Efficacy of ANFIS-GOA technique in flood prediction: a case study of Mahanadi river basin in India

Abinash Sahoo, Sandeep Samantaray and Siddhartha Paul
Department of Civil Engineering, NIT Silchar, Assam, India
*Corresponding author. E-mail: samantaraysandeep963@yahoo.com

ABSTRACT

Accurateness in flood prediction is of utmost significance for mitigating catastrophes caused by flood events. Flooding leads to severe civic and financial damage, particularly in large river basins, and mainly affects the downstream regions of a river bed. Artificial intelligence (AI) models have been effectively utilized as a tool for modelling numerous nonlinear relationships and is suitable to model complex hydrological systems. Therefore, the main purpose of this research is to propose an effective hybrid system by integrating an Adaptive Neuro-Fuzzy Inference System (ANFIS) model with meta-heuristic Grey Wolf Optimization (GWO) and Grasshopper Optimization Algorithm (GOA) for flood prediction in River Mahanadi, India. Robustness of proposed meta-heuristics are assessed by comparing with a conventional ANFIS model focusing on various input combinations considering 50 years of monthly historical flood discharge data. The potential of the AI models is evaluated and compared with observed data in both training and validation sets based on three statistical performance evaluation factors, namely root mean squared error (RMSE), mean squared error (MSE) and Wilmott Index (WI). Results reveal that robust ANFIS-GOA outperforms standalone AI techniques and can make superior flood forecasting for all input scenarios.

Key words: ANFIS-GOA, ANFIS-GWO, flood, hydrological system, modelling

INTRODUCTION

Forecasting various hydrological phenomena is of significant concern in the field of hydrology and is pivotal for appropriate water resources development and disaster management. Every year, substantial public and financial damages, as well as fatalities, are caused by dangerous storms worldwide, specifically in areas subjected to monsoon weather and regions with slow growth of water conservancy schemes (Jiang et al. 2013; Wang et al. 2015; Yu et al. 2015). Flood prediction and forecasting act as the essential practices to control flood events across the globe (Young 2002; Campolo et al. 2003; Moore et al. 2005; Jiang et al. 2016; Rath et al. 2017; Panigrahi et al. 2018; Samantaray & Sahoo 2020). Evidence of this information is a complex investigation that has concerned researchers over past decades. Upstream circumstances intensely influence the flood flows in downstream zones; therefore, a flood forecasting model needs to be developed which can detect accurately eloquent fundamental connection amid downstream and upstream situations. Artificial neural networks (ANNs) act as suitable models for the problem mentioned above. Over past decades, ANN and Adaptive Neuro-Fuzzy Inference System (ANFIS) models have been comprehensively utilized in a variety of engineering applications involving hydrology such as simulation of rainfall-runoff process (Wu & Chau 2011), model groundwater problems (Sahoo et al. 2005; Taormina et al. 2012), forecast streamflow (El-Shafie et al. 2007; Shu & Ouarda 2008) and modelling water quality (Singh et al. 2009; Yan et al. 2010). In the last decade, numerous developments have been made to improve both the enactment and consistency of ANN tools. In recent times, attention has been
shifted from focusing on the applicability of ANN tools to importance on refining estimation capability and clarifying the inner conduct of ANN tools (Maier & Dandy 2000; Sudheer & Jain 2004; Araghinejad 2013).

Chen et al. (2006) proposed the construction of a flood forecast model using ANFIS in the Choshui River, Taiwan, and compared its performance with a back-propagation neural network (BPNN). Obtained results demonstrated that ANFIS was effective and reliable to construct a flood forecasting model with better accuracy. Bisht & Jangid (2011) used ANFIS to develop river stage-discharge models at the Dhawalaishwaram barrage site in Andhra Pradesh, India. Based on the comparison of observed and estimated data, outcomes revealed that ANFIS performed better in predicting river flow discharge compared to customary models. Rezaeiazadeh et al. (2014) used ANFIS, ANN, multiple linear regression, and multiple nonlinear regression to forecast the peak flow of Khor Shrin catchment, positioned in the Fars region, Iran, on a daily basis. Predictive capabilities of the proposed model were evaluated and it was observed that ANFIS performed superiorly for predicting daily flow discharge at the proposed site with spatially distributed rainfall as input. Ghorbani et al. (2016) investigated the usability of two diverse ANNs, which include multilayer perceptron (MLP) and RBFN, and a comparison is made with support vector machine (SVM) for predicting streamflow in Zarrinehrud River, Iran on a monthly basis. Results indicated that the SVM model was more certain and consistent than MLP and RBFN in river flow prediction. Zhou et al. (2019) proposed a recurrent ANFIS entrenched with GA and least square estimator that helped in optimizing model constraints to make multi-step ahead flood forecasts of Three Gorges Reservoir, China. The results demonstrated that the proposed model significantly improved the accuracy of flood forecasts.

Selection of model input (for example, pre-processing of data utilizing data-mining methods), model parameter optimization and post-processing of model output (for example, real time correction, ensemble forecasts) are key motivations and significant constituents in multi-step ahead hydrological forecasts. Several climatological parameters have a significant impact on the performance of models while dealing with multi-step ahead forecasts in application to real world problems. While making multi-step ahead flood forecasts, models with different problems (e.g. model instability/overfitting) will fail in tracing flow traces carefully, particularly during peak flows, because of an increase in forecast horizon. Therefore, an effective algorithm is necessary to determine an optimum network parameter setting for improving the reliability and stability of forecasting models.

Optimization algorithms such as differential evolution (DE), particle swarm optimization (PSO), genetic algorithm (GA) and GWO have been developed and integrated with data-driven models for forecasting various water resources and environmental problems (Senapati et al. 2007; Guo et al. 2014; Zhang et al. 2014; Prasad et al. 2017; Yaseen et al. 2017; Ewees & Elaziz 2018; Dehghani et al. 2019a, 2019b; Tikhamarine et al. 2020). Mirjalili et al. (2014) introduced a GWO algorithm to optimize an MLP network showing superior performance compared to GA, PSO, Evolution Strategy (EA) and Ant Colony Optimization (ACO). Tikhamarine et al. (2020) proposed different efficient hybrid neural network models combining GWO with ANN, SVM, and MLR to improve precision and ability in forecasting streamflow on a monthly basis at Aswan High Dam in the Nile River. The results revealed that integrated techniques used in their study outperformed standard ANN, SVM, and MLR techniques and made improved forecasts throughout training and testing periods for monthly inflow. However, GWO has a few disadvantages such as low solving accuracy, unsatisfactory ability of local searching, and slow convergence rate. In recent times, a new meta-heuristic nature inspired optimization algorithm called GOA was introduced by Saremi et al. (2017). GOA is based on the swarming behaviour of grasshoppers. It is utilized to improve ANFIS model performance, which has previously not been explored for flood prediction. This algorithm is classified as a multi-solution algorithm in optimization problems having higher accurateness and avoids local optima. It has proved to be an influential algorithm in challenging problems which deal with unidentified search spaces (Saremi et al. 2017; Mirjalili et al. 2018). There are several applications of GOA integrated with AI models in researches involving various fields of science and engineering: selecting harmonic elimination in low-frequency voltage source inverter (Steczek et al. 2020); approximate flyrock distance in mine blasting (Fattahi & Hasanipanah 2021); estimating the parameter of photovoltaic modules on the basis of single diode models (Montano et al. 2020); neural assessment of heating load (HL) of residential buildings (Moayedi et al. 2019); optimal deployment of wireless sensor networks (Dehgbouch & Debbat 2021); prediction of pipe burst in urban water distribution systems (Alizadeh et al. 2019) and many more. GOA also has specific application in modelling various hydrological parameters such as evaluation of rainfall temporal variability (Farrokhi et al. 2020); monthly prediction of groundwater level (Seifi et al. 2020); forecasting short-term hydrological drought (Nabipour et al. 2020) and optimization of the non-linear muskingum flood routing model (Khalifeh et al. 2020).
The present research utilizes robust ANFIS-GWO and ANFIS-GOA models for flood prediction in the Mahanadi river basin, India, and outcomes achieved are assessed with conventional ANFIS and ANN models. Based on a literature survey, it is observed that no research has been carried out for predicting flood events using the robust ANFIS-GOA technique. The novelty of this research is the application of ANFIS-GOA in flood prediction, this has been carried out by the authors. Also, this study aims for the sensitivity analysis for three different artificial intelligence tools to forecast monthly flood water levels. In the present research, special attention is paid to modelling parameter optimization. This research also places emphasis on various input combinations for different scenarios which has a strong impact on the desired model output.

**STUDY AREA**

The Mahanadi River (Figure 1) flows in central India, rising in the hills of the southeastern state of Chhattisgarh, and mainly flowing through Odisha state. Mahanadi has a total course of 858 km (494 km in Odisha) and has an estimated drainage area of 141,600 km² (65,580 km² in Odisha), which is about 42% of the Odisha state. Mahanadi lies within 20.11°N 81.91°E coordinates. It is known for its devastating floods, causing much misery to life and property, as recorded in history. It is a significant river for the state of Odisha. It originates from the south of Sihawa town in Dhamtari district of Chhattisgarh and finally discharges to the Bay of Bengal at False Point of Jagatsinghpur, Odisha. In the present research, Jondhra and Kesinga gauge stations of Mahanadi river basin are selected for predicting flood events.

**METHODOLOGY**

**ANFIS**

On the basis of fuzzy sets theory, fuzzy logic (FL) works where there is no brusque or obvious margin. Unlike two-valued Boolean logic, FL is multi-valued and deals with degrees of membership and truth. It utilizes any rational value from a real number set between 0 (totally false) and 1 (totally true), identified as its membership value (MV) and the function which correspond to those specified values is called the membership function (MF) (Das et al. 2019). MF can take the shape of triangle, trapezoid, Gaussian or sigmoid and are chosen on the basis of usability. It possesses coherent functions such as AND, OR, NOT. Every function possesses its own description on the

![Figure 1 | Proposed gauge station.](http://iwaponline.com/h2open/article-pdf/4/1/137/917228/h2oj0040137.pdf)
basis of MV theory. Fuzzy set of laws are additional significant elements of FL that narrate fuzzy sets to one another. The IF-THEN rule suggests that if rule is correct as well as precursor, that resultant is also correct. To train ANFIS indicates finding out constraints in its architecture utilizing an optimization function. In the training period, principle and resultant constraints are used, as revealed in Figure 2. For obtaining efficient outcomes using ANFIS, it is requisite to train its network. Since growth in application of ANFIS, various training techniques have been recommended for achieving improved outcomes. Amid various Fuzzy Inference Systems, Takagi-Sugeno (Takagi & Sugeno 1985) is among the most regularly utilized. A distinctive rule set having dual fuzzy ‘IF-THEN’ rule is classified as:

\[
\text{IF} (x \text{ is } A_1) \text{ AND } (y \text{ is } B_1) \text{ THEN } (f_1 = p_1x + q_1y + r_1) \tag{1}
\]

\[
\text{IF} (x \text{ is } A_2) \text{ AND } (y \text{ is } B_2) \text{ THEN } (f_2 = p_2x + q_2y + r_2) \tag{2}
\]

where \(A_1, A_2, B_1, B_2\) are MVs of input variables \(x\) and \(y\), correspondingly; \(p_1, q_1, r_1\) and \(p_2, q_2, r_2\) are constraints of output functions \(f_1\) and \(f_2\). NF systems are an outcome of conjunction of ANN and FL. ANFIS

---

**Figure 2** | (a) Architecture, (b) flow chart of ANFIS.
comprises of five layer MLP wherein the BP algorithm is utilized for modifying primarily nominated MFs whereas the least mean square algorithm regulates unidentified factors of linear output functions.

GWO

Mirjalili et al. (2014) proposed GWO algorithm mimicking social behaviour and hierarchy of grey wolves. GWO is a novel meta-heuristic optimization algorithm. In general, the wolves pack is distributed into four categories: Alpha ($\alpha$), Beta ($\beta$), Delta ($\delta$) and Omega ($\omega$). Alpha ($\alpha$) wolf is the most dominant wolf and is the leader of the wolves pack. The level of domination goes on decreasing from $\alpha$ to $\omega$ as presented in Figure 3. The mechanism involved in GWO is carried out by splitting a solution set into four groups for a specified optimization problem. $\alpha$, $\beta$ and $\delta$ wolves are the first three solutions, whereas residual solutions fall in the group of $\omega$ wolves. For implementing this mechanism, the hierarchical step is updated in each iteration on the basis of three optimal solutions. A representation of the updated position is demonstrated in Figure 4. The significant approach involved in GWO is to search, encircle, hunt, and finally attack the prey.

Previous to hunting procedure, grey wolves encircle the victim. The encircling conduct of grey wolves is represented in the following equation:

$$X(t + 1) = X_P(t) - A \cdot D$$

(3)

Figure 3 | The hierarchical leadership structure of grey wolves.

Figure 4 | Flow chart of ANFIS-GWO.
where $\tilde{X}(t + 1)$ is the next position of any wolf, $\tilde{X}_p(t)$ is the grey wolf position vector, $t$ is the present iteration, $\tilde{A}$ is the coefficient of matrix and $\tilde{D}$ is the distance that separates wolf and victim. This is estimated using the following equation:

$$
\tilde{D} = |\tilde{C} \cdot \tilde{X}_p(t) - \tilde{X}(t)| \quad (4)
$$

$$
\tilde{A} = 2 \tilde{a} \tilde{r}_1 - \tilde{a} \quad (5)
$$

$$
\tilde{C} = 2 \tilde{r}_2 \quad (6)
$$

where $\tilde{r}_1$ and $\tilde{r}_2$ are arbitrarily produced from $\langle 0, 1 \rangle$.

The preceding equations allow a solution for relocating around prey in a two-dimensional search space. Nevertheless, this is insufficient for simulating the societal intellect of grey wolves. For simulating the prey, the solution is achieved by $\alpha$ wolf, which is nearer to the position of prey, however, the global optimum solution is not known. Hence it is anticipated that the topmost three results have a better awareness about their position, so the remaining wolves must be prepared for updating their positions utilizing the following equations:

$$
\tilde{X}(t + 1) = \frac{\tilde{X}_1 + \tilde{X}_2 + \tilde{X}_3}{3} \quad (7)
$$

where $\tilde{X}_1$, $\tilde{X}_2$ and $\tilde{X}_3$ are calculated using the following equations:

$$
\tilde{X}_1 = \tilde{X}_a(t) - A_1 \cdot \tilde{D}_a, 
\tilde{X}_2 = \tilde{X}_b(t) - A_2 \cdot \tilde{D}_b, 
\tilde{X}_3 = \tilde{X}_d(t) - A_3 \cdot \tilde{D}_d \quad (8)
$$

$$
\tilde{D}_a = |\tilde{C}_1 \cdot \tilde{X}_a(t) - \tilde{X}|, 
\tilde{D}_b = |\tilde{C}_2 \cdot \tilde{X}_b(t) - \tilde{X}|, 
\tilde{D}_d = |\tilde{C}_3 \cdot \tilde{X}_d(t) - \tilde{X}| \quad (9)
$$

The encircling and attacking of prey repeatedly continues until an optimal solution is achieved or it reaches maximum iterations.

**GOA**

GOA is a newly developed meta-heuristic optimization algorithm based on grasshoppers swarming behaviour (Saremi et al. 2017). The function to analyze the swarming behaviour is well-defined as follows (Mafarja et al. 2019):

$$
X_i = S_i + G_i + A_i \quad (10)
$$

where $X_i$ is the $i$th grasshopper position, $S_i$ is classical interaction, $G_i$ is gravity force and $A_i$ is wind advection. $S_i$ is expressed as:

$$
S_i = \sum_{j=1}^{N} s(d_{ij}) \frac{\tilde{d}_{ij}}{} \quad (11)
$$

where $s$ designates social forces and $d_{ij}$ is the distance between the $i$th and $j$th grasshoppers.

$$
\tilde{d}_{ij} = \frac{x_j - x_i}{d_{ij}} \quad (12)
$$

Here, $s$ can be defined as:

$$
s(t) = fe^{-t} - e^{-t} \quad (13)
$$

where $f$ is intensity of attraction and $l$ is scale of attractive length. $G$ is expressed as:

$$
G = -g\tilde{e}_k \quad (14)
$$
where \( \vec{e}_b \) is the unit vector and \( g \) is the gravitational constant. \( A_i \) is expressed as:

\[
A_i = u \vec{e}_w
\]

where \( \vec{e}_w \) is the unit vector and \( u \) is continuous drift. Lastly, a new location is computed using the following equation:

\[
X_i = \sum_{j=1}^{N} s(|x_j - x_i|) \frac{x_j - x_i}{d_{ij}} - g \vec{e}_b + u \vec{e}_w,
\]

where \( N \) is the number of grasshoppers. Moreover, an improved function is used for updating the grasshopper’s location:

\[
X_i^d = c \left[ \sum_{j=1}^{N} c \frac{u b_d - l b_d}{2} s(|x_j^d - x_i^d|) \frac{x_j^d - x_i^d}{d_{ij}} \right] + T_d
\]

where \( l b_d \) is the lower bound; \( u b_d \) is the upper bound; \( T_d \) is the \( D \)th dimension value; and \( c \) is a declining coefficient, repulsion zone and attraction zone. Figure 5 represents the flowchart of GOA. Equation (17) is computed on the basis of location of the best grasshopper (target), its present location and the location of all other grasshoppers.

To balance exploitation and exploration, parameter \( c \) is needed to be reduced in proportion to the number of iterations. This mechanism promotes exploitation as there is an increase in iteration count. It is relevant to emphasize the variable \( C \) insertion calculated with Equation (18):

\[
c = c_{max} - \frac{c_{max} - c_{min}}{L}
\]

**Pseudo-code of GOA**

Initialize the swarm \( X_i \) (\( i = 1, 2, 3, \ldots, n \))

- Initialize \( c_{max}, c_{min}, \) and maximum number of iteration
- Calculate the fitness of each search agent
- \( T = \) the best search agent

\[ \textbf{while} \ (1 < \text{maximum number of iteration}) \]

- Update using Eq. (18)
  - for each search agent
    - Normalize the distance between grasshoppers in [1, 4]
    - Update the position of the current search agent by the Equation (17)
    - Bring the current search agent back if it goes outside the boundaries
  - Update \( T \) if there is a better solution

- \( 1 = 1 + 1 \)

\[ \textbf{end while} \]

Return \( T \)

**Proposed hybrid methodology**

In the present study, GOA and GWO algorithms were applied for evaluating optimum values and training ANFIS. For developing ANFIS-GOA and ANFIS-GWO, optimum GOA and GWO parameters can be set on
the basis of abundant parametric studies. For creating ANFIS-GOA and ANFIS-GWO, two codes were generated in MATLAB. In ANFIS-GOA and ANFIS-GWO, GOA and GWO helps hybrid approaches to possess a closer relationship concerning input and output. Established robust techniques can estimate more precise outcomes for nonlinear problems. Major work in a hybrid modelling approach is the appropriate selection of GOA and GWO parameters. For evaluating the best values of GOA and GWO parameters, a trial and error approach was employed and optimum values were obtained. GOA parameters are given in Table 1. Also, Figures 4 and 6 shows the training process of ANFIS by GOA and GWO algorithms.

**Preparation of data set**

Precipitation (P) data was obtained from CWC, Bhubaneswar, whereas temperature (T), solar radiation (S$_r$), humidity (H), evapotranspiration loss (E$_l$), absorption loss (A$_l$) and percolation loss (P$_l$) data were collected from IMD, Pune, for the period of 1970–2019. Data from 1970 to 2004 were used for training, and 2005–2019 for testing purposes. Required monthly data were converted from daily data which is necessary in the training and testing model. The following arrangements were applied as input:

- **Scenario 1:** P, T, S$_r$, H
- **Scenario 2:** P, T, S$_r$, H, E$_l$
- **Scenario 3:** P, T, S$_r$, H, E$_l$, A$_l$

![Flow chart of GOA.](image-url)
Table 1 | Parameters of GOA to train the ANFIS model

| Parameter                              | Values  |
|----------------------------------------|---------|
| Maximum iterations number              | 1,000   |
| Number of search agents                | 50      |
| $c_{\text{max}}$                       | 0.8     |
| $c_{\text{min}}$                       | $1 \times 10^{-7}$ |
| Attraction intensity $f$               | 0.5     |
| Attraction longitude scale $l$         | 1.5     |

Scenario 4: P, T, S, H, E, A, P

Input and output were arranged in a manner so that all data fell within a computed range previous to training. This process is called normalization where normalized values fall within 0 and 1. The equation employed for scaling normalized data is:

\[ I_t = \frac{I - I_{\text{min}}}{I_{\text{max}} - I_{\text{min}}} \quad (19) \]
where $I_t$ = converted data series, $I$ = actual input data set, $I_{\text{min}}$ = minimum of actual input data set, and $I_{\text{max}}$ = maximum of actual input data set.

**Evaluating standards**

Willmott’s index, MSE and RMSE are assessing measures for finding the best model. For a selection of the best model for the proposed area of study, MSE, RMSE must be minimum while WI is maximum.

Willmott’s index (WI)\footnote{Willmott et al. 1985} is given by:

$$\text{Willmott's index (WI)} = 1 - \frac{\left(\sum_{i=1}^{N} (x_{i,\text{comp}} - \bar{x}_{\text{comp}})^2\right)}{\left(\sum_{i=1}^{N} (x_{i,\text{obs}} - \bar{x}_{\text{obs}})^2\right)}$$

The Willmott’s index value stipulates the percentage of deviation in one variable clarified by another variable (Willmott et al. 1985):

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^{N} (x_{i,\text{comp}} - x_{i,\text{obs}})^2$$

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (x_{i,\text{comp}} - \bar{x}_{\text{comp}})(x_{i,\text{obs}} - \bar{x}_{\text{obs}})}$$

where $x_{i,\text{comp}}$ = simulated data; $x_{i,\text{obs}}$ = actual data; $\bar{x}_{\text{comp}}$ = mean simulated data; and $\bar{x}_{\text{obs}}$ = mean actual data.

**RESULTS AND DISCUSSION**

A comparative performance of four scenarios through ANFIS are assessed in this study at Jondhra, Kesinga, respectively (Supplementary Material, Tables S1 and S2). Five different transfer functions are used for ANFIS for finding the best model that can proficiently predict flood in the proposed study area. It can be observed from the results that all membership functions produced reasonable outcomes for flood prediction. For a fair comparison, various models and parameters affecting flood prediction are taken for both cases. The results show that the Gbell function gives the best value of Willmott index for all four scenarios. Among the four scenarios, model four found the prominent value of WI 0.94453 and 0.94856 (for testing), and 0.96097 and 0.96474 (for training) phases at Jondhra and Kesinga gauge station.

The ANFIS-GWO results are assessed in Supplementary Material, Tables S3 and S4 for Jondhra, Kesinga station. For computation of performance Tri, Trap, Gbell, Gauss, Pi functions are taken into consideration. The Gbell function shows the best model which possesses MSE training 0.00326, MSE testing 0.00175, RMSE training 0.04997, RMSE testing 0.01012 and WI training 0.98139, testing 0.96001 in Scenario IV at Jondhra station. Accordingly for Scenarios I, II, and III, the best values of WI are 0.95288, 0.95589, and 0.95803 in the testing phases, respectively. Similarly, at Kesinga gauge station the best values of WI are 0.95406, 0.95678, 0.95889, and 0.96208 for Scenarios I, II, III, IV in the testing phases.

Correspondingly, a hybrid ANFIS-GOA algorithm is proposed for model development at Jondhra, Kesinga station (Supplementary Material, Tables S5 and S6). The results reaveal that the best values of WI are 0.99158, 0.98993, 0.98806, 0.98638 during the training phase for Scenarios I, II, III, IV at Jondhra basin, respectively. Similary, for Kesinga basin the prominent values of WI are 0.99263, 0.99001, 0.98814, 0.98654 while considering Scenarios I, II, III, IV as input criteria, respectively. For both stations Scenario IV showed a prominent performance over the other three scenarios.

The above results show that when evapotranspiration loss is added to normal climatology indices (Scenario II) it impacts more accuraracy than Scenario I. When we consider absorption loss (Scenario III), we get better results than Scenario II. The overall results show that Scenario IV is more prominent than the other three scenarios. Senario IV is different from Scenario III in terms of percolation loss. Therefore, losses are the key constraint toward flood forecasting. Similarly, on the other hand, the new hybrid model (ANFIS-GOA) showed better performance than ANFIS, ANFIS-GWO model. However, the obtained results showed that ANFIS-GOA perfomed best among other models. It has a lower coefficient of variation compared to other
models and optimization algorithms. The major contribution of this research is the assessment of flood prediction potential by hybrid models based on parametric effects owing to regulatory weights and parameters formed in the training phase of models.

Assessment of outcome for different models

Appraisals of ANFIS, ANFIS-GWO and ANFIS-GOA model at training and testing periods for all proposed gauge stations are presented in Figures 7 and 8. The paramount value of WI for ANFIS, ANFIS-GWO and ANFIS-GOA models was 0.96097, 0.98139, and 0.99158, respectively, for Jondhara station. Similarly, for Kesinga

![Figure 7](http://iwaponline.com/h2open/article-pdf/4/1/137/917228/h2oj0040137.pdf)

**Figure 7** | Observed versus predicted flood model using: (a) ANFIS, (b) ANFIS-GWO and (c) ANFIS-GOA from (i) Jondhara (ii) Kesinga gauge station (training).
station the preeminent value of WI was 0.96474, 0.98216, and 0.99263 for ANFIS, ANFIS-GWO and ANFIS-GOA models, respectively. A detailed graphical representation of actual versus predicted flood with respect to WI for different gauge stations in Figure 7.

**Assessment of actual flood versus simulated flood at Jondhara and Kesinga**

Figures 9 and 10 reveal the variation of actual versus simulated flood. Outcomes reveal that projected peak floods are 5,241, 5,372 and 5,442 m³/s for ANFIS, ANFIS-GWO and ANFIS-GOA, respectively, against the observed
Figure 9 | Monthly observed versus simulated flood model using: (a) ANFIS, (b) ANFIS-GWO and (c) ANFIS-GOA from Jondhra gauge station.
Figure 10 | Monthly observed versus simulated flood model using: (a) ANFIS, (b) ANFIS-GWO and (c) ANFIS-GOA from Kesinga gauge station.
peak of 5,568 m³/s for Jondhra gauge site. For Kesinga, projected peak floods are 5,036, 5,145 and 5,199 m³/s for ANFIS, ANFIS-GWO and ANFIS-GOA, respectively, contrary to the observed peak of 5,308 m³/s, as presented in Figure 10. This indicates a significant impact on flood and was found to be beneficial for flash flood regions with a predictive flood index.

Figure 11 shows a box-plot of observed and simulated flood values from 1970 to 2019 in Kesinga and Jondhra gauge stations. Assessment of ANFIS, ANFIS-GWO and ANFIS-GOA techniques with observed values reveals that the ANFIS-GOA method can make high flood estimations. Moreover, for ANFIS the box area is minimum which approves lower accurateness of ANFIS in comparison to other models.

A histogram showing the ratio of simulated and observed flood values for ANFIS, ANFIS-GWO and ANFIS-GOA models have been presented for assessing the frequency of datum points in a number of selected error bins. Here, the total number of months binned on the x-axis has been analysed where the probability of occurrence for any given time series has been checked. A close investigation of simulated and observed flood by ANFIS-GOA and its relative models is shown in Figures 12 and 13. These signify probability distribution of data. It is significant to understand that these plots are very necessary for representing probability occurrence of a specified flood value inside a particular interval. Based on this amount of model precision, it is apparent that probability distribution of predicted flood values by the ANFIS-GOA model were very near to the observed flood values for most intervals as presented in Figures 12 and 13.

Comparison of model performance

MSE, RMSE and WI indicators are used for evaluating the performance of ANFIS, ANFIS-GWO and ANFIS-GOA models for two gauge basins. An assessment of the performance indicators is listed in Table 2, illustrating

![Image of box plots]

**Figure 11** | Comparison of box plot for various techniques.
the efficiency of each model. Evaluating flood is very significant and hence proposed the methods applied in the present study are important to demonstrate flood prediction information. Therefore, calculation of RMSE, WI, and MSE values is vital to predict flood. It is apparent that the ANFIS-GOA model executed well compared to ANFIS and ANFIS-GWO for four scenarios. Evaluation and assessment are conducted for studying the performance of the models.

The obtained results clearly show the advantages of GOA to solve real-world problems with unidentified search space. The success of GOA is due to many reasons. In the preliminary steps of optimization, the exploration capability of GOA is high because of the huge repulsion rate amid grasshoppers. This helps GOA in broadly exploring search space and discovering its favorable areas. Then, in the final steps of optimization, exploitation is high because of higher attraction forces amid grasshoppers. This behaviour causes a local search and improves the accurateness of solutions found in the exploration stage. GOA efficiently balances exploitation and exploration, primarily focusing on avoiding local optima, and then conjunction. Suggestion of an adaptive comfort zone coefficient is the reason behind this behaviour. Steady declining of this constituent carries grasshoppers nearer

---

**Table 2 | Comparison of performance index**

| Gauge station | Techniques | MSE          | RMSE          | WI            |
|---------------|------------|--------------|---------------|---------------|
|               |            | Training     | Testing       | Training      | Testing       |
| Jondhra       | ANFIS      | 0.00491      | 0.00398       | 0.06014       | 0.02008       | 0.96097       | 0.94453       |
|               | ANFIS-GWO  | 0.00326      | 0.00175       | 0.04997       | 0.01012       | 0.98139       | 0.96001       |
|               | ANFIS-GOA  | 0.00165      | 0.00052       | 0.01889       | 0.0019        | 0.99158       | 0.9741        |
| Kesinga       | ANFIS      | 0.00436      | 0.00332       | 0.05885       | 0.01928       | 0.96474       | 0.94856       |
|               | ANFIS-GWO  | 0.00313      | 0.00165       | 0.04889       | 0.00937       | 0.98216       | 0.96208       |
|               | ANFIS-GOA  | 0.00162      | 0.00048       | 0.01765       | 0.00183       | 0.99263       | 0.97502       |
to target which is proportional to the number of iterations. Finally, the suggested target chasing mechanism necessitates GOA for saving the best solution attained so far as target, and driving grasshoppers towards it with a hope to improve its accurateness or find a superior one in the search space. In view of simulation, result, discussions, and analysis of this research, we believe that GOA is capable of solving several optimization problems efficiently. It considers a specified optimization problem as a black box, thus it does not require gradient information of search space. Hence, it can be employed to any optimization problem in various fields conditional on appropriate problem formulation.

The above results indicated that performance of ANFIS-GOA in terms of RMSE and WI value is prominent compared to hybrid ANN with KNN (Kan et al. 2020); ANFIS-GA, ANFIS-PSO, ANFIS-ACO (Azad et al. 2018); hybrid deep learning ConvLSTM (Moishin et al. 2021); FSF-ARIMA (Banihabib et al. 2020) and ANFIS (Rezaeianzadeh et al. 2014).

Sensitivity analysis
In this research, the sensitivity of precipitation (P), temperature (T), solar radiation (Sr), humidity (H), evapotranspiration loss (El), absorption loss (Al), percolation loss (Pl) and constraints toward flood prediction through different machine learning approaches are discussed. At first, we developed a model using precipitation, temperature, solar radiation, humidity parameters and found the model efficacy at two proposed stations. Second, evapotranspiration loss was included with the previous model and found better results than Scenario 1. Similarly, inclusion of absorption loss (Al) with Scenario 2 for all techniques found more performance value than Scenario 2. While we considered percolation loss (Pl) as an input constraint (Scenario 4) with the previous arrangement (Scenario 3) for model development, it gives preeminent value performance for all five machine learning approaches during both training and testing phases. Moreover, we found that all three losses (evapotranspiration loss, absorption loss, percolation loss) possess a sensitive effect towards flood prediction for all five proposed machine learning algorithms. Also, it was found that the ANFIS-GOA model gives more sensitive performance value than other proposed machine learning approaches. As the proposed study area is within a flood prone region, development of ANN models will aid in assessing flood discharge. These results suggest the most appropriate methods for developing environmental concerns for estimating flood in the stations of any flood region. However, a combination of techniques needs to be examined for improving the conjoint modelling techniques for the future.

Limitations and future scope
A major disadvantage of applying hybrid machine learning algorithms is that the training time of the model increases after hybridization of machine learning and meta-heuristic algorithms, particularly while dealing with certain complex problems. Also, they are classifier specific methods and depend on a combination of different feature selection methods. Taking into consideration the advantages and disadvantages of various algorithms, conjoining search strategies of different algorithms for generating a novel algorithm is a burning research matter. The prediction performances of data-driven models are subject to quality and quantity of data. This study is conducted in a specific location (Jondhra and Kesinga gauge stations of Mahanadi river basin). The scope of the present study can be extended by applying ML models to various other geographical locations. Selection of best input combinations for a particular model can possibly vary with changes in default model operators. This assessment of selecting best input combinations using different approaches could be an interesting subject for future research. Moreover, in the direction of future research, it is significant to mention that not all rules in model architecture are vital; hence, it is essential to reduce the complexity of trained models by removing non-contributing rules leading to a decrease in the computational cost of the network. To improve proposed methods more state-of-the-art AI methods, for example the probabilistic and ensemble forecasting methods, could be combined with data-driven models for reducing uncertainties in multistep ahead flood forecast in future research.

CONCLUSION
This study investigates the potential of new hybrid models combining GOA and GWO algorithms with the ANFIS (ANFIS-GOA and ANFIS-GWO) model for prediction of flood events. To achieve this objective, Jondhra and Kesinga stations located on the Mahanadi River were chosen as the case study. GOA and GWO were developed to optimize the parameters of ANFIS and were then compared with simple ANN and ANFIS models. The
outcomes of ANFIS, ANFIS-GWO and ANFIS-GOA models was compared and assessed on the basis of their training and testing performances based on performance indicators.

1. The results revealed that integration of AI models with optimization algorithms proved to be more effective and accurate compared to customary AI.

2. Comparing the results of ANFIS, ANFIS-GWO and ANFIS-GOA models, it could be inferred that the values of MSE and RMSE of ANFIS-GOA models were less than those of ANFIS-GWO and AI models. However, the WI value of the ANFIS-GOA model has higher sensitivity compared to other applied models. Hence, it was observed that ANFIS-GOA and ANFIS-GWO models improved prediction accurateness of the conventional ANFIS model with the ANFIS-GOA model performing slightly better than ANFIS-GWO for all input combinations.

3. It appears that application of a meta-heuristic optimization algorithm in the training period of the ANFIS technique can decrease the flaw of conventional optimization algorithms and lead to more consistent outcomes. Also, it was proved that climatic losses sensitively affect flood prediction.

4. Additionally, in certain case studies, it is essential to develop a real-time forecasting model, and therefore, time required to execute the model should be less. Hence, application of more newly developed search algorithms is necessary for achieving a fast convergence process as well.

**DATA AVAILABILITY STATEMENT**

All relevant data are included in the paper or its Supplementary Information.

**REFERENCES**

Alizadeh, Z., Yazdi, J., Mohammadiun, S., Hewage, K. & Sadiq, R. 2019 Evaluation of data driven models for pipe burst prediction in urban water distribution systems. *Urban Water Journal* 16 (2), 136–145.

Araghinejad, S. 2013 *Data-driven Modeling: Using MATLAB® in Water Resources and Environmental Engineering*. Springer Science & Business Media, Dordrecht, Heidelberg, London.

Azad, A., Farzin, S., Kashi, H., Sanikhani, H., Karami, H. & Kisi, O. 2018 Prediction of river flow using hybrid neuro-fuzzy models. *Arabian Journal of Geosciences* 11 (22), 1–14.

Banihabib, M. E., Bandari, R. & Valipour, M. 2020 Improving daily peak flow forecasts using hybrid Fourier-series autoregressive integrated moving average and recurrent artificial neural network models. *AI* 1 (2), 263–275.

Bisht, D. C. S. & Jangid, A. 2011 Discharge modelling using adaptive neuro-fuzzy inference system. *International Journal of Advanced Science & Technology* 51, 99–114.

Campolo, M., Soldati, A. & Andreussi, P. 2005 Artificial neural network approach to flood forecasting in the River Arno. *Hydrological Sciences Journal* 48, 381–398.

Chen, S., Lin, Y., Chang, L. & Chang, F. 2006 The strategy of building a flood forecast model by neuro-fuzzy network. *Hydrological Processes* 20, 1525–1540.

Das, U. K., Samantaray, S., Ghose, D. K. & Roy, P. 2019 Estimation of Aquifer Potential Using BPNN, RBFN, RNN, and ANFIS. *Smart Intelligent Computing and Applications*. Springer, Singapore, pp. 569–576.

Deghbouch, H. & Debbat, F. 2021 A hybrid bees algorithm with grasshopper optimization algorithm for optimal deployment of wireless sensor networks. *Inteligencia Artificial* 24 (67), 18–35.

Dehghani, M., Riahi-Madvar, H., Hooshyaripor, F., Mosavi, A., Shamshirband, S., Zavadskas, E. K. & Chau, K. W. 2019a Prediction of hydropower generation using grey wolf optimization adaptive neuro-fuzzy inference system. *Energies* 12, 289–309.

Dehghani, M., Scifi, A. & Riahi-Madvar, H. 2019b Novel forecasting models for immediate-short term to long-term inflow prediction by combining ANFIS and grey wolf optimization. *Journal of Hydrology* 576, 698–725.

El-Shafie, A., Taha, M. R. & Noureldin, A. 2007 A neuro-fuzzy model for inflow forecasting of the Nile river at Aswan high dam. *Water Resources Management* 21, 533–556.

Ewees, A. A. & Elaziz, M. A. 2018 Improved adaptive neuro-fuzzy inference system using gray wolf optimization: a case study in predicting biochar yield. *International Journal of Intelligent Systems* 29, 924–940.

Farrokhi, A., Farzin, S. & Mousavi, S. F. 2020 A new framework for evaluation of rainfall temporal variability through principal component analysis, hybrid adaptive neuro-fuzzy inference system, and innovative trend analysis methodology. *Water Resources Management* 34 (10), 3363–3385.

Fattahi, H. & Hasanipanah, M. 2021 An integrated approach of ANFIS-grasshopper optimization algorithm to approximate flyrock distance in mine blasting. *Engineering with Computers* 1–13.

Ghorbani, M. A., Zadeh, H. A., Isazadeh, M. & Terzi, O. 2016 A comparative study of artificial neural network (MLP, RBF) and support vector machine models for river flow prediction. *Environmental Earth Sciences* 75, 476–490.
Guo, J., Zhou, J., Lu, J., Zou, Q., Zhang, H. & Bi, S. 2014 Multi-objective optimization of empirical hydrological model for streamflow prediction. *Journal of Hydrology* 511, 242–253.

Jiang, P., Gautam, M. R., Zhu, J. & Yu, Z. 2013 How well do the GCMs/RCMs capture the multi-scale temporal variability of precipitation in the Southwestern United States? *Journal of Hydrology* 479, 75–85.

Jiang, P., Yu, Z., Gautam, M. R. & Acharya, K. 2016 The spatiotemporal characteristics of extreme precipitation events in the western United States. *Water Resources Management* 30, 4807–4821.

Kan, G., Liang, K., Yu, H., Sun, B., Ding, L., Li, J., He, X. & Shen, C. 2020 Hybrid machine learning hydrological model for flood forecast purpose. *Open Geosciences* 12 (1), 813–820.

Khalifeh, S., Esmaili, K., Khodashenas, S. & Akbarifard, S. 2020 Data on optimization of the non-linear Muskingum flood routing in Kardeh River using Goa algorithm. *Data in Brief* 30, 1–7.

Mafarja, M., Aljaraif, I., Faris, H., Hammouri, A. I., Ala’M, A. Z. & Mirjalili, S. 2019 Binary grasshopper optimization algorithm approaches for feature selection problems. *Expert Systems with Applications* 117, 267–286.

Maier, H. R. & Dandy, G. C. 2000 Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications. *Environmental Modelling and Software* 15, 101–124.

Mirjalili, S., Mirjalili, S. M. & Lewis, A. 2014 Grey wolf optimizer. *Advances in Engineering Software* 69, 46–61.

Mirjalili, S. Z., Mirjalili, S., Saremi, S., Faris, H. & Aljaraif, I. 2018 Grasshopper optimization algorithm for multi-objective optimization problems. *Applied Intelligence* 48 (4), 805–820.

Moayedih, H., Nguyen, H. & Foong, L. K. 2019 Nonlinear evolutionary swarm intelligence of grasshopper optimization algorithm and gray wolf optimization for weight adjustment of neural network. *Engineering with Computers* 37, 1265–1275.

Moishin, M., Deco, R. C., Prasad, R., Raj, N. & Abdulla, S. 2021 Designing deep-based learning flood forecast model with ConvLSTM hybrid algorithm. *IEEE Access* 9, 50982–50993.

Montano, J., Tobón, A. F., Villegas, J. P. & Durango, M. 2020 Grasshopper optimization algorithm for parameter estimation of photovoltaic modules based on the single diode model. *International Journal of Energy and Environmental Engineering* 11, 367–375.

Moore, R. J., Bell, V. A. & Jones, D. A. 2005 Forecasting for flood warning. *Comptes Rendus Geoscience* 357, 203–217.

Nabipour, N., Dehghani, M., Mosavi, A. & Shamshirband, S. 2020 Short-term hydrological drought forecasting based on different nature-inspired optimization algorithms hybridized with artificial neural networks. *IEEE Access* 8, 15210–15222.

Panigrahi, B. K., Das, S., Nath, T. K. & Senapati, M. R. 2018 An application of data mining techniques for flood forecasting: application in rivers Daya and Bhargavi, India. *Journal of the Institute of Engineers (India): Series B* 99, 331–342.

Prasad, R., Deo, R. C., Li, Y. & Maraseni, T. 2017 Input selection and performance optimization of ANN-based stream forecasts in the drought-prone Murray Darling Basin region using IIS and MODWT algorithm. *Atmospheric Research* 197, 42–63.

Rath, A., Samantaray, S., Bhoi, K. S. & Swain, P. C. 2017 Flow forecasting of Hirakud reservoir with ARIMA model. In: 2017 *International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS)*. IEEE, India, pp. 2952–2960.

Rezaeianzadeh, M., Tabari, H., Yazdi, A. A., Isik, S. & Kalin, L. 2014 Flood flow forecasting using ANN, ANFIS and regression models. *Neural Computing and Applications* 25, 25–37.

Sahoo, G. B., Ray, C. & Wade, H. F. 2005 Pesticide prediction in ground water in North Carolina domestic wells using artificial neural networks. *Ecological Modelling* 183, 29–46.

Samantaray, S. & Sahoo, A. 2020 Estimation of flood frequency using statistical method: Mahanadi river basin, India. *H2Open Journal* 3 (1), 189–207.

Saremi, S., Mirjalili, S. & Lewis, A. 2017 Grasshopper optimization algorithm: theory and application. *Advances in Engineering Software* 105, 30–47.

Seifi, A., Ehteram, M., Singh, V. P. & Mosavi, A. 2020 Modeling and uncertainty analysis of groundwater level using six evolutionary optimization algorithms hybridized with ANFIS, SVM, and ANN. *Sustainability* 12 (10), 4023–4065.

Senapati, M. R., Vijaya, I. & Dash, P. K. 2007 Rule extraction from radial basis functional neural networks by using particle swarm optimization. *Journal of Computer Science* 3 (8), 592–599.

Shu, C. & Ouarda, T. 2008 Regional flood frequency analysis at ungauged sites using the adaptive neuro-fuzzy inference system. *Journal of Hydrology* 349, 31–43.

Singh, K. P., Basant, A., Malik, A. & Jain, G. 2009 Artificial neural network modeling of the river water quality – a case study. *Ecological Modelling* 220, 888–895.

Steczek, M., Jefimowski, W. & Szlag, A. 2020 Application of grasshopper optimization algorithm for selective harmonics elimination in low-frequency voltage source inverter. *Energies* 13 (25), 6426–6442.

Sudheer, K. P. & Jain, A. 2004 Explaining the internal behaviour of artificial neural network river flow models. *Hydrological Processes* 18, 833–844.

Takagi, T. & Sugeno, M. 1985 Fuzzy identification of systems and its applications to modeling and control. *IEEE Transactions on Systems, Man and Cybernetics* 1, 116–132.

Taormina, R., Chau, K-W. & Sethi, R. 2012 Artificial neural network simulation of hourly groundwater levels in a coastal aquifer system of the Venice lagoon. *Engineering of Artificial Intelligence* 25, 1670–1676.
Tikhamarine, Y., Souag-Gamane, D., Ahmed, A. N., Kisi, O. & El-Shafie, A. 2020 Improving artificial intelligence models accuracy for monthly streamflow forecasting using grey wolf optimization (GWO) algorithm. *Journal of Hydrology* **582**, 1–16.

Wang, Z., Lai, C., Chen, X., Yang, B., Zhao, S. & Bai, X. 2015 Flood hazard risk assessment model based on random forest. *Journal of Hydrology* **527**, 1130–1141.

Willmott, C. J., Ackleson, S. G., Davis, R. E., Feddema, J. J., Klink, K. M., Legates, D. R., O'donnell, J. & Rowe, C. M. 1985 Statistics for the evaluation and comparison of models. *Journal of Geophysical Research: Oceans* **90** (C5), 8995–9005.

Wu, C. L. & Chau, K. W. 2011 Rainfall–runoff modeling using artificial neural network coupled with singular spectrum analysis. *Journal of Hydrology* **399**, 394–409.

Yan, H., Zou, Z. & Wang, H. 2010 Adaptive neuro fuzzy inference system for classification of water quality status. *Journal of Environmental Science* **22**, 1891–1896.

Yaseen, Z. M., Ebtehaj, I., Bonakdari, H., Deo, R. C., Mehr, A. D., Mohtar, W. H. M. W. & Singh, V. P. 2017 Novel approach for streamflow forecasting using a hybrid ANFIS-FFA model. *Journal of Hydrology* **554**, 263–276.

Young, P. C. 2002 Advances in real–time flood forecasting. *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences* **360**, 1433–1450.

Yu, Z., Jiang, P., Gautam, M. R., Zhang, Y. & Acharya, K. 2015 Changes of seasonal storm properties in California and Nevada from an ensemble of climate projections. *Journal of Geophysical Research: Atmospheres* **120**, 2676–2688.

Zhang, F., Dai, H. & Tang, D. 2014 A conjunction method of wavelet transform-particle swarm optimization-support vector machine for stream flow forecasting. *Journal of Applied Mathematics* **2014**, 1–10.

Zhou, Y., Guo, S. & Chang, F-J. 2019 Explore an evolutionary recurrent ANFIS for modelling multi-step-ahead flood forecasts. *Journal of Hydrology* **570**, 345–355.

First received 23 March 2021; accepted in revised form 6 June 2021. Available online 12 July 2021