Daily sediment yield prediction using hybrid machine learning approach

---

.

ABSTRACT

In this study, four different soft computing AI techniques were tested for the prediction of sediment yield based on hydro-meteorological variables at Jondhara station, Seonath stream in Rajnandgaon district, India. In order to fulfill this purpose, the models namely, multilayer perceptron (MLP), support vector machine (SVM), multilayer perceptron coupled with genetic algorithm (MLP-GA), and support vector machine coupled with genetic algorithm (SVM-GA) models were employed. To select the optimal input variables, a statistical method such as the Gamma test was considered among several methods. Based on the results of the analysis, all models were evaluated by using the following statistical indices: Coefficient of Correlation (CC), room mean square error (RMSE) and percent bias (PBAIS). Overall, the performance of the studied models indicates that all of them are capable of simulation sediments yield at Jondhara station, Seonath river basin in a satisfactory manner. Comparison of results showed that the MLP-GA with CC = 0.988, RMSE = 0.006 and PBIAS = 0.000 in training period and CC= 0.990, RMSE = 0.007 and PBIAS = 0.000 in testing period for S-6 model and CC = 0.986, RMSE = 0.025 and PBIAS = -0.001 in training period and CC = 0.988, RMSE = 0.029 and PBIAS = -0.001 in testing period for S-13 model were able to yield better results than the other models considered. Furthermore, an SVM model is also observed to have some advantages over MLP models and SVM-GA models since it can represent the output data in a continuous manner by fitting a linear regression function to the output data, which has the advantage of making the model more precise than MLP and SVM-GA models.

Keywords: Stage–discharge–sediment modeling, Gamma test, MLP, SVM, Genetic Algorithm, Sediment Yield, Seonath stream.

1. INTRODUCTION

To manage water resources effectively, it is necessary to understand sediment transport and to predict the quantity of sediment transported/yield. For sustainable agriculture and the environment, soil erosion is a major problem that needs to be addressed. The presence of event-based sediment loads is an important element of our ability to plan and manage protection for river systems, regardless of the contaminants in the rivers, or the ecosystems within them. We need to be able to make accurate predictions of the magnitude and variability of these loads.

The problem has been studied in a number of ways in an attempt to reduce the complexity of the problem. To accomplish this, several practical techniques have been developed that do not require an extensive knowledge of algorithms and theory in order to solve it. As a result, it is widely known that classical models such as multilinear regression (MLR) and sediment rating curves (SRC) could be used for model development of suspended sediments according to classical equations[1, 2]. In McBean and Al-Nassri [3], the authors claimed that linking streamflow records to the sediment concentration (SRC) method could bring about uncertainty, so they proposed developing regression-based models for estimating sediment concentration from streamflow records in order to eliminate this uncertainty. Meanwhile, due to the hysteresis effect, the relationship between sediment concentration and streamflow, as described by Lopes and Ffolliott [4], is very complex when it comes to sediment concentration and streamflow.

A heuristic learning approach has also been utilized for simulating the sediment yield that has been successful in simulating real-life environmental data[5–10].A new study by Kisi et al. [11] used statistical properties such as cross-correlations, auto-correlations and partial auto-correlations of the data series to identify an input vector that is unique to the ANN and is one that represents the sediment estimation process in a basin to the greatest extent possible. Stream discharge and sediment yield data from the stations Quebrada Blanca and Rio Valenciano in the United States were used to evaluate the method.

For time series modelling of suspended sediment concentration (SSC) in rivers[12], Rajaee et al.[13]have used artificial neural networks (ANNs), neuro-fuzzy models (NF), multi-linear regression models (MLRs), and conventional sediment rating curves (SRCs) models. In order to train these models, daily flood discharge data and SSC data from gauging stations along the Little Black River and Salt River in the USA were used. The results obtained by using ANN and NF models are consistent with the observed SSC values; however, they represent better results than those obtained by using MLR and SRC models. Singh et al. [14] used two hybrid machine learning, a combination of GA with Multilayer Perceptron (MLP-GA) and Support Vector Machine (SVM-GA). A comparison of the performances of four machine learning algorithms was performed on different sets of predictors using four different algorithms. The predictor combination containing sand, clay, Field Capacity, and Wilting Point gave the best accuracy out of all the machine learning models. Out of all the ML algorithms, the SVM-GA algorithm performed the best. There was a higher efficiency demonstrated by the SVM-GA algorithm than the MLP-GA algorithm.

Ibrahim et al. [15], stating that Artificial Intelligence (AI) appears to be the most effective alternative solution because of the dramatic increase in the development and development of artificial intelligence models for forecasting and modeling unstable patterns in hydrological fields over the past three decades. Despite this, AI models require optimization in tandem to contribute to achieving the best results. This therefore leads to the desire for hybrid models to be created by combining AI models with optimization techniques to achieve the best results. According to this comprehensive study, machine learning can be categorized into three different types, along with the optimization techniques that are most commonly implemented, and then we will investigate the different types of AI models used for different hydrology fields as well as the most frequently used optimization methods.

On the other hand, when compared to the hybrid multilayer perceptron model, the hybrid multilayer perceptron model performs relatively poorly.Heddam et al. [16]investigated the utility and capabilities of four different artificial intelligence models for the purpose of assessing their effectiveness. According to the results of the AI models, there is a strong and useful basis for predicting pigment concentrations by utilizing simple and well-defined indicators of water quality as variables for forecasting pigment concentrations. As a study was carried out by Özgeret al. [17], both SVM models and MLP models were used to predict drought, and it was also compared whether the empirical decomposition method and the different wavelet networks could be utilized to predict drought more effectively.Sayariet al. [18] evaluated five artificial intelligence and hybrid AI algorithms to predict water infiltration. Comparing hybrid AI models with conventional AI models, he found that hybrid AI models were more efficient than conventional AI models.

Based on review of artificial intelligence's potential applications in a wide range of fields, we concluded that MLP, SVM, and hybrid models that combine these with genetic algorithms can be used to develop novel prediction models using artificial intelligence. In order to develop a new approach for forecasting daily sediment yield, an advanced approach needs to be developed.Hence, our hypothesis was to develop a novel hybrid model for the develop stage-discharge models and discharge-sediment models as well as Sediment yield modeling under changing climate for Seonath basin based on MLP, SVM, MLP-GA, and SVM-GA models. Moreover, the gamma test was also used to develop superior input/output combinations.

2. MATERIAL AND METHODS

2.1 Study Area Description and Collection of Data

Seonath River originates near Panabaras village in the Rajnandgaon District Chhattisgarh India. Basin is located between latitude 20^0 16' N to 22^0 41' N & Longitude 80^0 25' E to $82^0$35' E (Fig. 1.). The basin area of the river up to its confluence with the Mahanadi River covers an area of 30,860 square kilometers. This river traverses a length of 380 kilometers during its journey. Its main tributaries include Kharun, Arpa, Hamp, Agar, and Maniyari Rivers. There is a wide variation in the amount of mean average rainfall received in the basin each year, ranging from 1005 mm to 1255 mm. The data is collected from Seonath sream at Jondhara station in Bilaspur district.

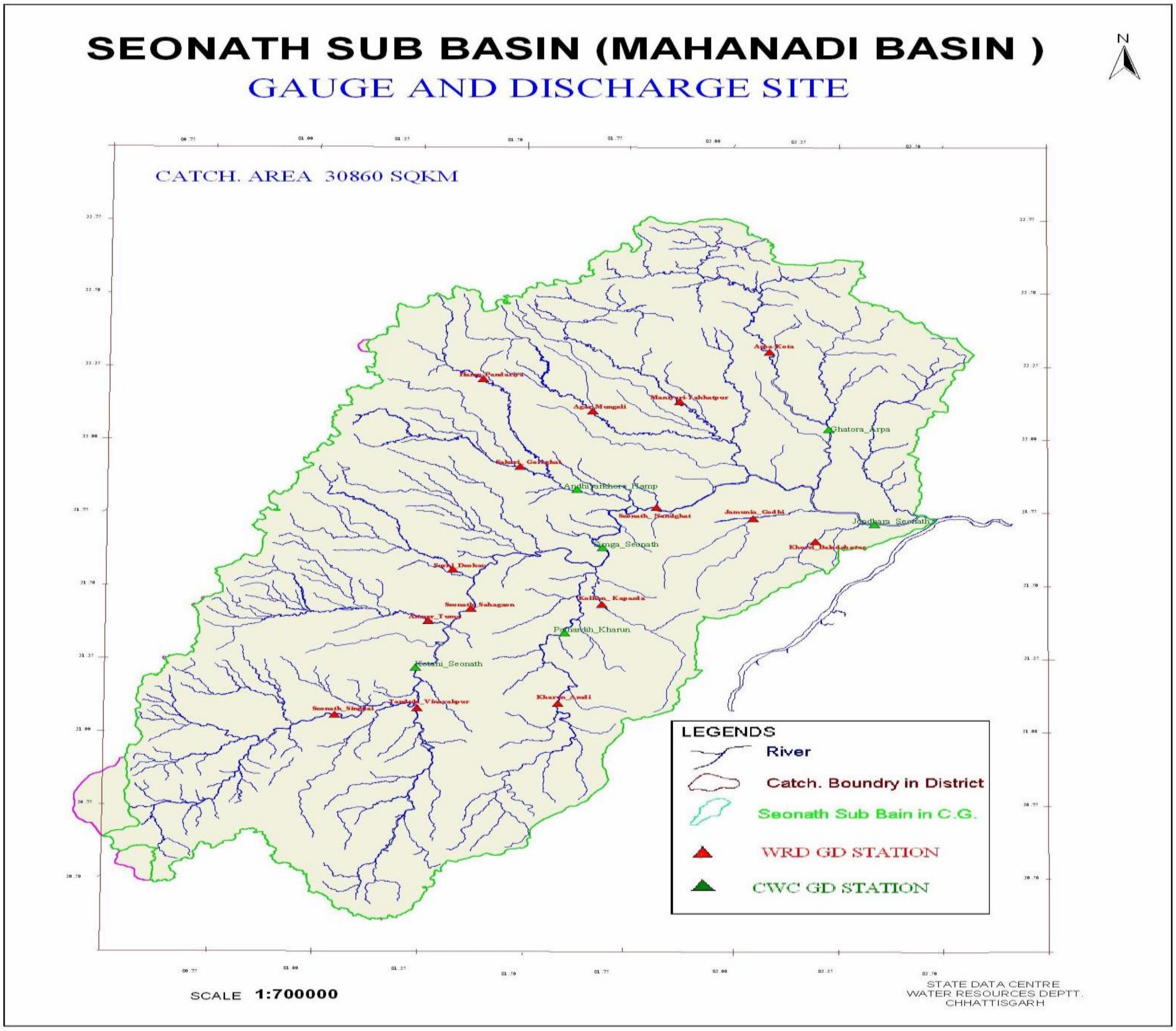


Fig. 1. Study Area

The details of data used in model are given Table 1. The details of the study's methodology are explained in Fig. 2. The measurement were performed by institutions and daily sediment yield were collected by the institutions.

Table 1.List of hydrological gauge stations in Seonath Basin

Station	District	Latitude	Longitude	Stream	Data Available	Typeof Site
Jondhara	Bilaspur	21°42'47"	82°21'30"	Seonath	2000-2020	GDSQ

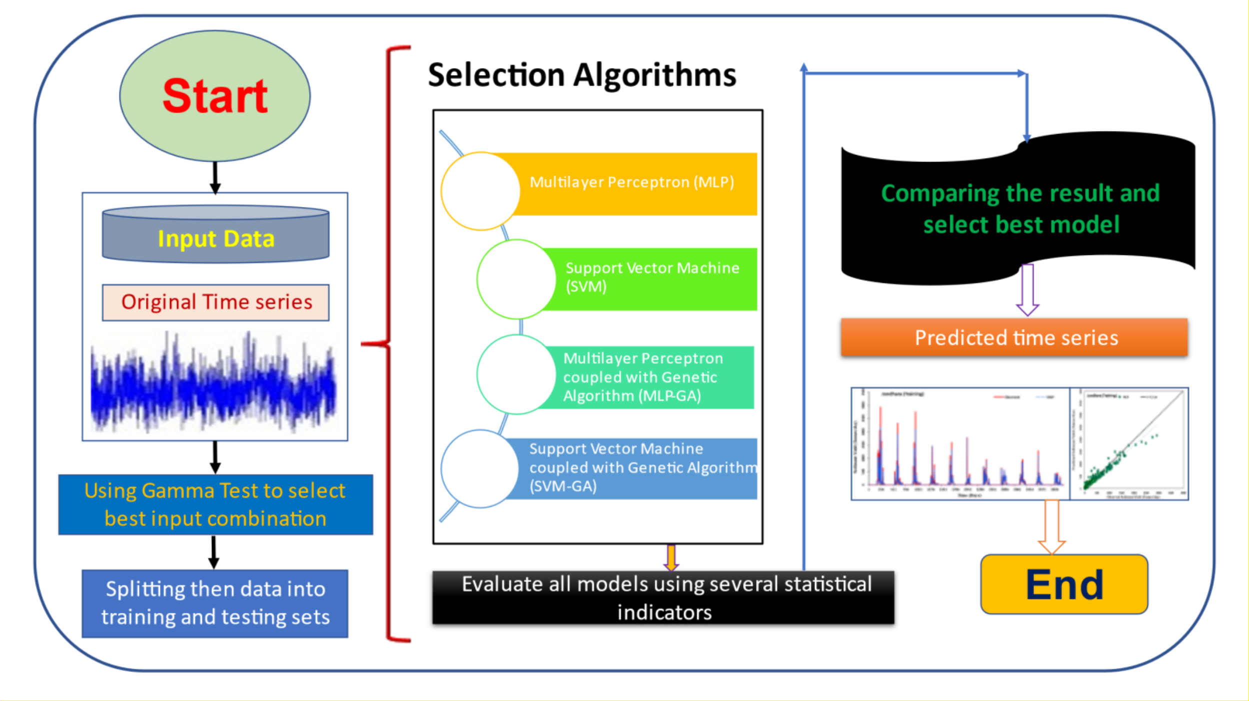


Fig. 2. Flowchart of the methodology

2.2 Artificial intelligence models

2.2.1 Multilayer perceptron neural networks (MLP)

A neural network consists of layers of neurons arranged in parallel [19].Neurons are located in the hidden layers of an ANN, which is composed of an input layer, a hidden layer, and an output layer. During the training phase, neurons are connected by weights to neurons in subsequent layers. There are two activation functions that can be used in order to analyse the features present in the input data. These are sigmoid activation functions and linear activation functions. These two functions are commonly used in the hidden, layer as well as output layer [20, 21]. A multilayer perceptron with a back propagation algorithm is considered to be the most common and popular kind of network, which will be the subject of this study. A number of prediction problems have been successfully tackled with the help of the backpropagation training algorithm [22–24]. Ghorbani et al. [25] elaborated more details about the MLP method.

2.2.2Support vector machine (SVM)

The support-vector machine (SVM, also known as the support-vector network) is a programming model for supervised learning which consists of an associated learning algorithm to analyze the data in order to make classifications and regressions This technique was developed with his colleagues at AT&T Bell Laboratories by Vladimir Vapnik [26, 27]. SVMs can be considered to be one of the most reliable methods for predicting outcomes, since they are based on statistical learning frameworks or on the V-C theory proposed by Cortes and Vapnik [28]. As a result of a set of training examples that have each been marked as belonging to a specific category, the SVM training algorithm builds a model by assigning each new example to one of the two categories based on the mark on the training examples. It is therefore a non-probabilistic binary linear classifier (although methods like Platt scaling exist for implementing SVM in a probabilistic classification environment). This method of machine learning maps the training examples to points in space in such a manner as to increase the width of the gap between the two categories. After mapping the new examples into this space, the predictions are made to determine whether they belong to a specific category based on the side of the gap from which they come [29].

2.2.3 Hybrid artificial intelligence Algorithm Based on GA

GAs are stochastic optimization methods that don't require the use derivatives and are encouraged by natural assortment in the fields of genomics and evolution in biology. In many ways, it is finest to other optimization methods in terms of performance. In terms of continuous and discrete optimization, this method can be used to solve problems in both areas. Unlike the artificial intelligence method, the GA method is less likely to cause the user to get stuck in local least than the AI method. The computational model is based on a population-inspired approach. This is a population genetics-inspired algorithm for learning how to learn new things. Traditionally, it has been used primarily as an optimization function and that has been found to be a valuable global optimization method, particularly for multi-model and non-continuous processes that require global optimization. An overview of the suggested hybrid algorithms can be seen in Figures 3a and 3b which show a schematic representation of them. Figure 3 shows how the programs/algorithms is executed. The proposed model proposes a hybrid AI (MLP and SVM) learning technique using the GA tool to optimise the hyperparameters of the network through integration with artificial intelligence. A chromosome of hyperparameters is encoded on an encoding matrix that allows the GA to tune each hyperparameter on the network. In the final step of the process, as a consequence of the GA procedure, the AI technique is used to train the network. All the AI models were sun in MATLAB v. 2018b.

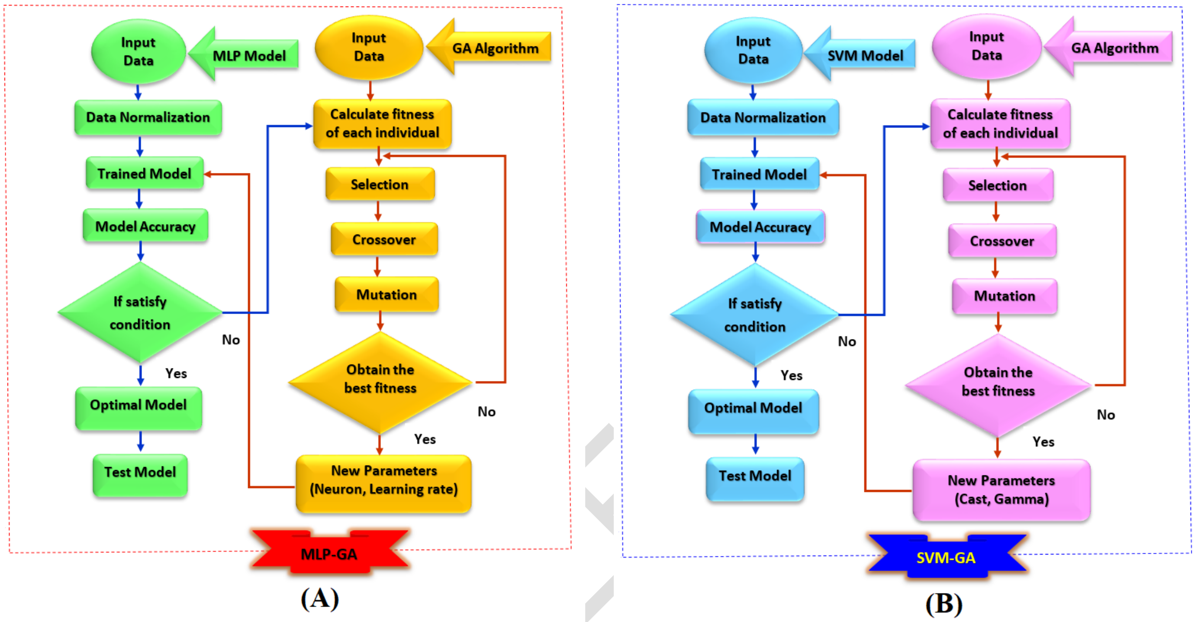


Fig. 3. Flow diagram of hybrid MLP-GA and SVM-GA algorithms

2.3 Model development

To predict the daily sediment yield, it is a complex procedure, which is dependent on various parameters such as rainfall behavior, runoff, soil properties, and vegetative cover, etc. It is also dependent on a particular time lag for the purposes of modelling. Thus, for the purpose of developing sediment yield prediction models, different combinations of runoff and sediment yield were combined.

$$Q_{i,j} = f\left(Q_{i,j}, Q_{i,j-1}, Q_{i,j-2}, \ldots \ldots \ldots Q_{i,j-m}, S_{i,j-1}, S_{i,j-2}, S_{i,j-2}, \ldots \ldots S_{i,j-m}\right) \qquad (1)$$

where, $Q_{i,j}$ is the runoff for j^{th} day of the i^{th} year, S_{ij} is the sediment concentration for j^{th} day of the i^{th} year and m standards for time lag which is taken as three in present study. Hence, different combinations of runoff and sediment concentration need to be considered when developing sediment concentration prediction models. Various input variables were employed to effectively predict the daily sediment yield presented in Table 2 in accordance with the significant correlation between the inputs and output.

Table 2. Combination of input- output variable

Model	Output	Inputs variable
S-1	S_t	Q_{t-1}, Q_{t-2}, Q_{t-3}, Q_{t-4}, Q_{t-5},R_t, R_{t-1}, R_{t-2}, R_{t-3}, R_{t-4}, R_{t-5}
S-2	S_t	Q_{t-1}, Q_{t-2}, Q_{t-3}, Q_{t-4},R_t, R_{t-1}, R_{t-2}, R_{t-3}, R_{t-4}
S-3	S_t	Q_{t-1}, Q_{t-2}, Q_{t-3},R_t, R_{t-1}, R_{t-2}, R_{t-3}
S-4	S_t	Q_{t-1}, Q_{t-2},R_t, R_{t-1}, R_{t-2}, R_{t-3}
S-5	S_t	Q_{t-1}, Q_{t-2},R_t, R_{t-1}, R_{t-2}

S-6	**S_t**	**Q_{t-1}, Q_{t-2}, Q_{t-3},R_t, R_{t-1}**
S-7	S_t	Q_{t-1}, Q_{t-2}, Q_{t-3},R_t
S-8	S_t	Q_{t-1},R_t, R_{t-1}
S-9	S_t	Q_{t-1}, R_t, R_{t-1}, R_{t-2}
S-10	S_t	Q_{t-1}, R_t, R_{t-1}, R_{t-2}, R_{t-3}
S-11	S_t	Q_t, Q_{t-1}, Q_{t-2}, Q_{t-3}, Q_{t-4}, Q_{t-5}, S_{t-1}, S_{t-2}, S_{t-3}, S_{t-4}, S_{t-5}
S-12	S_t	Q_t, Q_{t-1}, Q_{t-2}, Q_{t-3}, S_{t-1}, S_{t-2}, S_{t-3}
S-13	**S_t**	**Q_t, Q_{t-1}, Q_{t-2}, S_{t-1}, S_{t-2}, S_{t-3}**
S-14	S_t	Q_t, Q_{t-1}, Q_{t-2}, S_{t-1}, S_{t-2}
S-15	S_t	Q_t, Q_{t-1}, Q_{t-2}, Q_{t-3}, S_{t-1}

Where S_t, Q_t and R_t = Present day sediment yield (mg/L), discharge (m^3/sec) and rainfall (mm).

2.4 Best input Variable selection: Gamma Test (GT)

In any hydrological environment, there are many complex, dynamic, and non-uniform processes occurring constantly. A Gamma test establishes an impartial and multi-objective way of determining each input parameter's significant potential. Scholars use a tedious and time-consuming trial-and-error method to determine the ideal input combination [30, 31]. Therefore, to resolve this problem, a novel approach Gamma Test, is used to evaluate the ideal input variables in a data set, introduced by Stefansson et al. [32]. It is competent enough to create a trustworthy and smooth model. The gamma test can be used to determine whether a continuous, nonlinear model has the least possible standard error for each set of input-output data by examining its variance. [6, 33–39]. The two-gamma test statistic, gamma value (Γ) and V-ratio, are used to select the number of input variables. The relationship between the inputs *(x)* and output *(y)* variables is determined by Eq. (2):

$$y = Gx + \Gamma \tag{2}$$

Where, *G* and Γ denote the gradient and intercept of the line of regression (*x* = 0), *y* describes the output. Another indicator, i.e., V-ratio (VR):

$$VR = \frac{\Gamma}{\sigma^2(y)} \tag{3}$$

Here, Γ = gamma function, and $\sigma^2(y)$ = output variance. All combinations of inputs could be tested using the Gamma test in order to discover the input combination with the lowest absolute Gamma value. When m scalar inputs are present, there are 2m - 1 possible combinations. We can produce a superior mathematical model if the gamma, standard error, and V-ratio are below zero; when the values of gamma, standard error, and V-ratio are lower, we have a higher chance of model consistency. Input pairings were selected from those that had the lowest gamma, standard error, and V-ratio values [1, 14, 31, 34, 40].

2.5 Selection of Input variable

The selection of the optimal input variables is a crucial stage in modeling for the best result of the chosen models. Various input combinations of sediment yield (mg/L), discharge (m^3/sec) and rainfall (mm) with multi lag were used to determine the best input combination for present day sediment yield prediction. Table 3 lists the values of the two gamma test indicators mask, gamma value along with the standard error for each input pairings for Jondhara station for present day sediment yield. The lower gamma test statistics show that an input combination performs better. Out of 20 feasible combinations, for prediction of present-day sediment yield at Jondhara station, S-6 (Q_{t-1}, Q_{t-2}, Q_{t-3},R_t, R_{t-1}) and S-13 (Q_t, Q_{t-1}, Q_{t-2}, S_{t-1}, S_{t-2}, S_{t-3}) input combination was selected.

Table 3. Selection of Input variable based on Gamma Test

Model	**Input variable**	**Jondhara**	
		Gamma	**SE**
S-1	S_t	0.6041	0.0612
S-2	S_t	0.6276	0.0597
S-3	S_t	0.6256	0.0592
S-4	S_t	0.5378	0.0608

S-5	S_t	0.6048	0.0588
S-6	$\mathbf{S_t}$	**0.5207**	**0.0580**
S-7	S_t	0.5529	0.0603
S-8	S_t	0.6105	0.0600
S-9	S_t	0.5619	0.0595
S-10	S_t	0.5607	0.0598
S-11	S_t	0.5322	0.0550
S-12	S_t	0.5927	0.0590
S-13	$\mathbf{S_t}$	**0.4030**	**0.0481**
S-14	S_t	0.5079	0.0552
S-15	S_t	0.5326	0.0503
S-16	S_t	0.6032	0.0490
S-17	S_t	0.5679	0.0547
S-18	S_t	0.5650	0.0601
S-19	S_t	0.5916	0.0594
S-20	S_t	0.5905	0.0593

2.6 Model performance evaluation

For predicting the daily stream discharge and sediment of the Jondhara, Ghatora and Simga station, the developed models were evaluated qualitatively and quantitatively. A qualitative evaluation of the models was conducted by comparing predicted and observed daily runoff for the years 2000 and 20022 in order to assess their qualitative performance. In order to evaluate the quantitative performance of the model, statistical and hydrological indices like correlation coefficient (CC), root mean square error (RMSE) and percent bias (PBIAS) were estimated. For this study, the acceptable threshold for correlation coefficient (CC) is 0.9 and above. A model between 0.8 and 0.9 is considered to be good, and one below 0.8 is regarded as unsatisfactory. For PBIAS, between 0 to 10 is considered to be very good, 10 to 15 is considered to be good, 15 to 25 is considered to be fair and more than 25 is regarded as unsatisfactory or inadequate. A better and excellence model is one that has the minimum value of root mean square error (RMSE) and percent bias (PBIAS) and higher values of coefficients of correlation (CC) [41–46].

$$\mathrm{CC} = \frac{\sum_{i=1}^{N}\left(S_{t_i}^{\mathrm{Obs}} - \overline{S_{t_1}^{\mathrm{Obs}}}\right)\left(S_{t_i}^{\mathrm{Pre}} - \overline{S_{t_1}^{\mathrm{Pre}}}\right)}{\sqrt{\sum_{i=1}^{N}\left(S_{t_i}^{\mathrm{Obs}} - \overline{S_{t_1}^{\mathrm{Obs}}}\right)^2 \sum_{i=1}^{N}\left(S_{t_i}^{\mathrm{Pre}} - \overline{S_{t_1}^{\mathrm{Pre}}}\right)^2}} \tag{4}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(S_{t_i}^{\mathrm{Obs}} - S_{t_i}^{\mathrm{Pre}}\right)^2} \tag{5}$$

$$\mathrm{RMSE} = \frac{\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(S_{t_i}^{\mathrm{Pre}} - S_{t_i}^{\mathrm{Obs}}\right)} \times 100}{\sum_{i=1}^{N}\left(S_{t_i}^{\mathrm{Obs}}\right)} \tag{6}$$

Where, all parameters are indicated as follows: $S_{t_i}^{\mathrm{Obs}}$ is an observed or actual value of sediment yield, $S_{t_i}^{\mathrm{Pre}}$ is simulated or forecasted value of sediment yield, $\overline{S_{t_1}^{\mathrm{Pre}}}$ and $\overline{S_{t_1}^{\mathrm{Obs}}}$ are the mean value of observed or actual and simulated or forecasted samples of sediment yield, respectively and N is the total number of data points.

3. RESULTS AND DISCUSSION

Daily sediment yield prediction models using different AI technique such as MLP, SVM, MLP-GA and SVM-GA models were developed using above best input combination i.e., S-6 and S-13 model. The daily rainfall and runoff data of 22 years (2000 to 2022) was used for model development. Eleven years data (2002-2012) were used for training and remaining eight years data (2013-2020) were used for testing of developed models.

The well-established Artificial Intelligence predictive models (i.e., the MLP, SVM, MLP-GA and SVM-GA models) are evaluated numerically through several performance statistics that are acceptable norms in hydrological research community. Table 4 tabulates the performance assessment in the training and testing phase of daily sediment yield modeling. The best input combinations for the MLP, SVM, MLP-GA and SVM-GA are designated as (Model S-6, and Model S-13), in their order of the best to the worst performance. For S-6 model, the best fit-of-goodness metrics including CC, RMSE, and PBIAS values, were found 0.983, 0.044 (mg/L) and -0.001 (%) for the MLP with 3-51-1 Architecture, 0.991, 0.117 (mg/L) and -0.003 (%) for SVM model with Cast: 15 & Gamma: 0.20 Architecture, 0.988, 0.006 (mg/L) and 0.000 (%) for MLP-GA model with 3-24-1 Architecture, and 1.000, 0.100 (mg/L), and -0.003 (%) for SVM-GA model with Cast: 19 & Gamma: 0.25 Architecture, respectively for training period which is very reliable and accurately. For testing period, CC, RMSE, and PBIAS values, were found 0.986, 0.051 (mg/L) and -0.002 (%) for the MLP, 0.990, 0.137 (mg/L) and -0.005 (%) for SVM, 0.990, 0.007 (mg/L) and 0.000 (%) for MLP-GA, and 0.999, 0.118 (mg/L) and -0.005 (%) for SVM-GA model, respectively which is very reliable and accurately and closely fitted to observed value.

Table 4. Results of Sediment yield modelling at Jondhara

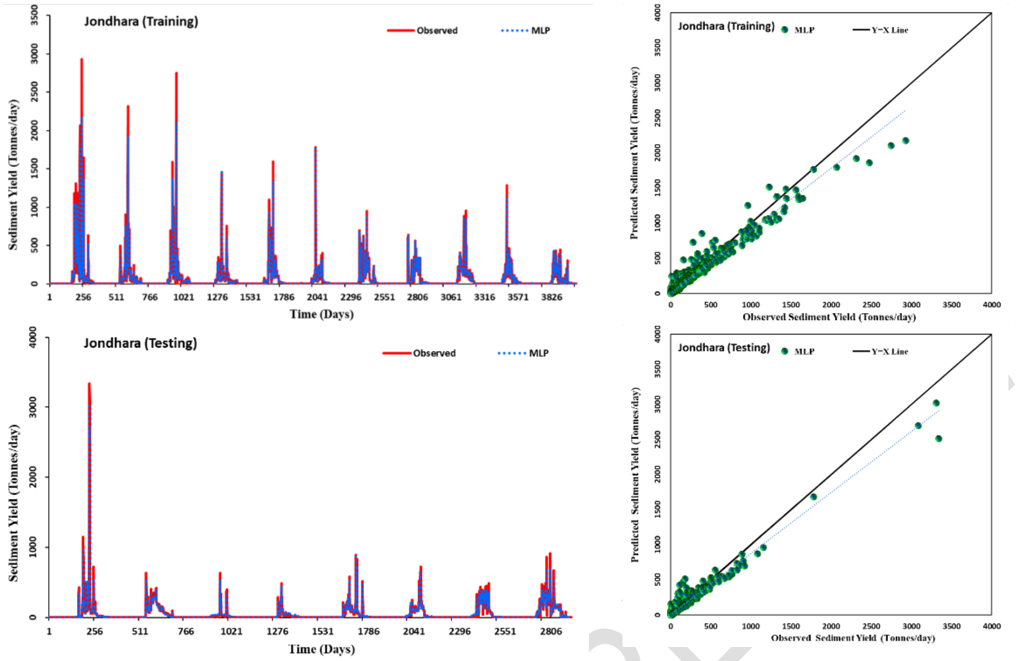
Input	Algorithms	Architecture	Training			Testing		
			CC	**RMSE**	**PBIAS**	**CC**	**RMSE**	**PBIAS**
S-6	MLP	3-51-1	0.983	0.044	-0.001	0.986	0.051	-0.002
	SVM	Cast: 15, Gamma: 0.20	0.991	0.117	-0.003	0.990	0.137	-0.005
	MLP-GA	3-24-1	0.988	0.006	0.000	0.990	0.007	0.000
	SVM-GA	Cast: 19, Gamma: 0.25	1.000	0.100	-0.003	0.999	0.118	-0.005
S-13	MLP	5-29-1	0.935	0.050	-0.001	0.974	0.097	-0.004
	SVM	Cast: 27, Gamma: 0.20	0.989	0.217	-0.006	0.994	0.078	-0.003
	MLP-GA	5-38-1	0.986	0.025	-0.001	0.988	0.029	-0.001
	SVM-GA	Cast: 14, Gamma: 0.22	0.998	0.108	-0.003	0.997	0.127	-0.005

The best input combinations with stream discharge and sediment yield with lag for the MLP, SVM, MLP-GA and SVM-GA are designated as Model S-13, in their order of the best to the worst performance. For S-13 model, the best fit-of-goodness metrics including CC, RMSE, and PBIAS values, were found 0.935, 0.050 (mg/L) and -0.001 (%) for the MLP with 5-29-1 Architecture, 0.989, 0.217 (mg/L) and -0.006 (%) for SVM model with Cast: 27 & Gamma: 0.20 Architecture, 0.986, 0.025 (mg/L) and -0.001 (%) for MLP-GA model with 5-38-1 Architecture, and 0.998, 0.108 (mg/L), and -0.003 (%) for SVM-GA model with Cast: 14 & Gamma: 0.22 Architecture, respectively for training period which is very reliable and accurately. For testing period, CC, RMSE, and PBIAS values, were found 0.974, 0.097 (mg/L) and -0.004 (%) for the MLP, 0.994, 0.078 (mg/L) and -0.003 (%) for SVM, 0.988, 0.029 (mg/L) and -0.001 (%) for MLP-GA, and 0.997, 0.127 (mg/L) and -0.005 (%) for SVM-GA model, respectively which is very reliable and accurately and closely fitted to observed value.

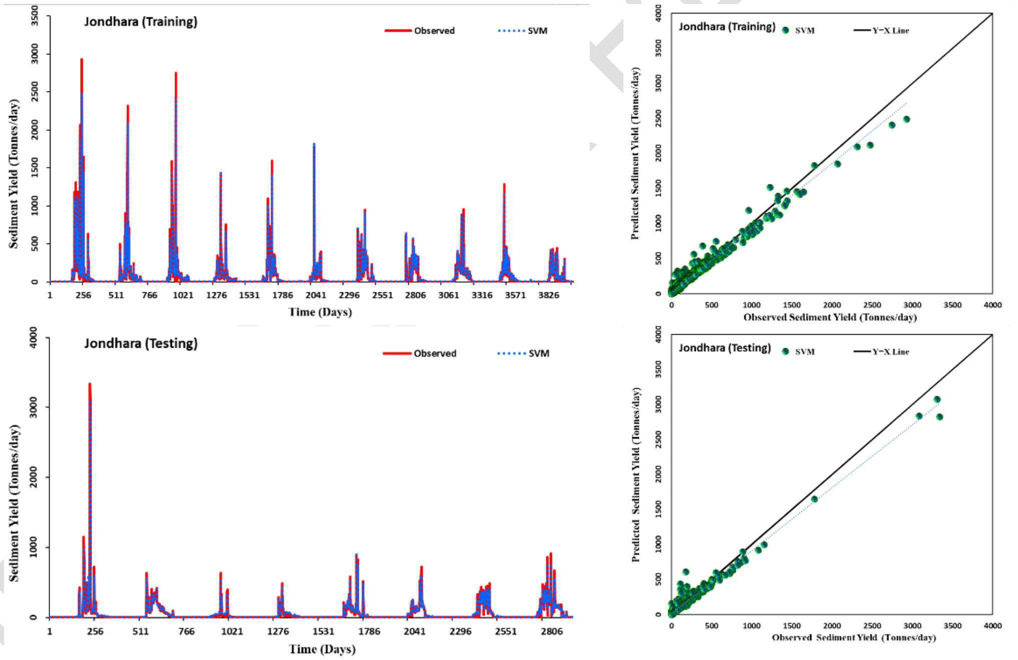
However, in terms of quantitative values of statistical metrics presented in Table 4, the MLP-GA was found to perform better compared to MLP, SVM, and SVM-GA in training and testing phase at Jondhara stations in most of the three statistical metrics for both scenario (S-6 and S-13 model).

The temporal variation of observed and predicted daily sediment yield along with their scatter plots for the MLP, SVM, MLP-GA and SVM-GA models during the training and testing period are plotted in Figure2 for S-6 model and Figure 3 for S-13 model, respectively. As shown in Figure 4 to 5, the estimated and observed values of sediment yield during the training and testing period for Jondhara station during the development of various artificial intelligence models can be seen using the various developed models. Based on the graphs and scatterplots presented in the table, it is evident that the developed models underestimate the sediment yield on a daily basis. For all models, the regression line and the best fit (1:1) line are relatively close to each other. However, for the MLP-GA model, which has less scattered estimates and very closed to reference line than other models, the regression line and best fit (1:1) are more closely related. The regression line is exactly below the best fit (1:1) for S-6 and S-13 input combination of MLP, SVM, MLP-GA and SVM-GA models, which means that all the models under-predict the daily stream discharge values at Jondhara station.

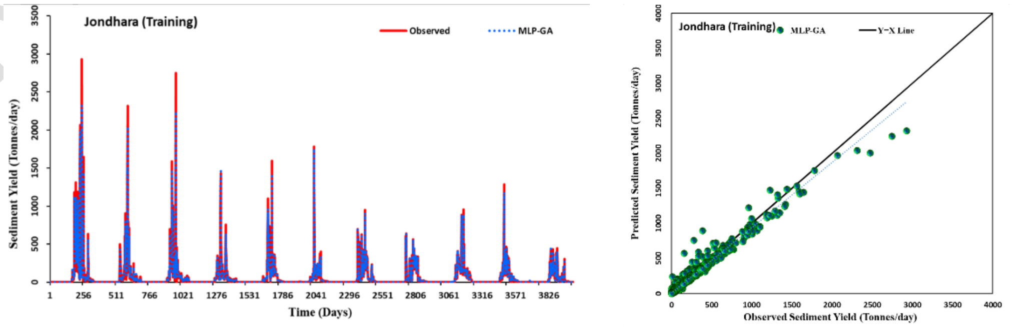
a)

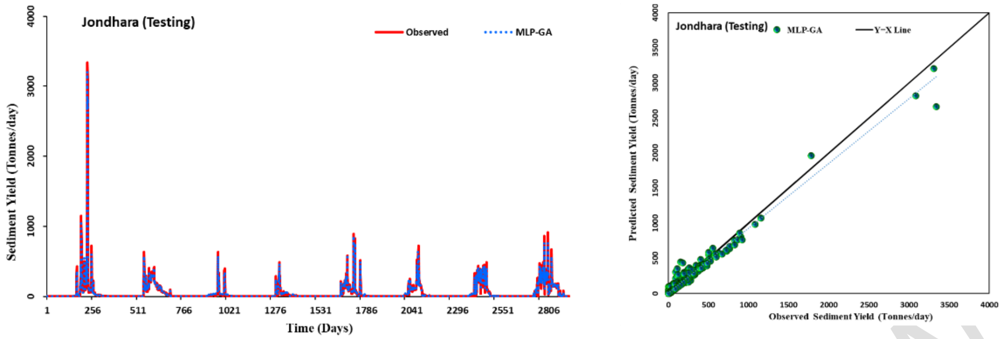

Jondhara (Training)
Observed
MLP
Sediment Yield (Tonnes/day)
Time (Days)
Y=X Line
Predicted Sediment Yield (Tonnes/day)
Observed Sediment Yield (Tonnes/day)
Jondhara (Testing)
Observed
MLP
Sediment Yield (Tonnes/day)
Time (Days)
Y=X Line
Predicted Sediment Yield (Tonnes/day)
Observed Sediment Yield (Tonnes/day)

b)


Jondhara (Training)
Observed
SVM
Sediment Yield (Tonnes/day)
Time (Days)
Y=X Line
Predicted Sediment Yield (Tonnes/day)
Observed Sediment Yield (Tonnes/day)
Jondhara (Testing)
Observed
SVM
Sediment Yield (Tonnes/day)
Time (Days)
Y=X Line
Predicted Sediment Yield (Tonnes/day)
Observed Sediment Yield (Tonnes/day)

c)


Jondhara (Training)
Observed
MLP-GA
Sediment Yield (Tonnes/day)
Time (Days)
Y=X Line
Predicted Sediment Yield (Tonnes/day)
Observed Sediment Yield (Tonnes/day)



d)

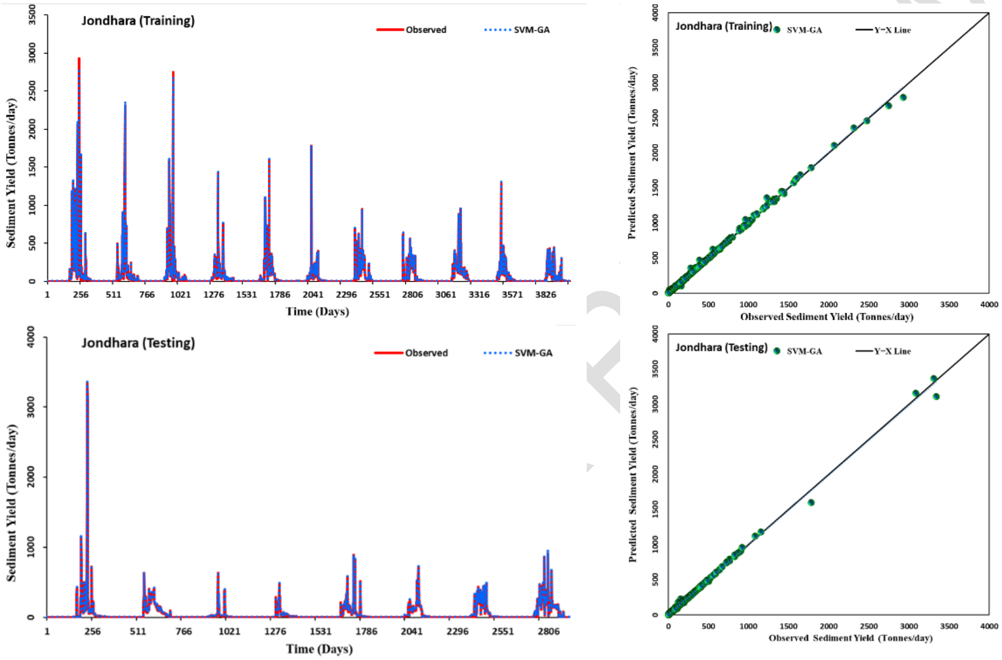
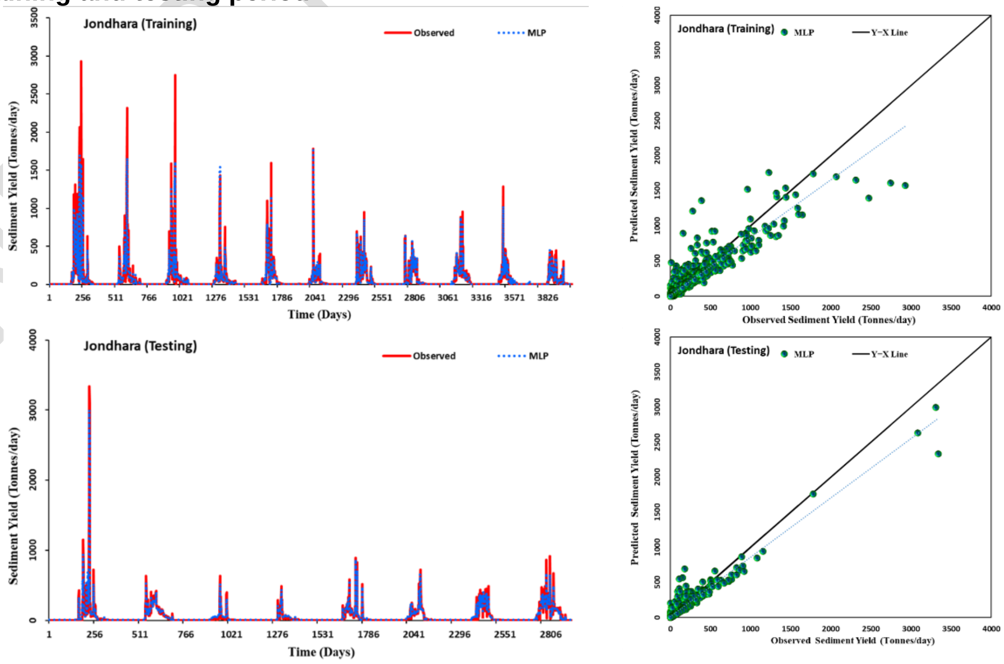
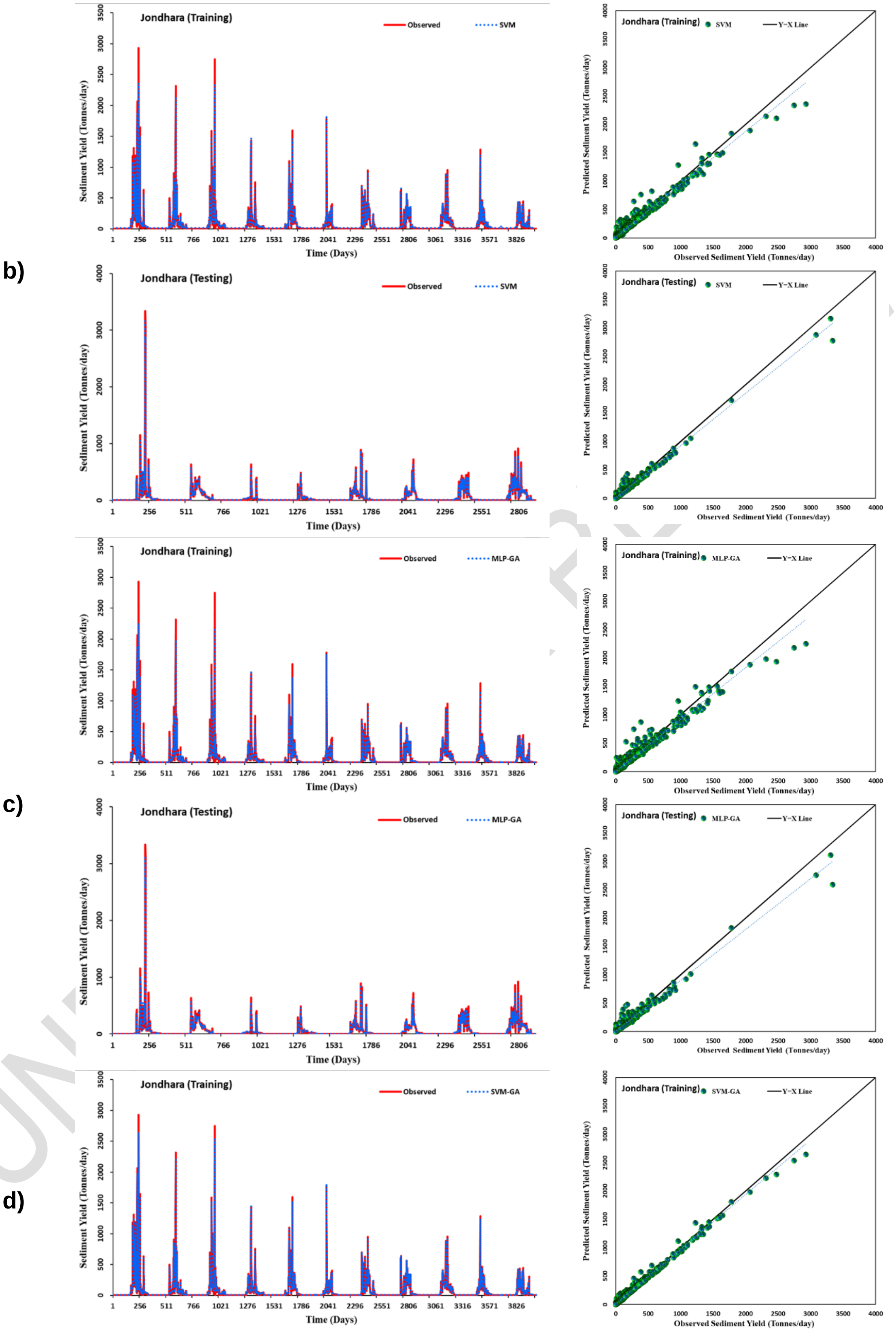


Fig. 4. Line and scatter diagram of (a) MLP, (b) SVM, (c) MLP-GA and (d) SVM-GA models model (S-6) during training and testing period

a)




b)
c)
d)
Jondhara (Training)
Jondhara (Testing)
Observed
SVM
MLP-GA
SVM-GA
Y=X Line
Sediment Yield (Tonnes/day)
Time (Days)
1
256
511
766
1021
1276
1531
1786
2041
2296
2551
2806
3061
3316
3571
3826
0
500
1000
1500
2000
2500
3000
3500
4000
Predicted Sediment Yield (Tonnes/day)
Observed Sediment Yield (Tonnes/day)

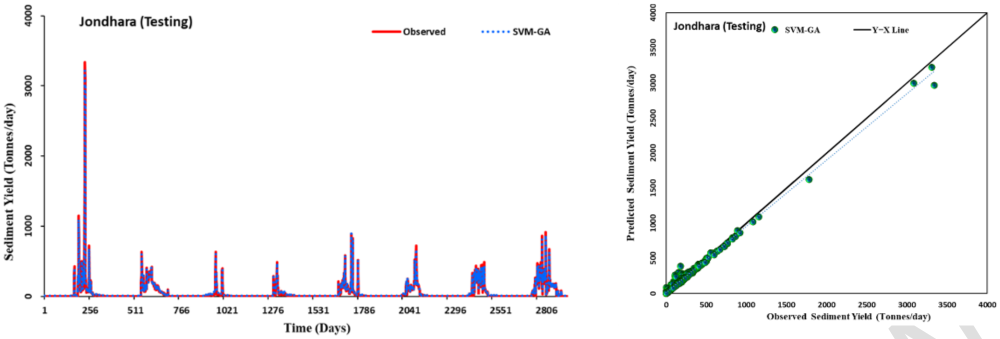


Fig. 5. Line and scatter diagram of (a) MLP, (b) SVM, (c) MLP-GA and (d) SVM-GA models model (S-13) during training and testing period

4. CONCLUSION

In this study, four conceptual models viz. MLP, SVM, MLP-GA and SVM-GA model have been examined for modelling sediment yield of Jondhara station at Seonath River basin in Chhattisgarh-India. As rain-runoff-sediment processes have a highly non-linear nature, these models may be a good candidate for evaluation because of their highly non-linear nature. The daily sediment yield information from year 2000 to 2020 were used to train and test the all models. The possible 20 input variables to all the models were comprised of S-6 (Q_{t-1}, Q_{t-2}, Q_{t-3},R_t, R_{t-1}) and S-13 (Q_t, Q_{t-1}, Q_{t-2}, S_{t-1}, S_{t-2}, S_{t-3}) combination of total daily rainfall, daily stream discharge and daily sediment yield, lagged for one two and three days. In the selection of the best input combination, the gamma test (GT) was used to test the statistical parameters that inform the choice of the best input combination. Based on the statistical evaluation measures such as CC, RMSE and PBIAS, the performance of each model structure was assessed. According to the findings of the present study, the following conclusions can be drawn:

1. Based on the performance indices, among artificial neural networks MLP-GA model (3-24-1) for S-6 model and MLP-GA model (5-38-1) for S-13 model outperformed than MLP, SVM, SVM-GA models in daily sediment yield prediction for Jondhara river.
2. The results showed that an increase (decrease) in rainfall, led to an increase (decrease) in runoff and sediment yield.
3. Results also suggest that SVM model can be a good alternative for sediment estimation on field where quick results are expected in least computational time.

REFERENCES

1. Shukla, R., Kumar, P., Vishwakarma, D.K., Ali, R., Kumar, R., Kuriqi, A.: Modeling of stage-discharge using back propagation ANN-, ANFIS-, and WANN-based computing techniques. Theor. Appl. Climatol. (2021). https://doi.org/10.1007/s00704-021-03863-y
2. Kisi, O.: Suspended sediment estimation using neuro-fuzzy and neural network approaches/Estimation des matières en suspension par des approches neurofloues et à base de réseau de neurones. Hydrol. Sci. J. 50, null-696 (2005). https://doi.org/10.1623/hysj.2005.50.4.683
3. McBean, E.A., Al‐ Nassri, S.: Uncertainty in Suspended Sediment Transport Curves. J. Hydraul. Eng. 114, 63–74 (1988). https://doi.org/10.1061/(ASCE)0733-9429(1988)114:1(63)
4. Lopes, V.L., Ffolliott, P.F.: Sediment rating curves for a clear-cut ponderosa pine watershed in northern Arizona. J. Am. Water Resour. Assoc. 29, 369–382 (1993). https://doi.org/10.1111/j.1752-1688.1993.tb03214.x
5. Malik, A., Kumar, A., Kisi, O., Shiri, J.: Evaluating the performance of four different heuristic approaches with Gamma test for daily suspended sediment concentration modeling. Environ. Sci. Pollut. Res. 26, 22670–22687 (2019). https://doi.org/10.1007/s11356-019-05553-9
6. Singh, V.K., Kumar, D., Kashyap, P.S., Kisi, O.: Simulation of suspended sediment based on gamma test, heuristic, and regression-based techniques. Environ. Earth Sci. 77, 708 (2018).

https://doi.org/10.1007/s12665-018-7892-6

7. Adnan, R.M., Parmar, K.S., Heddam, S., Shahid, S., Kisi, O.: Suspended Sediment Modeling Using a Heuristic Regression Method Hybridized with Kmeans Clustering. Sustainability. 13, 4648 (2021). https://doi.org/10.3390/su13094648
8. Shamaei, E., Kaedi, M.: Suspended sediment concentration estimation by stacking the genetic programming and neuro-fuzzy predictions. Appl. Soft Comput. 45, 187–196 (2016). https://doi.org/10.1016/j.asoc.2016.03.009
9. Samantaray, S., Sahoo, A.: Prediction of suspended sediment concentration using hybrid SVM-WOA approaches. Geocarto Int. 37, 5609–5635 (2022). https://doi.org/10.1080/10106049.2021.1920638
10. Sahraei, S., Alizadeh, M.R., Talebbeydokhti, N., Dehghani, M.: Bed material load estimation in channels using machine learning and meta-heuristic methods. J. Hydroinformatics. 20, 100–116 (2018). https://doi.org/10.2166/hydro.2017.129
11. Kisi, Ö.: Constructing neural network sediment estimation models using a data-driven algorithm. Math. Comput. Simul. 79, 94–103 (2008). https://doi.org/10.1016/j.matcom.2007.10.005
12. Vafakhah, M.: Comparison of cokriging and adaptive neuro-fuzzy inference system models for suspended sediment load forecasting. Arab. J. Geosci. 6, 3003–3018 (2013). https://doi.org/10.1007/s12517-012-0550-5
13. Rajaee, T., Mirbagheri, S.A., Zounemat-Kermani, M., Nourani, V.: Daily suspended sediment concentration simulation using ANN and neuro-fuzzy models. Sci. Total Environ. 407, 4916–4927 (2009). https://doi.org/https://doi.org/10.1016/j.scitotenv.2009.05.016
14. Singh, V.K., Panda, K.C., Sagar, A., Al-Ansari, N., Duan, H.-F., Paramaguru, P.K., Vishwakarma, D.K., Kumar, A., Kumar, D., Kashyap, P.S., Singh, R.M., Elbeltagi, A.: Novel Genetic Algorithm (GA) based hybrid machine learning-pedotransfer Function (ML-PTF) for prediction of spatial pattern of saturated hydraulic conductivity. Eng. Appl. Comput. Fluid Mech. 16, 1082–1099 (2022). https://doi.org/10.1080/19942060.2022.2071994
15. Ibrahim, K.S.M.H., Huang, Y.F., Ahmed, A.N., Koo, C.H., El-Shafie, A.: A review of the hybrid artificial intelligence and optimization modelling of hydrological streamflow forecasting. Alexandria Eng. J. 61, 279–303 (2022). https://doi.org/10.1016/j.aej.2021.04.100
16. Heddam, S., Sanikhani, H., Kisi, O.: Application of artificial intelligence to estimate phycocyanin pigment concentration using water quality data: a comparative study. Appl. Water Sci. 9, 164 (2019). https://doi.org/10.1007/s13201-019-1044-3
17. Özger, M., Başakın, E.E., Ekmekcioğlu, Ö., Hacısüleyman, V.: Comparison of wavelet and empirical mode decomposition hybrid models in drought prediction. Comput. Electron. Agric. 179, 105851 (2020). https://doi.org/https://doi.org/10.1016/j.compag.2020.105851
18. Sayari, S., Mahdavi-Meymand, A., Zounemat-Kermani, M.: Irrigation water infiltration modeling using machine learning. Comput. Electron. Agric. 180, 105921 (2021). https://doi.org/https://doi.org/10.1016/j.compag.2020.105921
19. Kawamoto, A.H., McClelland, J.L., Rumelhart, D.E.: Explorations in Parallel Distributed Processing: A Handbook of Models, Programs, and Exercises. Am. J. Psychol. 102, 435 (1989). https://doi.org/10.2307/1423065
20. Deo, R.C., Samui, P., Kim, D.: Estimation of monthly evaporative loss using relevance vector machine, extreme learning machine and multivariate adaptive regression spline models. Stoch. Environ. Res. Risk Assess. 30, 1769–1784 (2016). https://doi.org/10.1007/s00477-015-1153-y
21. Lima, A.R., Cannon, A.J., Hsieh, W.W.: Forecasting daily streamflow using online sequential extreme learning machines. J. Hydrol. 537, 431–443 (2016). https://doi.org/10.1016/j.jhydrol.2016.03.017
22. Mohanty, S., Jha, M.K., Raul, S.K., Panda, R.K., Sudheer, K.P.: Using Artificial Neural Network Approach for Simultaneous Forecasting of Weekly Groundwater Levels at Multiple Sites. Water Resour. Manag. 29, 5521–5532 (2015). https://doi.org/10.1007/s11269-015-1132-6
23. Günay, M.E.: Forecasting annual gross electricity demand by artificial neural networks using predicted values of socio-economic indicators and climatic conditions: Case of Turkey. Energy Policy. 90, 92–101 (2016). https://doi.org/10.1016/j.enpol.2015.12.019
24. Kawamoto, A.H.: Application of artificial neural network for predicting hourly indoor air temperature and relative humidity in modern building in humid region. Am. J. Psychol. 102, 435–438 (2023). https://doi.org/10.2307/1423065

25. Ghorbani, M.A., Khatibi, R., Hosseini, B., Bilgili, M.: Relative importance of parameters affecting wind speed prediction using artificial neural networks. Theor. Appl. Climatol. 114, 107–114 (2013). https://doi.org/10.1007/s00704-012-0821-9
26. Vapnik, V.: Statistical learning theory. John Wiley & Sons, Inc. Oxford, UK (1998)
27. Vapnik, V.: The nature of statistical learning theory. Springer science & business media (1999)
28. Cortes, C., Vapnik, V.: Support-vector networks. Mach. Learn. 20, 273–297 (1995). https://doi.org/10.1007/BF00994018
29. Ma, Y., Liu, H., Zhai, G., Huo, Z.: Financial Risk Early Warning Based on Wireless Network Communication and the Optimal Fuzzy SVM Artificial Intelligence Model. Wirel. Commun. Mob. Comput. 2021, 1–20 (2021). https://doi.org/10.1155/2021/7819011
30. Markuna, S., Kumar, P., Ali, R., Vishwkarma, D.K., Kushwaha, K.S., Kumar, R., Singh, V.K., Chaudhary, S., Kuriqi, A.: Application of Innovative Machine Learning Techniques for Long-Term Rainfall Prediction. Pure Appl. Geophys. (2023). https://doi.org/10.1007/s00024-022-03189-4
31. Singh, A.K., Kumar, P., Ali, R., Al-Ansari, N., Vishwakarma, D.K., Kushwaha, K.S., Panda, K.C., Sagar, A., Mirzania, E., Elbeltagi, A., Kuriqi, A., Heddam, S.: An Integrated Statistical-Machine Learning Approach for Runoff Prediction. Sustainability. 14, 8209 (2022). https://doi.org/10.3390/su14138209
32. Stefánsson, A., Končar, N., Jones, A.J.: A note on the Gamma test. Neural Comput. Appl. 5, 131–133 (1997). https://doi.org/10.1007/BF01413858
33. Seifi, A., Riahi, H.: Estimating daily reference evapotranspiration using hybrid gamma test-least square support vector machine, gamma test-ANN, and gamma test-ANFIS models in an arid area of Iran. J. Water Clim. Chang. 11, 217–240 (2020). https://doi.org/10.2166/wcc.2018.003
34. Kumar, A., Singh, V.K., Saran, B., Al-Ansari, N., Singh, V.P., Adhikari, S., Joshi, A., Singh, N.K., Vishwakarma, D.K.: Development of Novel Hybrid Models for Prediction of Drought- and Stress-Tolerance Indices in Teosinte Introgressed Maize Lines Using Artificial Intelligence Techniques. Sustainability. 14, 2287 (2022). https://doi.org/10.3390/su14042287
35. Singh, V.K., Singh, B.P., Kumar, A., Vivekanand: A Comparative Study of Artificial Intelligence and Conventional Techniques for Rainfall-Runoff Modeling. Int. J. Agric. Eng. Vol 10, No 2. 10, 441–449 (2017). https://doi.org/DOI: 10.15740/HAS/IJAE/10.2/441-449
36. Malik, A., Kumar, A., Kim, S., Kashani, M.H., Karimi, V., Sharafati, A., Ghorbani, M.A., Al-Ansari, N., Salih, S.Q., Yaseen, Z.M., Chau, K.W.: Modeling monthly pan evaporation process over the Indian central Himalayas: application of multiple learning artificial intelligence model. Eng. Appl. Comput. Fluid Mech. 14, 323–338 (2020). https://doi.org/10.1080/19942060.2020.1715845
37. Kumar, M., Kumar, P., Kumar, A., Elbeltagi, A., Kuriqi, A.: Modeling stage–discharge–sediment using support vector machine and artificial neural network coupled with wavelet transform. Appl. Water Sci. 12, 87 (2022). https://doi.org/10.1007/s13201-022-01621-7
38. Bajirao, T.S., Kumar, P., Kumar, M., Elbeltagi, A., Kuriqi, A.: Potential of hybrid wavelet-coupled data-driven-based algorithms for daily runoff prediction in complex river basins. Theor. Appl. Climatol. 145, 1207–1231 (2021). https://doi.org/10.1007/s00704-021-03681-2
39. Kuriqi, A., Ali, R., Pham, Q.B., Montenegro Gambini, J., Gupta, V., Malik, A., Linh, N.T.T., Joshi, Y., Anh, D.T., Nam, V.T., Dong, X.: Seasonality shift and streamflow flow variability trends in central India. Acta Geophys. 68, 1461–1475 (2020). https://doi.org/10.1007/s11600-020-00475-4
40. Piri, J., Amin, S., Moghaddamnia, A., Keshavarz, A., Han, D., Remesan, R.: Daily Pan Evaporation Modeling in a Hot and Dry Climate. J. Hydrol. Eng. 14, 803–811 (2009). https://doi.org/10.1061/(ASCE)HE.1943-5584.0000056
41. Vishwakarma, D.K., Kumar, R., Pandey, K., Singh, V., Kushwaha, K.S.: Modeling of Rainfall and Ground Water Fluctuation of Gonda District Uttar Pradesh, India. Int. J. Curr. Microbiol. Appl. Sci. 7, 2613–2618 (2018). https://doi.org/10.20546/ijcmas.2018.705.302
42. Vishwakarma, D.K., Pandey, K., Kaur, A., Kushwaha, N.L., Kumar, R., Ali, R., Elbeltagi, A., Kuriqi, A.: Methods to estimate evapotranspiration in humid and subtropical climate conditions. Agric. Water Manag. 261, 107378 (2022). https://doi.org/10.1016/j.agwat.2021.107378
43. Sagar, A., Hasan, M., Singh, D.K., Al-Ansari, N., Chakraborty, D., Singh, M.C., Iquebal, M.A., Kumar, A., Malkani, P., Vishwakarma, D.K., Elbeltagi, A.: Development of Smart Weighing Lysimeter for Measuring Evapotranspiration and Developing Crop Coefficient for Greenhouse Chrysanthemum. Sensors. 22, 6239 (2022). https://doi.org/10.3390/s22166239
44. Kumar, R., Manzoor, S., Vishwakarma, D.K., Al-Ansari, N., Kushwaha, N.L., Elbeltagi, A.,

Sushanth, K., Prasad, V., Kuriqi, A.: Assessment of Climate Change Impact on Snowmelt Runoff in Himalayan Region. Sustainability. 14, 1–23 (2022). https://doi.org/10.3390/su14031150

45. Vishwakarma, D.K., Kumar, R., Kumar, A., Kushwaha, N.L., Kushwaha, K.S., Elbeltagi, A.: Evaluation and development of empirical models for wetted soil fronts under drip irrigation in high-density apple crop from a point source. Irrig. Sci. (2022). https://doi.org/10.1007/s00271-022-00826-7

46. Kushwaha, N.L., Rajput, J., Elbeltagi, A., Elnaggar, A.Y., Sena, D.R., Vishwakarma, D.K., Mani, I., Hussein, E.E.: Data Intelligence Model and Meta-Heuristic Algorithms-Based Pan Evaporation Modelling in Two Different Agro-Climatic Zones: A Case Study from Northern India. Atmosphere (Basel). 12, 1654 (2021). https://doi.org/10.3390/atmos12121654