Contents lists available at ScienceDirect

MethodsX

journal homepage: www.elsevier.com/locate/mex

Method Article

Prediction of Flood Discharge Using Hybrid PSO-SVM Algorithm in Barak River Basin

Sandeep Samantaray^{a,*}, Abinash Sahoo^b, Ankita Agnihotri^b

^a Department of Civil Engineering, NIT Srinagar, India

^b Department of Civil Engineering, NIT Silchar, India

ARTICLE INFO

Method name: PSO-SVM

Keywords: Bpnn Svm Pso-svm Flood

ABSTRACT

A crucial necessity in integrated water resource management is flood forecasting. Climate forecasts, specifically flood prediction, comprise multifaceted tasks as they are dependant on several parameters for predicting the dependant variable, which varies from time to time. Calculation of these parameters also changes with geographical location. From the time when Artificial Intelligence was first introduced to the field of hydrological modelling and prediction, it has produced enormous attention in research aspects for additional developments to hydrology. This study investigates the usability of support vector machine (SVM), back propagation neural network (BPNN), and integration of SVM with particle swarm optimization (PSO-SVM) models for flood forecasting. Performance of SVM solely depends on correct assortment of parameters. So, PSO method is employed in selecting SVM parameters. Monthly river flow discharge for a period of 1969 - 2018 of BP ghat and Fulertal gauging sites from Barak River flowing through Barak valley in Assam, India were used. For obtaining optimum results, different input combinations of Precipitation (P_t), temperature (T_t), solar radiation (Sr), humidity (H_t), evapotranspiration loss (E_l) were assessed. The model results were compared utilizing coefficient of determination (R²) root mean squared error (RMSE), and Nash-Sutcliffe coefficient (N_{SF}). The most important results are highlighted below.

- First, the inclusion of five meteorological parameters improved the forecasting accuracy of the hybrid model.
- Second, model comparison specifies that hybrid PSO-SVM model executed superior performance with RMSE- 0.04962 and NSE- 0.99334 compared to BPNN and SVM models for monthly flood discharge forecasting.
- Third, applied optimization algorithm has easy implementation, simple theory, and high computational efficacy.

Results revealed that PSO-SVM could be utilised as an improved alternate method for flood forecasting as it provided a higher degree of reliability and accurateness.

* Corresponding author. E-mail addresses: samantaraysandeep963@gmail.com (S. Samantaray), abinash_rs@civil.nits.ac.in (A. Sahoo).

https://doi.org/10.1016/j.mex.2023.102060

Received 6 November 2022; Accepted 2 February 2023

Available online 7 February 2023







^{2215-0161/© 2023} The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)

Specifications table

Subject area:	Engineering
More specific subject area:	Geology and Hydrology, Flood
Name of your method:	PSO-SVM
Name and reference of original method:	C Zhou, K Yin, Y Cao, B Ahmed, (2016) Application of time series analysis and PSO-SVM model in predicting the
	Bazimen landslide in the Three Gorges Reservoir, China. Engineering Geology, 204, 108–120
Resource availability:	MATLAB (validation and Methodology), Origin pro-2020 (Results and graphical analysis), GIS (Study Area)

Introduction

Floods are responsible for significant destruction of social and economic amenities and also loss of human life. Flooding possess a potential threat to heavily inhabited regions positioned downstream of and next to major river bodies. Precise flood flow forecasting is a necessity to reduce flooding risk and is significant for water resources planning and management. But, precise river discharge forecasting is a challenging issue as analysis of river flood level is a multifaceted, dynamic procedure categorized by temporal and spatial variations. Many prediction methods, which entail a huge volume of data to forecast, have been suggested for mitigating or preventing impact of flood conditions. Due to changes in global climatic conditions, there is a rise in occurrence of flood extremes around the world [10,18] in rural catchments, increasing sternness of flood hazards.

In India, many destructive flooding occurred recently, like the 2008 Bihar flood in River Kosi, 2005 Mumbai flood in River Mithi, cloudburst in Uttarakhand in 2013, Chennai flood in 2015 and Assam in 2022. Therefore, a reliable flood forecasting method is needed and very essential for helping the local and high-level government in reducing the damage and loss of life and property. Different ANN (artificial neural network) models provide good forecasting results for various problems. Prediction of flood discharge is done using ANNs at different gauging stations [1,28,29]. Yet, these models were subjected to slow learning and local convergence. Hence, they could not achieve agreeable performance while resolving complicated hydrological processes. In recent times, SVM model has emerged as an efficacious computational approach as it overcomes several of ANN's shortcomings and is therefore found in many applications for various areas of engineering and science [17,18,46]. Even in the field of hydrological modelling, investigators have extensively applied SVM.

Han et al. [14] described applicability of SVM over Bird Creek watershed, USA. They addressed a few significant problems on development and usability of SVM to forecast flood events considering an optimum number of input parameters and combinations. Study results showed that SVM can be a reliable technique used for different rainfall inputs and performed well under different evaluation criteria. Nikam and Insom et al. [17] applied a novel method to develop an improved SVM model based on particle filter to estimate suitable model parameters for enhanced flood classification system. Based on obtained results proposed method provided a more accurate analysis and performed superior to standard SVM model. Latt and Wittenberg [21] examined performance of stepwise-MLP (multi-layer perceptron) and ANN models for multi-step flood forecasts of Chindwin River in north Myanmar. ANN models performed superiorly compared to SMLP models. Liu et al. [23] proposed combined stacked auto-encoders (SAE) with BPNN for stream flow prediction and compared with SVM, RBFN, BPNN, and ELM (extreme learning machine) models. Outcomes showed that projected hybridised model performed much better than other models. Macharyulu et al. [52] applied CNN and MLP models for developing a flood prediction model considering hydrological data from Subarnarekha river basin. Various artificial intelligence approaches are applied to predict different climatic indices at various watersheds in India ([31–33]). In several case studies, SVM

With an increasing number of applications in forecasting flood events, SVM is a dependable technique for computing extensive non-linear time-series data [15,19]. SVM's performance mainly depends on hyperparameter and kernel function choices [25]. Most frequently utilised kernel function is RBF kernel which is used to develop SVM models (Raghavendra and [8]). However, there are certain drawbacks related to SVM; while determining these parameters, sufferings from dimensionality and large time consumption are experienced during data analysis [53]. To overcome such drawbacks, scientists have developed several evolutionary optimisation algorithms inspired by nature which include ACO (ant colony optimisation), GA (genetic algorithm), PSO, ABC (artificial bee colony), DE (differential evolution), etc. Integrated models have been developed utilising metaheuristic algorithms for determining parameters of techniques like ANFIS (adaptive neuro-fuzzy inference system), SVM, MLP, etc. Hence, in this study, PSO algorithm is applied to determine optimum SVM parameters in an SVM-based (PSO-SVM) flood forecasting model.

Wang et al. [45] used a combination of SVM with PSO (SVM-PSO) model to forecast annual reservoir inflow of Fengtan reservoir situated downstream of Yuan river basin. Obtained results were compared with ANN model which discovered that SVM-PSO model showed better forecasting performance. Guo et al. [13] applied SVM-PSO for monthly stream flow estimation of Changjiang River basin and compared obtained results with ANN and traditional SVM model. Findings verified that developed SVM model had a better ability to generalize and higher prediction accurateness. Wang et al. [44] applied simple SVM and hybrid PSO-SVM models to forecast monthly runoff of Dahuofang reservoir. Results revealed that hybrid model produced more accurate predictions compared to simple SVM model at proposed study area. Ch et al. [5] explored efficiency of hybrid SVM model combined with quantum-behaved PSO (SVM-QPSO) to predict monthly streamflows of Polavaram and Vijayawada stations of Andhra Pradesh, India. Based on evaluation measures proposed SVM-QPSO model performed better compared to other conventional models with better accuracy. Li et al. [22] developed SVM model based on boosting algorithm and kernel PCA for forecasting flood occurrences of Wangjiaba station at River Huaihe. Results revealed that developed ensemble model enhanced flood forecasting precision efficiently. Ebtehaj

et al. [9] evaluated average stream velocity below sediment deposition constraints in partly occupied rounded storm sewers utilising SVM model combined with FFA (firefly algorithm). Enactment of hybrid tool was contrasted with those of SVM, genetic programming (GP), and ANN models, and findings indicate that SVM-FFA predicted dens metric particle Froude number at settlement limit very precisely.

Zaini et al. [47] applied a conjoint SVM model combined with PSO for forecasting short-term daily river flow at Upper Bertam watershed. Obtained results were compared against simple SVM model, which revealed that hybridized SVM-PSO model outperformed the simple SVM model and provided more accurate results. Zaini et al. [48] investigated potential of two hybrid models, namely bat algorithm integrated with BPNN (Bat-BPNN) and SVM-PSO, for forecasting monthly stream flow at River Kuantan situated in Malaysia and compared with traditional BPNN and SVM models. Outcomes specified that hybrid Bat-BPNN and SVM-PSO produced better performance than regular BPNN and SVM models, respectively, in stream flow forecasting. Chen et al. [7] proposed an improved GA combined with BPNN (IGA-BPNN) to predict water levels at middle and lower stretches of River Han, China. The performance of IGA-BPNN model was compared with traditional artificial intelligence models which revealed that IGA-BPNN improved prediction accuracy and model stability for desired study area. Robati and Iranmanesh [50] developed a model for studying the inflation rate in Iran using ANFIS-PSO model. Applied method was found to be effective and reliable based on the statistical measures. Kao et al. [54] applied Long Short-Term Memory and feed forward neural network combined with Encoder-Decoder (FFNN-ED and LSTM-ED) models for forecasting multi-step ahead flood using hydrological data from Shihmen Reservoir, Taiwan. Results revealed that LSTM-ED produced more accurate and reliable flood forecasting compared to FFNN-ED. Adikari et al. [55] compared potential of LSTM, Convolutional Neural Network (CNN), and Wavelet-ANFIS in forecasting flooding and drought events. They found that CNN showed best performance in forecasting flood whereas WANFIS in forecasting drought, irrespective of the climatic condition of the study region. Xiujia et al. [51] investigated the efficiency of SVM, BPNN, Wavelet Neural Network (WNN), SVM-PSO, BPNN-PSO, and WNN-PSO for calculating Crop Water Deficit Index (CWDI) to predict drought of regional crops in Fuxin City, China. Compared outcomes indicated that WNN-PSO model performed superior compared to all other models. Kamali and Asghari [56] applied SVM-SPO model to investigate impact of droughts on groundwater storage and measure the effect of impending drought occurrences on Najafabad aquifer in Iran based on variations in climatic conditions. Agnihotri et al. [1] proposed ANFIS combined with ACO algorithm for flood prediction at Matijuri gauging site of Barak basin, India. Outcomes revealed that proposed optimization algorithm had good ability of enhancing accuracy of standard ANFIS. Aswad et al. [57] applied Random Forest, Decision Tree (DT), and Decision Jungle for predicting flood process. Their research findings show that DT provided highest accuracy amongst the three algorithms.

In this work, a hybrid optimisation technique known as PSO-SVM is introduced to predict flood discharge and performance of the model is evaluated based on quantitative statistical measures. Also, a comparison is made between the hybrid and conventional approaches. Here two conventional techniques BPNN and SVM are applied to predict flow and compared with the results of PSO-SVM algorithm. PSO algorithm being a robust technique discovers out search space in a proficient way and helps in finding out optimal set of SVM parameters. The novelty in this research lies in the inclusion of evapotranspiration loss parameters for flood forecasting. Though this parameter had a little impact on flood, however, it was observed that addition of evapotranspiration loss improved model accuracy and prediction results. In literature survey, it was found that there was no inclusion of evapotranspiration loss in flood forecasting studies. Also, the importance of this work lies in the identified study region. Since Assam receives major flooding every year because of glaciers melting in summer that accords with monsoon rainfall resulting in an escalation of downstream, causing the annual floods. There is no previous study conducted in the particular area which can be a flood sensitive region because of its location.

Study area

Barak River is a 900 km river that flows through Nagaland, Assam, Mizoram, and Manipur states of India and via Bangladesh into Bay of Bengal. It originates from its source at Village Liyai Kullen situated in Manipur. Out of the total length, 524 km flows through India, 31 km on Indo-Bangladesh boundary, and rest through Bangladesh. River Barak lies in the southern portion of Assam (Fig. 1). It falls between 24°8′ to 25°8′ N latitudes and 92,⁰15′ to 93,⁰15′ E longitudes and is the second largest river in North-East India. Average annual precipitation in this valley is around 2400–4100 mm, with 80–85% from April to the middle of October. Two gauging stations BP ghat and Fulertal are considered in this study.

Methodology

BPNN

BPNN was first introduced to the field of neural networks (NNs) by Rumelhart and McClelland [27]. It is a multi-layer FFNN trained by error BP algorithm and is one of the most extensively utilized ANN models [6,12,31,34]. Learning rule involved in BPNN is to utilize steepest descent technique and repetitively adjust biases and weights of network using reverse iteration to minimize summation of squared errors [4,35].

Elementary training procedure comprises of two approaches. First approach involves forward propagation, while the second is error back propagation. In forward propagation, computation of error output is executed from input to output. Here, input data affects output node via a hidden layer, and output is produced by non-linear conversion. If actual output does not relate with simulated



Fig. 1. Study area map (Barak Valley).





Fig. 2. Continued

output, then error is back propagated for eliminating as much of it as conceivable. This is utilized for passing output error via hidden nodes, layer by layer, utilizing error signals acquired from every layer as a source to adjust weights, hence allocating error to every node in each layer. Adjustment of bias and joint strength amidst layers, error gets diminished alongside direction of gradient, and after repetitive learning preparation, network constraints are found for satisfying minimal error. BPNN utilized in present study is specified by following equation:

$$y = f_o \left[\sum_h w_{oh} \times f_h \left(\sum w_{hi} x_i + b_h\right) + b_o\right]$$
(1)

where x_i is self-regulating training and testing input variables, y is output variable, w_{hi} is weight from input to hidden; w_{oh} is weight from hidden to output; b_o and b_h is bias for output and hidden layers, in respective order; f_o and f_h are transfer functions for output and hidden layers. Architecture and flowchart of BPNN are presented in Fig. 2a and b.

SVM

Vapnik [42] proposed SVM for the first time for solving regression and classification kind of problems. It is a very capable method for data regression and classification [38,39]. Conventional machine learning techniques involve numerous difficulties, such as local

minima; low convergence rate, demand for various training data, and under fitting/over fitting [24,30]. Operating based on structural risk minimisation (SRM), SVM incapacitates these problems. SVM's output function is given by:

$$f(x) = \sum_{i=1}^{N} (\alpha_i - \alpha_i^*) K(x, z) + b_i$$
(2)

where $(\alpha_i - \alpha_i^*)$ specifies Lagrange factors, K(x, z) is inner product core function described according to Mercer theorem and b_i is bias value.

SVM is a computer-based algorithm that learns using examples to find preeminent function of classifier or hyperplane for splitting two classes in input space [36,37]. Fig. 3a and b demonstrate the architecture and flowchart of SVM model.

PSO

Kennedy and Eberhart [20] were the first to introduce PSO algorithm in 1995. PSO is employed in this study as a prevalent optimization tool. It is a swarm intelligence where birds cooperate in a swarm for searching an optimum place to obtain food. Predominantly it is based on inspiration from social movement of bird flocks and then recognized as a basic model on basis of birds probing for foodstuff. A population in a D-dimensional target search space is formed by a bird flock whose amount is m and known as "particle". If solution space of optimisation problem is D dimensions, space vector $X_i = (x_{i1}, x_{i2}, x_{i3}, \dots, x_{iD})$ is presented as *i*-th element, where X_i is location of *i*-th element, $i = 1, 2, \dots, m$ and also a conceivable explanation. It is being broadly used for solving variable optimization problems as it is very easy to understand and has great convergence rapidity [11,16,43,49]. In addition to SVM, PSO is presented for obtaining optimum constraints in SVM.

Velocity and position of particles are iterated for finding subsequent equations:

$$v_{id}^{t+1} = \omega \cdot v_{id}^{t} + c_1 \cdot r_1 \cdot \left(p_{id} - x_{id}^{t} \right) + c_2 \cdot r_2 \cdot \left(p_{gd} - x_{id}^{t} \right), \tag{3}$$

$$x_{id}^{t+1} = x_{id}^t + v_{id}^{t+1}$$
(4)

where $V_i = (v_{1i}, v_{2i}, v_{3i}, \dots v_{Di})$ is *i*-th particle velocity, and $P_i = (p_{1i}, p_{2i}, p_{3i}, \dots p_{Di})$ is ideal location of this element. Optimal swarm location is $P_g = (p_{1g}, p_{2g}, p_{3g}, \dots p_{Dg})$. Under *i*-th particle condition at *t*-th iteration, x_{id}^t and v_{id}^t are *d*-th position and constituent of speed. Positive coefficient $c_{1,22}$, r_1 and r_2 - random numbers ranging between 0 and 1; ω is inertial PSO algorithm weight. It supports multiple points search and can pact with continuous optimization problems. Therefore, we consider PSO algorithm for finding constraints of SVM to improve its performance. Flowchart of hybrid PSO-SVM structure is revealed in Fig. 4.

PSO technique in selecting SVM parameters

Performance of SVM depends on kernel function γ' parameter and regularization C' parameter. Hence C and γ must be selected carefully for increasing classification accurateness. PSO-based SVM model in learning stage is trained on basis of SRM for minimizing



Fig. 3. Architecture of SVM.



Fig. 3. Continued

training error. When there is an improvement in training error, γ and *C* are controlled using PSO. Most suitable parameters are the controlled parameters having the least error. Accordingly, optimum parameters are to be found. After finding optimised parameters of SVM, it is utilized in the retraining of SVM model. SVM classifier is now prepared for identification of new samples in testing period after passing the training period. Through the above feature selection from original dataset, a testing set is also selected. Then, imputation of testing patterns is done to the trained multi-layer SVM classifier. Following are the overall steps involved in parameter optimization of SVM model.



Fig. 4. Flowchart of hybrid PSO-SVM.



Fig. 5. Actual verses predicted graph for (a) BP Ghat, (b) Fulertal gauge station using BPNN during (i) training, and (ii) testing phases.

Step 1. Initialising parameters of particle swarm, comprising learning factors, particle number, weighting coefficient, particle velocity, and particle position.

Step 2. Encoding γ and *C* of SVM as the location of particles.

Step 3. Training SVM model with 50 training samples. Along with flying of particles, there is a variation in parameter pairs (γ , *C*).

Step 4. Calculating and assessing the particle's fitness value. Fitness value is utilised for evaluating validity of flood forecasting model with parameter combination γ and *C*. A lesser fitness value specifies greater flood forecasting accuracy.

Step 5. Terminating iteration procedure and obtaining optimum parameters of SVM if preferred accurateness is received, or continue to iterative computation.

Steps 6. and 7: Updating particle best, global best, of parameter and updating particle velocity with particle position.

Step 8. Validating SVM model with attained combinations of optimal parameter γ and *C*. The pseudo-code of PSO algorithm is presented below.

- 1. Initialisation of swarm size, position, velocity of all particles randomly
- 2. Compute the fitness of each swarm

3. Do

- 4. Compute fitness value of each particle
- 5. If fitness value is superior than pbest fix present value as new pbest
- 6. End

7. If fitness value is superior than *gbest* then select *gbest* as fitness value as among all

- particles
- 8. End
- 9. Update position of each particle using Eq. 8
- 10. Compute velocity of each particle using Eq. 9
- 11. End
- 12. While minimum error criteria or maximum iterations is not reached

Preparation of data set

Precipitation (P_t), solar radiation (S_r), temperature (T_t), humidity (H_t), and evapotranspiration loss (E_t) data (1969–2018 of monsoon months) are collected from Indian meteorological department. Data collected from 1969 to 2003 are applied for training, and from 2004 to 2018 are employed for testing. Following arrangements are considered as inputs:

Model	Input parameter	BPNN	SVM	PSO-SVM
M!1	Pt	BPNN!1	SVM!1	PSO-SVM!1
M!2	P _t , T _t	BPNN!2	SVM!2	PSO-SVM!2
M!3	P _t , T _t , S _r	BPNN!3	SVM!3	PSO-SVM!3
M!4	P_t, T_t, S_r, H_t	BPNN!4	SVM!4	PSO-SVM!4
M!5	P_t , T_t , S_r , H_t , E_t	BPNN!5	SVM!5	PSO-SVM!5

Evaluating criteria

RMSE, NSE, and R^2 are statistical assessment factors for finding best performing model. For the perfect model in this study, condition is RMSE must be least and (R^2) and NSE must be highest. The values of RMSE lies between 0 - ∞ , R^2 between 0 - 1.0 and NSE between - ∞ - 1.0.

R²(Samantaray et al. 2022a, b)

$$= 1 - \frac{\left(\sum_{i=1}^{N} I_{comp} - \bar{I}_{comp}\right)^{2}}{\left(\sum_{i=1}^{N} I_{obs} - \bar{I}_{obs}\right)^{2}}$$
(5)

NSE [58,59]

$$=\frac{\sum_{k=1}^{N}\left[I_{obs}-I_{comp}\right]^{2}}{\sqrt{\sum_{k=1}^{N}\left[I_{obs}-\bar{I}_{obs}\right]^{2}}}$$

RMSE ([60], Samantaray et al. 2022c)

$$= \sqrt{\frac{1}{n} \sum_{i=1}^{n} \left[I_{obs} - I_{comp} \right]^2}$$

Where

$$\begin{split} I_{comp} = & \text{Predicted data} \\ I_{obs} = & \text{Collected data} \\ \bar{I}_{comp} = & \text{Mean predicted data} \\ \bar{I}_{obs} = & \text{Mean collected data} \end{split}$$

Results and discussion

Three different techniques BPNN, SVM, PSO-SVM are employed here to assess their performance in flood discharge prediction. Several input scenarios are applied through BPNN algorithms to evaluate performance of model (Table 2). For assessing model performance, different constraints like R², RMSE, and MSE are considered. Five types of input scenarios in context to one output are

(6)

(7)

Table 1		
Parameters	used in PSO m	nodel.

Parameters	Value
Acceleration constants (c_1, c_2)	(2, 2)
Inertia weight ω	Varied linearly
Population size	50
Minimum error gradient	10 ⁻⁸

Table 2

Consequences of BPNN.

Station	Model	Architecture	NSE		RMSE		R ²		
			Train	Test	Train	Test	Train	Test	
	M!1	1-2-1	0.90582	0.88731	0.07752	0.03937	0.89965	0.88114	
BP Ghat	M!2	2-6-1	0.91705	0.89425	0.07590	0.03831	0.91002	0.88782	
	M!3	3-9-1	0.9218	0.89803	0.07354	0.03394	0.91774	0.89265	
	M!4	4-5-1	0.92839	0.91007	0.07268	0.03262	0.92386	0.90563	
	M!5	5-7-1	0.94001	0.91844	0.07201	0.03084	0.93636	0.91241	
	M!1	1-8-1	0.89534	0.88234	0.07576	0.03782	0.89012	0.87663	
Fulertal	M!2	2-3-1	0.90492	0.88936	0.07464	0.03673	0.89965	0.88119	
	M!3	3-2-1	0.91837	0.89197	0.07331	0.03374	0.91487	0.88465	
	M!4	4-7-1	0.9266	0.90562	0.07244	0.03129	0.92161	0.89923	
	M!5	5-4-1	0.93809	0.90865	0.07193	0.02325	0.93243	0.9008	

Table 3

Consequences of SVM.

Station	Input	NSE		RMSE		R ²	
		Train	Test	Train	Test	Train	Test
	M!1	0.92763	0.90498	0.06809	0.02232	0.92227	0.89968
BP Ghat	M!2	0.93927	0.90825	0.06794	0.02209	0.93382	0.90021
	M!3	0.94506	0.9143	0.06742	0.02175	0.93812	0.90827
	M!4	0.94772	0.92106	0.06601	0.02157	0.94065	0.91658
	M!5	0.95984	0.93991	0.06082	0.02134	0.95383	0.93483
	M!1	0.92406	0.90435	0.06788	0.02203	0.91887	0.89113
Fulertal	M!2	0.92943	0.90861	0.06782	0.02199	0.92215	0.90002
	M!3	0.9401	0.91437	0.06689	0.02168	0.93763	0.90773
	M!4	0.94847	0.92246	0.06462	0.02144	0.94001	0.91553
	M!5	0.95669	0.9389	0.05981	0.02035	0.95088	0.93437

applied to develop the model. Detail model performances of both gauge stations are specified in Table 2. Model BPNN:5 gives the preeminent value of performance amongst all the models. The paramount values of R^2 are 0.91241 and 0.9008 for BP Ghat, Fulertal gauge stations when model BPNN:5 is considered as input scenario during testing phase. Consistently for same input condition, excellent values of R^2 are 0.93636 and 0.93243 for BP Ghat, Fulertal gauge station while considering training phases. Actual predicted graph for both stations is revealed in Fig. 5.

Similarly, SVM techniques are applied with several combination of input scenarios for BP Ghat, Fulertal gauge station are specified in Table 3. On a similar way model, 5 found best models while considering SVM algorithms for both stations. Excellent values of R^2 are 0.93483, and 0.93437 for BP Ghat, Fulertal gauge station during testing phase. Consistently in case of training phase paramount value of R^2 for BP Ghat, Fulertal gauge stations are 0.95383, 0.95088 while P, T, S_r, H, E_l is considered as input scenario. Actual versus predicted graph for both stations is presented in Fig. 6.

Fig. 7 shows Actual verses predicted plots for PSO-SVM model of BP Ghat and Fulertal station. Results show that model PSO-SVM:5 gives the prominent result amongst all the scenarios for both the gauge station. The best value of R^2 for BP Ghat and Fulertal station is 0.98918 and 0.98762 during training phase. Similarly, during testing phase, the best R^2 value is 0.9638 and 0.96204 for station BP Ghat and Fulertal, respectively. Detailed performance results of PSO-SVM techniques are given below (Table 4).

Linear scale plots of actual vs. predicted floods for BP Ghat and Fulertal locations are shown in Fig. 8. Outcomes demonstrate that predictable peak flood discharge is 5054.36m³/S, 5163.94m³/S, 5328.58m³/S for BPNN, SVM, and PSO-SVM against actual peak 5451.79m³/S for the station BP Ghat. The approximated peak discharge is 5423.92m³/S, 5540.9m³/S, and 5717.55m³/S for BPNN, SVM, and PSO-SVM adjacent to actual peak 5849.15m³/S for Fulaertal division.

Using a boxplot spread of observed and predicted flood water levels have been demonstrated in Fig. 9. Boxplot using respective quartile values signifies degree of spread in predicted dataset. Quartile Q1 (25th percentile) represents the lower end of plot, and quartile Q3 (75th percentile) represents upper end. Median of data is denoted by second quartile Q2 (50th percentile) shown by a

Table 4





(b. i)



Fig. 6. Actual verses predicted graph for (a) BP Ghat, (b) Fulertal gauge station using SVM during (i) training and (ii) testing phases.

Station	Input	NSE		RMSE		\mathbb{R}^2	
		Train	Test	Train	Test	Train	Test
	M!1	0.96735	0.94721	0.05706	0.01805	0.96113	0.94114
BP Ghat	M!2	0.9746	0.95227	0.05599	0.01679	0.96882	0.94688
	M!3	0.97861	0.95936	0.05395	0.01425	0.97021	0.95325
	M!4	0.98408	0.96405	0.05261	0.01308	0.97774	0.95887
	M!5	0.99334	0.96998	0.04973	0.01069	0.98918	0.9638
	M!1	0.96736	0.94435	0.05556	0.01637	0.96002	0.93899
Fulertal	M!2	0.9698	0.94902	0.05539	0.0161	0.96229	0.94365
	M!3	0.97623	0.95537	0.05344	0.01378	0.96997	0.94991
	M!4	0.98164	0.96113	0.05092	0.01162	0.97556	0.95532
	M!5	0.98996	0.96849	0.04962	0.01039	0.98762	0.96204



Fig. 7. Actual versus predicted graph for (a) BP Ghat, (b) Fulertal gauge station using PSO-SVM during (i) training and (ii) testing phases.

vertical line. It is observed that median of observed and estimated flood of hybrid PSO-SVM model was nearly identical for both the stations. However, whiskers prolong to slightly different values for each instance.

Fig. 10 presents the histogram chart with 12 bins for the test and training, in BPNN, SVM, PSO-SVM modelling. The shape of histogram bears a resemblance to that of a normal distribution curve. The deviations of the PSO-SVM model in relation to observed flood values are very low as compared to simple SVM and BPNN models. The histogram plot also resembles the effective predictions by the PSO-SVM model.

Comparison of BPNN, SVM, and PSO-SVM models

Performances of modelling flood flow data are compared utilizing three methods namely BPNN, SVM, and PSO-SVM. Performance measure values are provided in Table 5, which specifies that performance of all models in predicting river discharge is satisfactory and these methods are appropriate to model river discharge data. Best R² values of two stations for BPNN models are 0.93636, 0.91241, and 0.93243, 0.9008 during training and testing phases respectively. Similarly, PSO-SVM models of two stations have R² values of 0.98318, 0.98098 and 0.9638, and 0.96204 for training and testing phases. Consequently, a comparison of R² values of flood forecasting models revealed that results of the PSO-SVM model are better than BPNN and SVM models.

Moreover, SVM performed slightly better than BPNN in monthly flood prediction. This is in accordance with Bafitlhile and Li [3]. Findings of present study revealed that proposed PSO-SVM model performed more accurately and consistently than other applied models considered in different studies [2,26,40,41]. Hence, it can be practically concluded that PSO-SVM showed better prediction

performance than other models since it is a robust hybrid machine learning approach that efficiently predicts natural hazards or other environmental constituents. Fig. 11 clearly indicates that performance of SVM-PSO in terms of N_{SE} value (testing phase) is best compared to BPNN and SVM models.

Discussion

Obtained results from the present study revealed that developed models could analyse data with fewer parameters, lesser calculation time, and good performance. Developed hybrid model can be an alternate time-consuming and cost-effective technique for predicting floods with relatively high efficacy. There are some limitations related to capturing, understanding, and analysing fullscale flood risks by just utilizing remotely sensed data and radar-based images. Moreover, floods cause substantial destruction to critical infrastructures, damaging roads, telecommunication networks, and bridges, making it challenging to reach people trapped in submerged zones. For minimizing effects of such destructions in the future, there is a vital requirement for strengthening early flood warning systems sustained by well-distributed fluvial discharge stations and hydro-meteorological networks. Development of flood prediction models and monitoring are influential methods to mitigate flood damage. The continuous improvement of ML techniques



Fig. 8. Linear scale graph for (i) BP Ghat, (ii) Fulertal gauge station using (a) BPNN, (b) SVM and (c) PSO-SVM.



(ii. b)

Fig. 8. Continued



Fig. 8. Continued



(i)

(ii)

Fig. 9. Box plot representation of actual v. predicted data for (i) BP Ghat, (ii) Fulertal gauge station.

Table 5	
Comparison of results for proposed station.	

Station Techniques		NSE		RMSE		R ²	
		Train	Test	Train	Test	Train	Test
	BPNN	0.94001	0.91844	0.07201	0.03084	0.93636	0.91241
BP Ghat	SVM	0.95984	0.93991	0.06082	0.02134	0.95383	0.93483
	PSO-SVM	0.99334	0.96998	0.04973	0.01069	0.98918	0.9638
	BPNN	0.93809	0.90865	0.07193	0.02325	0.93243	0.9008
Fulertal	SVM	0.95669	0.9389	0.05981	0.02035	0.95088	0.93437
	PSO-SVM	0.98996	0.96849	0.04962	0.01039	0.98762	0.96204

S. Samantaray, A. Sahoo and A. Agnihotri

over the past decades, together with the accessibility of long-term hydrological constraints like temperature, streamflow, and precipitation, has presented a great level of potential for predicting flood magnitude. The performance of ML models certainly has higher accuracy than physical and conventional statistical models. Errors in the datasets may partly explain the study's findings: limitations existing in geospatial floods archive may not be entirely correct because of digitisation errors or may not show the maximum extent of the floods at all times. Naturally, these shortcomings will also impact the performance of methods that are based on mechanistic models. The data limitations specify the requirement for continued reliable monitoring of flood events.



Fig. 10. Histogram chart of proposed prediction models for (i) BP Ghat, (ii) Fulertal gauge station.

17





(ii)

Fig. 10. Continued



(a) BP Ghat



(b) Fulertal

Fig. 11. Comparison of applied models.

Conclusion

Performances of three modelling methods are investigated in present research for providing suggestions regarding suitable methods for monthly flood forecasting. The studied techniques comprise BPNN, SVM, and PSO-SVM. An amalgam model on basis of integration of SVM and PSO is developed in this study for improving forecasting performances. PSO-SVM model was acquired by combining two new techniques SVM and PSO. Combination of SVM with PSO where PSO algorithm was utilized to select optimal parameters for SVM model. This resulted in better performance by the hybrid model than other conventional models applied in this study. Model comparison stipulates that hybrid PSO-SVM model achieved higher performance with RMSE- 0.04962, R² - 0.98918 and N_{SE}- 0.99334 considering both the stations for monthly flood discharge forecasting. Monthly flow discharge data of two gauge sites, i.e. BP ghat and Fulertal from Barak River over a period of 50 years (1969–2018), were utilized.

Moreover, this research concludes that projected PSO-SVM provides a legitimate substitute for applicability in different hydrological modelling scenarios. The study shows that proposed models can be successfully applied for flood forecasting in selected area of study. However, based on performance evaluation measures used in present study, it is observed that robust PSO-SVM model performed better than BPNN, SVM models. This knowledge of hybrid modelling technique can be utilized in different areas like weather, sediment concentration, water table depth and runoff forecasts for checking applicability of proposed model.

Table 1.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

Sandeep Samantaray: Conceptualization, Methodology, Software, Writing – original draft. Abinash Sahoo: Conceptualization, Methodology, Software, Writing – original draft. Ankita Agnihotri: Writing – review & editing.

Data availability

Data will be made available on request.

References

- A. Agnihotri, A. Sahoo, M.K. Diwakar, in: Flood Prediction Using Hybrid ANFIS-ACO Model: A Case Study, In Inventive Computation and Information Technologies, Springer, Singapore, 2022, pp. 169–180.
- [2] S. Anupam, P. Pani, Flood forecasting using a hybrid extreme learning machine-particle swarm optimization algorithm (ELM-PSO) model, Modeling Earth Systems and Environment 6 (1) (2020) 341–347.
- [3] T.M. Bafitlhile, Z. Li, Applicability of ε-support vector machine and artificial neural network for flood forecasting in humid, Semi-Humid and Semi-Arid Basins in China, Water (Basel) 11 (1) (2019) 85.
- [4] M. Buscema, Back propagation neural networks, Subst Use Misuse 33 (1998) 233-270.
- [5] S. Ch, N. Anand, B.K. Panigrahi, S. Mathur, Streamflow forecasting by SVM with quantum behaved particle swarm optimization, Neurocomputing 101 (2013) 18-23.
- [6] F.C. Chen, Back-propagation neural networks for non-linear self-tuning adaptive control, IEEE Control Syst Mag 10 (1990) 44–48.
- [7] N. Chen, C. Xiong, W. Du, C. Wang, X. Lin, Z. Chen, An Improved Genetic Algorithm Coupling a Back-Propagation Neural Network Model (IGA-BPNN) for Water-Level Predictions, Water (Basel) 11 (2019) 1795.
- [8] P.C. Deka, Support vector machine applications in the field of hydrology: a review, Appl Soft Comput 19 (2014) 372–386.
- [9] I. Ebtehaj, H. Bonakdari, S. Shamshirband, Z. Ismail, R. Hashim, New approach to estimate velocity at limit of deposition in storm sewers using vector machine coupled with firefly algorithm, J Pipeline Syst Eng Pract 8 (2017) 4016018.
- [10] T. Ezer, L.P. Atkinson, Accelerated flooding along the US East Coast: on the impact of sea-level rise, tides, storms, the Gulf Stream, and the North Atlantic Oscillations, Earth's Futur 2 (2014) 362–382.
- [11] H.M. Feng, C.Y. Chen, F. Ye, Evolutionary fuzzy particle swarm optimization vector quantization learning scheme in image compression, Expert Syst Appl 32 (2007) 213–222.
- [12] A.T.C. Goh, Back-propagation neural networks for modeling complex systems, Artif Intell Eng 9 (1995) 143-151.
- [13] J. Guo, J. Zhou, H. Qin, Q. Zou, Q. Li, Monthly streamflow forecasting based on improved support vector machine model, Expert Syst Appl 38 (2011) 13073-13081.
- [14] D. Han, L. Chan, N. Zhu, Flood forecasting using support vector machines, J hydroinformatics 9 (2007) 267–276.
- [15] Z. He, X. Wen, H. Liu, J. Du, A comparative study of artificial neural network, adaptive neuro fuzzy inference system and support vector machine for forecasting river flow in the semiarid mountain region, J Hydrol (Amst) 509 (2014) 379–386.
- [16] T.C. Huang, Y.M. Huang, S.C. Cheng, Automatic and interactive e-learning auxiliary material generation utilizing particle swarm optimization, Expert Syst Appl 35 (2008) 2113–2122.
- [17] P. Insom, C. Cao, P. Boonsrimuang, D. Liu, A. Saokarn, P. Yomwan, Y. Xu, A support vector machine-based particle filter method for improved flooding classification, IEEE Geosci Remote Sens Lett 12 (2015) 1943–1947.
- [18] P.P. Jena, C. Chatterjee, G. Pradhan, A. Mishra, Are recent frequent high floods in Mahanadi basin in eastern India due to increase in extreme rainfalls? J Hydrol 517 (2014) 847–862.
- [19] A. Kalra, S. Ahmad, A. Nayak, Increasing streamflow forecast lead time for snowmelt-driven catchment based on large-scale climate patterns, Adv Water Resour 53 (2013) 150–162.
- [20] J. Kennedy, R. Eberhart, Particle swarm optimization, in: Proceedings of ICNN'95-International Conference on Neural Networks IEEE, 1995, pp. 1942–1948.
- [21] Z.Z. Latt, H. Wittenberg, Improving flood forecasting in a developing country: a comparative study of stepwise multiple linear regression and artificial neural network, Water Resour Manag 28 (2014) 2109–2128.
- [22] S. Li, K. Ma, Z. Jin, Y. Zhu, A new flood forecasting model based on SVM and boosting learning algorithms, in: 2016 IEEE Congress on Evolutionary Computation (CEC) IEEE, 2016, pp. 1343–1348.

- [23] F. Liu, F. Xu, S. Yang, A flood forecasting model based on deep learning algorithm via integrating stacked autoencoders with BP neural network, in: 2017 IEEE third International conference on multimedia big data (BigMM) IEEE, 2017, pp. 58–61.
- [24] W. Lu, W. Wang, A.Y. Leung, S.M. Lo, R.K. Yuen, Z. Xu, H. Fan, Air pollutant parameter forecasting using support vector machines, in: Proceedings of the 2002 International Joint Conference on Neural Networks IJCNN'02 (Cat. No. 02CH37290), IEEE, 2002, pp. 630–635.
- [25] P.G. Nieto, E. García-Gonzalo, J.A. Fernández, C.D. Muñiz, Hybrid PSO–SVM-based method for long-term forecasting of turbidity in the Nalón river basin: a case study in Northern Spain, Ecol Eng 73 (2014) 192–200.
- [26] M. Rezaeianzadeh, H. Tabari, A. Yazdi, S. Isik, L. Kalin, Flood flow forecasting using ANN, ANFIS and regression models, Neural Computing and Applications 25 (1) (2014) 25–37.
- [27] D.E. Rumelhart, G.E. Hinton, R.J. Williams, Learning representations by back-propagating errors, Nature 323 (1986) 533-536.
- [28] A. Sahoo, S. Samantaray, S. Bankuru, D.K. Ghose, Prediction of Flood Using Adaptive Neuro-Fuzzy Inference Systems: a Case Study, in: Smart Intelligent Computing and Applications, Springer, Singapore, 2020, pp. 733–739.
- [29] A. Sahoo, S. Samantaray, D.K. Ghose, Stream flow forecasting in mahanadi river basin using artificial neural networks, Procedia Comput Sci 157 (2019) 168–174.
- [30] R. Salat, S. Osowski, Accurate fault location in the power transmission line using support vector machine approach, IEEE Trans power Syst 19 (2004) 979–986.
 [31] S. Samantaray, A. Sahoo, D.K. Ghose, Prediction of Sedimentation in an Arid Watershed Using BPNN and ANFIS, in: ICT Analysis and Applications, Springer,
- [37] S. Samantaray, A. Sahoo, D.K. Ghose, Infiltration Loss Affects Toward Groundwater Fluctuation Through CANFIS in Arid Watershed: a Case Study, in: Smart
 [32] S. Samantaray, A. Sahoo, D.K. Ghose, Infiltration Loss Affects Toward Groundwater Fluctuation Through CANFIS in Arid Watershed: a Case Study, in: Smart
- [32] S. Samantaray, A. Sanoo, D.K. Gnose, innitration Loss Affects Toward Groundwater Fluctuation Infough CANFIS in Arid Watersned: a Case Study, in: Smart Intelligent Computing and Applications, Springer, Singapore, 2020, pp. 781–789. b.
- [33] S. Samantaray, A. Sahoo, D.K. Ghose, Assessment of Groundwater Potential Using Neural Network: a Case Study, in: International Conference on Intelligent Computing and Communication, Springer, 2019, pp. 655–664. a.
- [34] S. Samantaray, A. Sahoo, D.K. Ghose, Assessment of runoff via precipitation using neural networks: watershed modelling for developing environment in arid region, Pertanika J Sci Technol 27 (4) (2019) 2245–2263 b.
- [35] S. Samantaray, A. Sahoo, Appraisal of Runoff Through BPNN, RNN, and RBFN in Tentulikhunti Watershed: a Case Study, in: Frontiers in Intelligent Computing: Theory and Applications, Springer, Singapore, 2020, pp. 258–267. a.
- [36] S. Samantaray, O. Tripathy, A. Sahoo, D.K. Ghose, Rainfall Forecasting Through ANN and SVM in Bolangir Watershed, in: Smart Intelligent Computing and Applications, Springer, IndiaSingapore, 2020, pp. 767–774. c.
- [37] S. Samantaray, A. Sahoo, D.K. Ghose, Assessment of Sediment Load Concentration Using SVM, SVM-FFA and PSR-SVM-FFA in Arid Watershed, India: a Case Study, KSCE J Civ Eng 24 (6) (2020) 1944–1957 d.
- [38] S. Samantaray, A. Sahoo, Estimation of Runoff Through BPNN and SVM in Agalpur Watershed, in: Frontiers in Intelligent Computing: Theory and Applications, Springer, Singapore, 2020, pp. 268–275.
- [39] S. Samantaray, A. Sahoo, Assessment of Sediment Concentration Through RBNN and SVM-FFA in Arid Watershed, India, in: Smart Intelligent Computing and Applications, Springer, Singapore, 2020, pp. 701–709.
- [40] C. Sudheer, R. Maheswaran, B.K. Panigrahi, S. Mathur, A hybrid SVM-PSO model for forecasting monthly streamflow, Neural Comput Appl 24 (2014) 1381–1389.
 [41] K. Tsakiri, A. Marsellos, S. Kapetanakis, in: Artificial Neural Network and Multiple Linear Regression For Flood Prediction in Mohawk River, 10, New York, Water, 2018, p. 1158.
- [42] V. Vapnik, The Nature of Statistical Learning Theory, Springer New York google scholar, New York, 1995.
- [43] J.G. Vlachogiannis, K.Y. Lee, Multi-objective based on parallel vector evaluated particle swarm optimization for optimal steady-state performance of power systems, Expert Syst Appl 36 (2009) 10802–10808.
- [44] L.X. Wang, L.N. Wang, G.F. Li, C. Luan, F.F. Sun, in: Application Research of Support Vector Machine Based On Particle Swarm Optimization in Runoff Forecasting, Applied Mechanics and Materials Trans Tech Publ, 2012, pp. 2303–2307.
- [45] W. Wang, X. Nie, L. Qiu, Support vector machine with particle swarm optimization for reservoir annual inflow forecasting, in: 2010 International Conference on Artificial Intelligence and Computational Intelligence, IEEE, 2010, pp. 184–188. b.
- [46] J.X. Xie, C.T. Cheng, K.W. Chau, Y.Z. Pei, A hybrid adaptive time-delay neural network model for multi-step-ahead prediction of sunspot activity, Int J Environ Pollut 28 (2006) 364–381.
- [47] N. Zaini, M.A. Malek, M. Yusoff, N.H. Mardi, S. Norhisham, Daily river flow forecasting with hybrid support vector machine-particle swarm optimization, in: IOP Conference Series: Earth and Environmental Science, IOP Publishing Ltd., 2018, pp. 1315–1755. a.
- [48] N. Zaini, M.A. Malek, M. Yusoff, S.F.C. Osmi, N.H. Mardi, S. Norhisham, Support vector machine and neural network based model for monthly stream flow forecasting, Int J Eng Technol 7 (2018) 683–688 b.
- [49] L. Zhao, Y. Yang, PSO-based single multiplicative neuron model for time series prediction, Expert Syst Appl 36 (2009) 2805–2812.
- [50] F.N. Robati, S. Iranmanesh, Inflation rate modeling: adaptive neuro-fuzzy inference system approach and particle swarm optimization algorithm (ANFIS-PSO), MethodsX 7 (2020) 101062.
- [51] C. Xiujia, Y. Guanghua, G. Jian, M. Ningning, W. Zihao, Application of WNN-PSO model in drought prediction at crop growth stages: a case study of spring maize in semi-arid regions of northern China, Computers and Electronics in Agriculture 199 (2022) 107155.
- [52] I.S. Macharyulu, D.P. Satapathy, A. Sahoo, S. Samantaray, N.R. Mohanta, A. Ray, Performance Evaluation of MLP and CNN Models for Flood Prediction, in: Intelligent System Design: Proceedings of INDIA 2022, Springer Nature Singapore, Singapore, 2022, pp. 273–281.
- [53] G.F. Lin, Y.C. Chou, M.C. Wu, Typhoon flood forecasting using integrated two-stage support vector machine approach, Journal of Hydrology 486 (2013) 334–342.
 [54] I.F. Kao, Y. Zhou, L.C. Chang, F.J. Chang, Exploring a Long Short-Term Memory based Encoder-Decoder framework for multi-step-ahead flood forecasting,
- Journal of Hydrology 583 (2020) 124631.
 [55] K.E. Adikari, S. Shrestha, D.T. Ratnayake, A. Budhathoki, S. Mohanasundaram, M.N. Dailey, Evaluation of artificial intelligence models for flood and drought forecasting in arid and tropical regions, Environmental Modelling & Software 144 (2021) 105136.
- [56] S. Kamali, K. Asghari, The effect of meteorological and hydrological drought on groundwater storage under climate change scenarios, Water Resources Management (2022) 1–19.
- [57] F.M. Aswad, A.N. Kareem, A.M. Khudhur, B.A. Khalaf, S.A. Mostafa, Tree-based machine learning algorithms in the Internet of Things environment for multivariate flood status prediction, Journal of Intelligent Systems 31 (1) (2022) 1–14.
- [58] D.P. Satapathy, H. Swain, A. Sahoo, S. Samantaray, S.C. Satapathy, Application of a Combined GRNNFOA Model for Monthly Rainfall Forecasting in Northern Odisha, India, in: Intelligent System Design: Proceedings of INDIA 2022, Springer Nature Singapore, Singapore, 2022, pp. 355–364.
- [59] A. Mishra, A. Sahoo, S. Samantaray, D.P. Satapathy, S.C. Satapathy, Monthly Runoff Prediction by Support Vector Machine Based on Whale Optimisation Algorithm, in: Intelligent System Design: Proceedings of INDIA 2022, Springer Nature Singapore, Singapore, 2022, pp. 329–338.
- [60] N.M. Kumar, I. Saikrishnamacharyulu, A. Sahoo, S. Samantaray, M.H. Kumar, A. Naik, S. Sahoo, Improving Streamflow Prediction Using Hybrid BPNN Model Combined with Particle Swarm Optimization, in: Intelligent System Design: Proceedings of INDIA 2022, Springer Nature Singapore, Singapore, 2022, pp. 299–308.