 0 and 1, which near to 1 indicates that the proposed model describes all the variability of the response around its mean. It is given by following equation:

$$R^2 = 1 - \frac{\sum (RF_{ipredicted} - RF_{iactual})^2}{\sum (RF_{iactual} - \text{mean}RF_{ipredicted})^2} \tag{4}$$

Development of artificial neural network (ANN) algorithm

Artificial Neural Network (ANN) has been widely utilized to demonstrate nonlinear and non-stationary time-series data in hydrology. It was found to produce good outcomes as compared to statistical models. ANN models were found to be influential prediction model for the relation amongst rainfall and river flow parameters (Aichouri et al. 2015).

In this study, a practice based on ANN is anticipated for the predicting river flow into hydropower plant. The purposed of predicting river flow into hydropower plant is to predict future water availability to be used in generating future hydropower generation and helps in the development and management of water resources at the hydropower. Input data used in the proposed ANN model include annual rainfall and annual river flow, in order to predict future river flow into hydropower plant. In this study, Backpropagation Algorithm is used to develop River Flow prediction model. Backpropagation Neural Network (BPNN) is the most popular technique in Multilayer Perceptron (MLP) learning algorithms (Seiffert and Wunsch 2010; Al-Allaf 2011). The development of ANN model involved three (3) processes namely, flow of data, steps in development process of ANN model, and the future River Flow as recommended by (Maier et al. 2010).

Firstly, the input values that bring forward to the network towards nodes in hidden layer and increased with weights of connecting nodes (initial weights and threshold levels are set up randomly) to compute values of hidden nodes by activation function. Usually, linear sigmoid function is used to calculate the hidden values, however in this study the hyperbolic tangent function was used to perform and arrangement with a tough learning task that requires the use of a large network and accelerate the learning process. Karlik and Olgac (Karlik and Olgac 2011) point out that hyperbolic tangent sigmoid shows better recognition accuracy and results when used in MLP to evaluate both nodes of hidden and output layers than Bipolar sigmoid (Log sigmoid), unipolar sigmoid, conic section and radial basis function (RBF). Activation function of hyperbolic tangent function is given by:

$$\tanh x = \frac{\sinh x}{\cosh x} = \frac{e^x - e^{-x}}{e^x + e^{-x}} = \frac{1 - e^{-2x}}{1 + e^{-2x}} \tag{5}$$

This function has range outputs between -1 and 1. Using hyperbolic tangent activation function the actual outputs of the neuron in the hidden layer and output layer were calculated. The values in hidden layer were calculated and

propagate to output layer node. After that the error in output node is calculated using computation of nonlinear error signal as followings:

$$e_k(p) = y_{d,k}(p) - y_k(p) \tag{6}$$

where p refers to the p^{th} training example presented to the perceptron, $y_{d,k}(p)$ is the desired output of the neuron k at iteration p and $y_k(p)$ is the real output produced by the network in output to the input, x_i . When error is identified, it will be used in the backward phase from output layer to the hidden layer to adjust the weights. Here the learning rate, β , of 0.3 and momentum, α , of 0.3 (where it can be a positive number range $0 \leq \alpha < 1$) are brought to equation and rate of change need to be identify to updated the new weight value. The weights, w_{ij} , and biases, b_i , are then adjusted using the following perceptron learning rule:

$$w_{ij}(p + 1) = w_{ij}(p) - \beta \delta_j y_i \tag{7}$$

$$w_{ij}(p + 1) = w_{ij}(p) - \beta \delta_j x_i \tag{8}$$

$$b_i(p + 1) = b_i(p) - \beta \delta_j \tag{9}$$

where is the δ_j error gradient in the hidden layer, p is the number of epoch. Then, error from hidden layer need to be propagated to the input layer, where weight between input and hidden layer can be updated and rate of progress should

be determined for each weights. It is important to calculate all errors before update any weight to prevent from invalid results. A details process of ANN model development has been discussed in previous studies by (Maier et al. 2010), where it is involved three process from flow of data, steps in ANN model development process, and the outcomes/ decisions production. A schematic view of the development of ANN Model is shown in Fig. 2.

A detail ANN parameter chosen in this study is elaborate as follows:

i. Input selection

Input selection is relevant being developed of ANN. Two (2) approaches has been defined by (Maier et al. 2010) to be used in input selection namely model-free and model-based approaches. This parameter is significant in light of the fact that it can survey the consequence correlation between potential model inputs and outputs. In this study, two (2) input parameters used are historical rainfall and river flow data used to predict future River Flow into hydropower plant. This method is important to avoid from redundant input that may lead to over fitting and training.

ii. Hidden layers and nodes

The three (3) or more layers in ANN design incorporate the input, hidden and output layers. In this study, a trial-and-error technique for hidden layer was

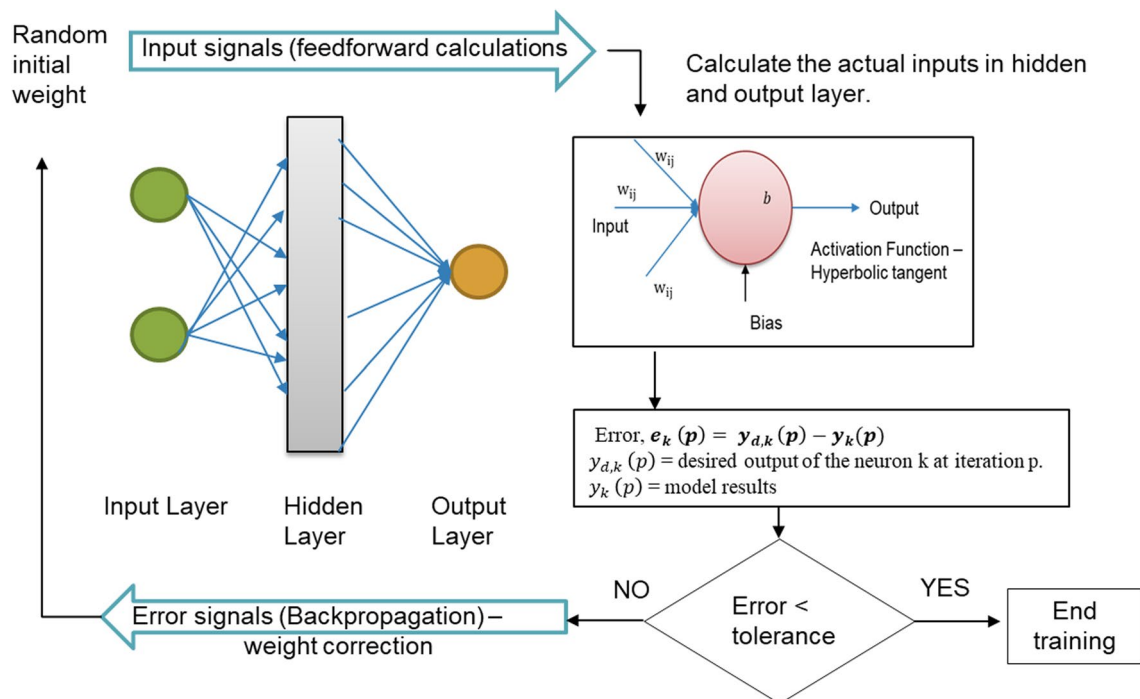


Fig. 2 Development of ANN Model

directed. The hidden layer recreated was at 1 and 2 layers with changing number of nodes at each layer.

iii. Learning Rate (β) and Momentum (α)

The purpose of learning rate (β) and momentum (α) are to expediate the training process while sustaining error reduction. In this study, learning rate, β , of 0.3 and momentum, α , of 0.3 were used as input. Instability of the system can be affected if the learning rate, β , are more than 1 (Osmi 2018).

iv. Initial weight

A nonzero random values and small values are used as initial weight to train ANN (Osmi 2018). Synaptic weights are utilized to store information. By using synaptic weights of normally distributed random numbers in the range of -1 to 1 , the number of initial weights can be adjusted.

v. Stopping criteria

Total sum of squared error and epoch number of simulations using training data are two (2) common method used as stopping criteria. In this study, 10 to 10,000 epochs were chosen to be utilized in the proposed ANN model. The ideal number of epochs is chosen dependent on the most reduced RMSE qualities and R^2 , which is around drawing nearer to 1.0.

vi. Activation function

Activation function was used to present nonlinearity to hidden layer where it can be linear, threshold or sigmoid functions. In this study, the sigmoid activation function was applied to the hidden layer since it combines nearly linear, curved, and constant behaviours subjected to the input values used (Cilimkovic 2015). (Karlik and Olgac 2011) stated that hyperbolic tangent sigmoid produced higher recognition accuracy and results when used in MLP to evaluate both nodes of hidden and output layers as compared to Bipolar sigmoid (Log sigmoid), unipolar sigmoid, conic section and radial basis function (RBF).

vii. Data partition

In ANN model development, three (3) data set are conducted namely as training, testing and validation. To estimate the unknown connection weight and to determine the adjusted weight and biases of a network, training set was used. Testing sets are used for calibration. This will prevent overtraining of system. In the meantime, to evaluate the generalization capacity of the trained model the validation set was used. In this study, 70% of all data are used for training and the remaining 30% of all data are used for validation and testing. In this study, the ANN architecture is design at three layers of network which are one input layer, one hidden layer, and one output layer.

Cuckoo Search (CS) algorithm

Cuckoo Search (CS) calculation is a meta-heuristic practice anticipated by (Yang et al. 2009). This calculation was empowered by the submit brood parasitism of cuckoo species by laying their eggs in the homes of other host winged animals. Some host home can keep direct differences. If an egg is found by the host winged creature as not of its own, it will by then either throw the dark egg or give up its home and manufacture another home elsewhere. The CS algorithm adheres to three idealize standards;

- (a) Each cuckoo lay each egg thusly and put its egg in subjectively picked home
- (b) The best homes with high quality of eggs will continue to the accompanying ages
- (c) The amount of open host homes is fixed, and the egg laid by a cuckoo is found by the host feathered animal with a likelihood $pa \in [0, 1]$.

For this situation, the host fledgling can either dispose of the egg or give up the home, and gather an absolutely new home. The standard c described above can be approximated by the part $pa \in [0, 1]$ of the n settles that are supplanted by new homes (with new irregular arrangements). While producing new arrangements x^{t+1} for cuckoo i , a Levy flight is performed using Eq. 10.

$$x_i^{t+1} = x_i^t + \alpha \oplus \text{levy}(\lambda) \quad (10)$$

where $\alpha > 0$ is the movement size, which should be related to the extents of the issue of interest. The item \oplus implies section savvy increases. The arbitrary stroll by Levy flight is progressively effective in investigating the pursuit space since its progression length is any longer over the long haul. Duty flight basically gives an arbitrary walk, while the sporadic development length is drawn from Levy dispersion appeared in Eq. 11.

$$\text{Levy} \approx u = t^{-\lambda}, 1 < \lambda \leq 3 \quad (11)$$

This has a limitless fluctuation with a boundless mean. The means were basically build in an irregular walk measure with a force law step-length circulation and a weighty tail. Levy flights have been used in many Search calculations. In Cuckoo Search calculation request flight is a critical part for nearby and worldwide search.

In this study, the procedure on training and testing, just as correspondence between cuckoo search calculation and ANN, required a specialized procedure to perform excellent preparing and get the best ANN model. In the initial step, the model will stack all the preparation information before CS begins to prepare, recognize the best weight and instate inclinations during the principal age. At that point these

weights are pass to ANN where examination is directed to locate the best arrangement utilizing in reverse way. CS will then update the weights with the best probable solution until the last cycle or epoch is achieved.

In this study, coding on hybrid CS-ANN model was written in Java NetBeans IDE 8.2. The proper number of nodes in a hidden layer ranges from $(2n^{\frac{1}{2}} + m)$ to (2^{n+1}) , where n is the number of input nodes and m is the number of output nodes (Fletcher and Goss 1993), (Olyaie et al. 2017). Figure 3 shows the anticipated Hybrid Cuckoo Search (CS)-Artificial Neural Network (ANN) process.

As it could be depicted in Fig. 3, the model first start by feeding the ANN model with the model input data and

processed by initializing the values of the weights and bias within the ANN structure, as shown in the first two blocks. Consequently, the model will generate the associated output accordingly, and hence, checked with the actual values with respect to the actual data. As presented in the third block, the evaluation of the model output and the actual data will be carried out, overall mean square error will be calculated and examined against the overall performance goal required, if higher, the model will be re-calculate again the values of the weights and biases that have been initialized at the beginning, hence, the Cuckoo searching algorithm will work in order to optimize the values of the estimated weights and biases to achieve the optimal ones.

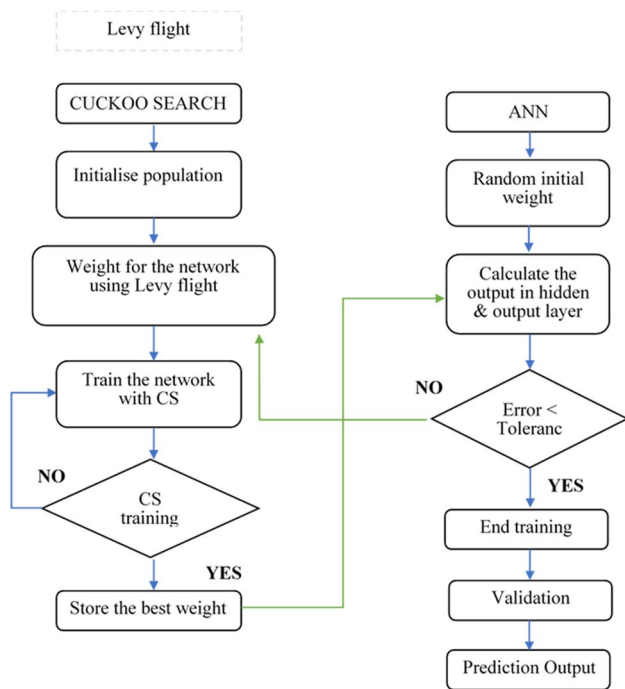


Fig. 3 Proposed hybrid cuckoo search (CS)-artificial neural network (ANN) process

Results and discussion

A. Validation of the proposed model

The goal of this paper is to establish a reliable prediction model to predict future river water flowing into hydropower plant. To utilize proposed technique to predict future river flow into hydropower plant, validation on the proposed model was led by performing correlations on the outcomes produced from the proposed ANN model, Hybrid CS-ANN and actual recorded river flow at year 1971 to 2017 during model training and testing as shown in Fig. 4. These values are also obtainable in Table 1, where two statistical tests namely Root-Mean-Square Error (RMSE) and Coefficient of Determination

Table 1 RMSE and R^2 values for ANN and Hybrid CS-ANN at 10 years ahead of prediction

Artificial Intelligence	RMSE		R^2	
	Training	Testing	Training	Testing
ANN	127.789	12.747	0.895	0.923
Hybrid CS-ANN	121.708	10.958	0.900	0.935

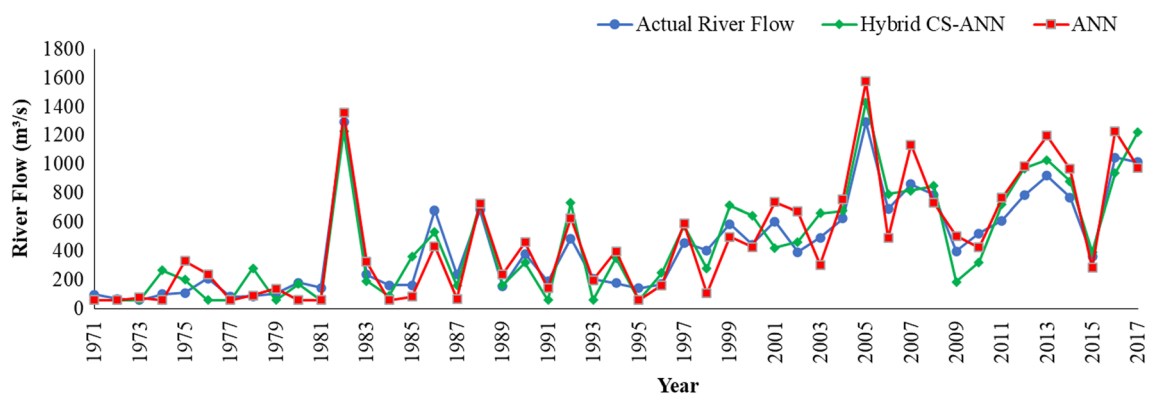


Fig. 4 Comparison between Stand-alone ANN and Hybrid CS-ANN

(R^2) were used to evaluate the strength of synthetic data generated.

In view of the results attained, it shows that despite the fact that both proposed models created comparative pattern when contrasted with genuine data, Hybrid CS-ANN model has figured out how to produce river flow values that are nearer to the actual recorded river flow values. This finding is likewise contrasted with the qualities produced by the independent ANN model. In Fig. 4 it shows the higher river flow obtained by the actual river flow, stand-alone ANN and hybrid CS-ANN is 1291.6 m³/s, 1575.6 m³/s and 1432.5 m³/s, respectively, in year 2005. The stand-alone ANN model value is more than actual river flow value and Hybrid CS-ANN in year 1975, 1976, 1982, 1990, 1994, 1997 in training stage and 2001, 2002, 2007, 2009, 2012–2014, 2016 during testing stage. In this way, it is demonstrated that river flow values produced by Hybrid CS-ANN model are superior to independent ANN model.

The R^2 stand-alone ANN during training and testing is 0.895 and 0.923, respectively. It was also found that R^2 of Hybrid CS-ANN model is better as compared to ANN model for both training and testing with 0.900 and 0.935, respectively, as shown in Table 1. Hybrid CS-ANN has improved by 1.3% as compared to performance of ANN prediction model. The hybrid ANN model generally perform better than stand-alone ANN (Wang et al. 2006). The superiority of Hybrid CS-ANN in fast convergence and optimum achievement and it could find a very good and acceptable estimation. It should be noted the higher performance of Hybrid CS-ANN in reaching better results than ANN (Khademikia et al. 2016). The computational time of Hybrid CS-ANN required a short period of time compared to stand-alone ANN based on Table 1. The performance advances made by the proposed Hybrid CS-ANN over the independent ANN might actually be accomplished

as a result of the hybridisation of the independent ANN, which improves the correspondence ability of the cuckoos to look for a superior area where the ideal home can impart data to the cuckoo (Chiroma et al. 2015). Hence, through this validation process, we can achieve that the proposed Hybrid CS-ANN is appropriate to be utilized for predicting future river flow into hydropower plant.

B. Prediction of river flow into hydropower plant

The developed Hybrid CS-ANN is proved outperform ANN. Therefore, this investigation further continues to utilize Hybrid CS-ANN in anticipating of future manufactured prediction of river flow into hydropower plant. Figure 5 shows graphical presentation of predicted river flow into hydropower plant for year 1971 to 2027.

Figure 5 shows the predicted river flow into hydropower plant from year 1971 to 2027, it shows that the value is fluctuate from year 2018 till 2027 due to the trend of Hybrid CS-ANN at the testing stage with most of the peak hydrographs met. Therefore, it can be concluded that the values of predicted river flow data from year 2018 to 2027 produced from the proposed Hybrid CS-ANN model develop in this study are satisfactory and acceptable. The analysis conducted are prediction of river flow using Hybrid Cuckoo Search–Artificial Neural Network. Usage of rainfall data and river flow data as well as incorporation of Artificial Intelligence technique in prediction of future river flow in this study has introduced an advanced technique in prediction of River Flow into hydropower plant as compared to traditional and commonly used statistical and stochastics approaches. The outcome of this study is the development of prediction model using stand-alone Artificial Neural Network and Hybrid Cuckoo Search-ANN algorithm for future river flowing into hydropower plant. Analyses have been conducted to investigate the insight of the performance, accuracy and robustness of this prediction model. In this study, the Hybrid Cuckoo Search-ANN model was found

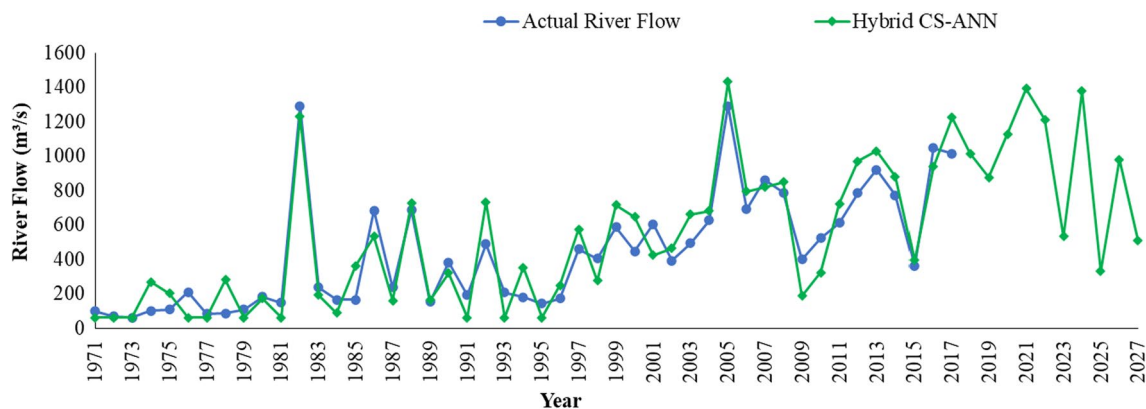


Fig. 5 Predicted River Flow into hydropower plant from year 1971 to 2027

to be more efficient than stand-alone ANN model in predicting the river flow into hydropower plant.

Conclusion and recommendation

This study is being commenced at a time when disagreement is being expressed within the modern computational tools. Utilisation of prediction using Hybrid Cuckoo Search–ANN technique introduced an advancement in river flow forecasting technique that extends beyond the traditional statistical and stochastics approaches. The significance of this study is to make contributions towards the prediction of future river flowing into hydropower plant to generate hydropower operation. Knowledge of future water availability in hydropower plant is pertinent for the sustainability of hydropower generation at Sultan Mahmud Shah Power Plant. In development of prediction model for river flow, Root-Mean-Square Error (RMSE) and coefficient of determination (R^2) were used as model performance indicator to evaluate competency of the proposed ANN and Hybrid CS-ANN models. In this study, it was found that at epoch 1000, the proposed ANN model attained R^2 of 0.895 at training stage and R^2 of 0.923 at testing stage. In addition, the proposed Hybrid CS-ANN model showed an improvement in R^2 as compared to ANN model with R^2 of 0.900 at training stage and R^2 of 0.935 at testing stage. RMSE for proposed ANN model is 127.79 m^3/s at training stage and 12.7 m^3/s at testing stage. Proposed Hybrid CS-ANN model showed a good RMSE value as compared to stand-alone ANN model with 121.7 m^3/s at training stage and 10.95 m^3/s at testing stage. Different input combinations of rainfall and river flow have been introduced, nevertheless, no noticeable improvement has attained. Several recommendations that can be proposed in this study for future work. Include installation of more Ground Rainfall stations at the study area for validation purposes. This is to avoid biasness incurred due to temporal resolution of scarce numbers of Ground Rain gauges at the study area. Prediction of future river flowing into hydropower plant can include additional meteorological and physical ground parameters such as cloud cover, pressure, temperature, humidity, ground level, soil type and others.

Acknowledgements The authors would also like to acknowledge technical support from Geoscience and Digital Earth Centre (INSTeG), Research Institute for Sustainable Environment (RISE), Universiti Teknologi Malaysia, and access to data from Department of Irrigation and Drainage Malaysia (DID) and Department of Survey and Mapping Malaysia (JUPEM).

Author Contributions Conceptualization was contributed by MAM and MNMR; Data curation was contributed by MAM; Formal analysis was contributed by WNCWZ and NZ; Funding acquisition was contributed

by MAM; Investigation was contributed by WNCWZ; Methodology was contributed by WNCWZ and MAM; Project administration was contributed by MAM; Resources were contributed by WNCWZ; Software was contributed by MNMR; Supervision was contributed by MAM and MNMR; Validation was contributed by WNCWZ, AE and NZ; Writing—original draft was contributed by WNCWZ, ANA and MS; Writing—review and editing was contributed by MAM, MNM, NZ, ANA, MS and AE. All authors have read and agreed to the published version of the manuscript.

Funding This MSc study is financially supported by BOLD Scholarship, College of Graduate School. Universiti Tenaga Nasional, Malaysia.

Availability of data and materials Not applicable.

Declarations

Conflicts of interest The authors declare no conflict of interest.

Consent to Participate Not applicable.

Consent to Publish Not applicable.

Ethical Approval
Not applicable.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Aichouri I, Hani A, Bougherira N et al (2015) River flow model using artificial neural networks. *Energy Procedia* 74:1007–1014
- Al-Allaf ONA (2011) Fast BackPropagation Neural Network algorithm for reducing convergence time of BPNN image compression. In: 2011 International Conference on Information Technology and Multimedia: “Ubiquitous ICT for Sustainable and Green Living”, ICIM 2011. IEEE, pp 1–6
- Amirabadizadeh M, Ghazali AH, Huang YF, Wayayok A (2016) Downscaling daily precipitation and temperatures over the Langat River Basin in Malaysia : a comparison of two statistical downscaling approaches. *Inter J Water Resour Environ Eng* 8:120–136. <https://doi.org/10.5897/IJWREE2016.0585>
- Azamathulla HM, Ghani AA, Chang CK et al (2010) Machine Learning Approach to Predict Sediment Load – A Case Study. *Clean (weinh)* 38:969–976. <https://doi.org/10.1002/CLEN.201000068>
- Azamathulla HM, Haghbi AH, Parsaie A (2016) Prediction of side weir discharge coefficient by support vector machine technique. *Water Supply* 16:1002–1016. <https://doi.org/10.2166/WS.2016.014>
- Bennett ND, Croke BFW, Guariso G et al (2013) Characterising performance of environmental models. *Environ Model Softw* 40:1–20. <https://doi.org/10.1016/j.envsoft.2012.09.011>

- Chiroma H, Abdul-Kareem S, Khan A et al (2015) Global warming: Predicting OPEC carbon dioxide emissions from petroleum consumption using neural network and hybrid cuckoo search algorithm. *PLoS ONE* 10:1–21. <https://doi.org/10.1371/journal.pone.0136140>
- Cilimkovic M (2015) Neural Networks and Back Propagation Algorithm. Institute of Technology Blanchardstown Blanchardstown Road North Dublin 15
- Essam Y, Ahmed AN, Ramli R, Chau KW, Ibrahim MSI, Sherif M, Sefelnasr A, El-Shafie A (2022) Investigating photovoltaic solar power output forecasting using machine learning algorithms. *Eng Appl Comput Fluid Mech* 16(1):2002–2034. <https://doi.org/10.1080/19942060.2022.2126528>
- Famili A, Shen W-M, Weber R, Simoudis E (1997) Data preprocessing and intelligent data analysis. *Intell Data Anal* 1:3–23
- Fletcher D, Goss E (1993) Forecasting with neural networks: an application using bankruptcy data. *Inform Manage* 24:159–167. [https://doi.org/10.1016/0378-7206\(93\)90064-Z](https://doi.org/10.1016/0378-7206(93)90064-Z)
- Ibrahim KSMH, Huang YF, Ahmed AN et al (2021) A review of the hybrid artificial intelligence and optimization modelling of hydrological streamflow forecasting. *Alex Eng J*. <https://doi.org/10.1016/j.aej.2021.04.100>
- Intergovernmental Panel on Climate Change (2007) *Climate Change 2007: Mitigation of Climate Change*. Cambridge University Press, Cambridge. <https://doi.org/10.1017/CBO9780511546013>
- Karlik B, Olgac AV (2011) Performance analysis of various activation functions in generalized MLP architectures of neural networks. *Inter J Artificial Intell Expert Syst* 1:111–122
- Kassem AA, Raheem AM, Khidir KM, Alkattan M (2019) Predicting of daily Khazir basin flow using SWAT and hybrid SWAT-ANN models. *Ain Shams Eng J* 11:435–443
- Khademikia S, Haghizadeh A, Godini H, Khorramabadi GS (2016) Artificial neural network-cuckoo optimization algorithm (ANN-COA) for optimal control of khorramabad wastewater treatment plant, Iran. *Civil Engineering Journal* 2(11):555–567. <https://doi.org/10.28991/cej-2016-00000058>
- Kuhn M, Johnson K (2013) *Applied predictive modeling*. Springer
- Lallament C (2019) Hydrometry: measuring the flow rate of a river, why and how? *Encyclopedia of the Environment* 1–9
- Maier HR, Jain A, Dandy GC, Sudheer KP (2010) Methods used for the development of neural networks for the prediction of water resource variables in river systems: current status and future directions. *Environ Model Softw* 25:891–909
- Najah Ahmed A, Binti Othman F, AbdulmohsinAfan H et al (2019) Machine learning methods for better water quality prediction. *J Hydrol (amst)* 578:124084. <https://doi.org/10.1016/j.jhydrol.2019.124084>
- Olyaie E, Abyaneh HZ, Mehr AD (2017) A comparative analysis among computational intelligence techniques for dissolved oxygen prediction in Delaware River. *Geosci Front* 8:517–527
- Osmi SFC (2018) Development Of Hybrid Meta Heuristic And Neural Network Models For Water Quality Prediction At Langat River Basin
- Owusu PA, Asumadu-sarkodie S (2016) A review of renewable energy sources, sustainability issues and climate change mitigation. *Cogent Eng* 3:1167990. <https://doi.org/10.1080/23311916.2016.1167990>
- Safi SK (2016) A comparison of artificial neural network and time series models for forecasting GDP in Palestine. *Am J Theoret Appl Statis* 5(2):58. <https://doi.org/10.11648/j.ajtas.20160502.13>
- Seiffert J, Wunsch DC (2010) Backpropagation and ordered derivatives in the time scales calculus. *IEEE Trans Neural Netw* 21:1262–1269. <https://doi.org/10.1109/TNN.2010.2050332>
- Shanmuganathan S (2016) Artificial Neural Network Modelling: An Introduction. In: Shanmuganathan S, Samarasinghe S (eds) *Artificial Neural Network Modelling*. Springer International Publishing, Cham, pp 1–14
- Tikhmarine Y, Souag-Gamane D, Najah Ahmed A et al (2020) Improving artificial intelligence models accuracy for monthly streamflow forecasting using grey Wolf optimization (GWO) algorithm. *J Hydrol (amst)* 582:124435. <https://doi.org/10.1016/j.jhydrol.2019.124435>
- Wang W, Gelder PHAJMV, Vrijling JK, Ma J (2006) Forecasting daily streamflow using hybrid ANN models. *J Hydrol (amst)* 324:383–399. <https://doi.org/10.1016/j.jhydrol.2005.09.032>
- Yang X, Deb S, Behaviour ACB (2009) Cuckoo Search via Levy Flights. *World Congress Nat Biologic Inspired Comput (NaBIC)* 2009:210–214. <https://doi.org/10.1109/NABIC.2009.5393690>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.