Journal of
Hydroinformatics

Check for updates

© 2021 The Authors Journal of Hydroinformatics Vol 23 No 4, 849 doi: 10.2166/hydro.2021.011

Application of gradient tree boosting regressor for the prediction of scour depth around bridge piers

B. M. Sreedhara[a], Amit Prakash Patil[b,*], Jagalingam Pushparaj[c], Geetha Kuntoji[d] and Sujay Raghavendra Naganna[a]

[a] Department of Civil Engineering, Siddaganga Institute of Technology, Tumakuru 572103, Karnataka, India
[b] Department of Civil Engineering, Annasaheb Dange College of Engineering and Technology, Ashta 416301, Maharashtra, India
[c] School of Computer Science and Engineering, Vellore Institute of Technology, Vellore 632014, Tamil Nadu, India
[d] Department of Civil Engineering, BMS College of Engineering, Bengaluru 560019, Karnataka, India
*Corresponding author. E-mail: amitpatil.nitk@gmail.com

ABSTRACT

Scour around bridge piers is a complex phenomenon and it is essential to assess or predict the scour hazard around bridge piers in tandem with completely understanding its mechanism. To date, there is no exact method for the estimation of scour depth. Nowadays, machine learning techniques are being recognized as effective tools for the prediction of scour depth using experimental data. In the present study, gradient tree boosting (GTB) technique was used for the prediction of scour depth around various pier shapes under different streambed conditions. Sediment size, sediment quantity, velocity, and flow time were used as input parameters to predict the scour depth under clear-water and live-bed scour conditions. The scour depth was predicted for different pier shapes such as, circular, rectangular, round-nosed and sharp-nosed shaped. The GTB model predicted scour depth values were compared with that of the group method of data handling (GMDH) technique. The performance of GTB and GMDH models were then evaluated based on statistical indices such as RRMSE, NNSE, WI, MNE, SI, and KGE. The study concludes that the GTB model performance was relatively superior to that of GMDH in the prediction of scour depth around different pier shapes.

Key words: clear-water scour, GMDH, gradient tree boosting, live-bed scour, scour prediction

HIGHLIGHTS

- An attempt has been made to predict the bridge scour for various pier shapes under different streambed conditions.
- Ensemble Machine Learning methods like Gradient Tree Boosting (GTB) and Group Method of Data Handling (GMDH) were used for prediction of bridge scour.
- The study concludes that GTB method can be successfully used for predicting bridge scour.

INTRODUCTION

In recent years, many cases of bridge failures have been attributed to hydraulic failure and scour around the piers (Wang *et al.* 2017). Generally, obstruction of flow causes scouring at the pier, abutment, and erosion control devices. Scouring generally involves the removal of sediment from the riverbed due to the erosive action of flowing water. The obstruction caused by the piers of the bridge to the flow of water creates vortexes and removes the material from the riverbed. This type of scouring, which occurs near the piers and abutments, is usually termed as local scour. Local scour near the pier is a dynamic phenomenon and depends on many factors like flow depth, approach flow velocity, attack angle of approach flow, size and gradation of bed material, shape and size of the pier, etc.

A system of vortices (Figure 1) is considered as the basic mechanism behind local scouring around bridge piers (Farooq & Ghumman 2019). The river water flowing at a particular velocity when it is obstructed by a pier comes to complete rest and creates a downward pressure gradient that forces the flow to move downwards like a jet of water. This downward impinging of water on the riverbed is the primary reason behind scouring. The temporal variation of scours at a bridge pier depends on many factors that are mentioned above. However, the accurate estimation of equilibrium depths of local scour is of immense importance for a bridge from a hydraulic design point of view. Equilibrium

This is an Open Access article distributed under the terms of the Creative Commons Attribution Licence (CC BY-NC-ND 4.0), which permits copying and redistribution for non-commercial purposes with no derivatives, provided the original work is properly cited (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Downloaded from http://iwa.silverchair.com/jh/article-pdf/23/4/849/910421/jh0230849.pdf
by guest

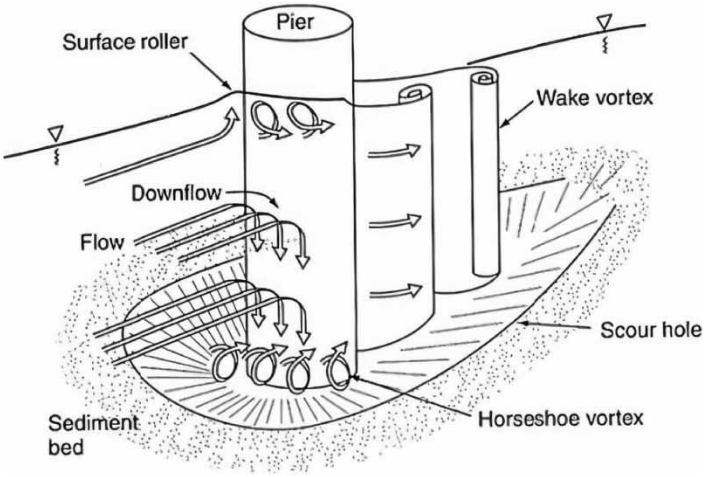

Figure 1 | Scouring mechanism near bridge pier (source: Sreedhara *et al.* 2019).

scour occurs when equilibrium is achieved between the erosive capabilities of the flow and the resistance to the movement of the bed material. The process of bridge scour is very complicated and makes the process of predicting bridge scour quite challenging (Johnson *et al.* 2015).

Most of the studies that were undertaken earlier used scaled laboratory experiments with dimensional analysis for studying the phenomenon of scouring. However, the inherent complexity involved in the problem could not accommodate the complexities of the natural river flow in a laboratory setting. Further, the lack of simple numerical and empirical models added to the quest to develop an accurate model for predicting scour depth around piers. To overcome these problems in the experimental and theoretical estimation of bridge scour, artificial intelligence has been widely used in recent years.

In recent years, machine learning models (Azmathullah *et al.* 2005; Babovic 2009; Karpatne *et al.* 2017; Chadalawada *et al.* 2020; Nearing *et al.* 2020) which allow computers to learn patterns from existing data sets without being explicitly programmed are being very widely used for this purpose (Azamathulla 2012). Artificial neural networks (ANN), which act as a universal approximator have been widely used for estimating bridge scour (Muzzammil 2008; Kaya 2010; Toth & Brandimarte 2011; Choi *et al.* 2015; Kızılöz *et al.* 2015). Raikar *et al.* (2016) applied ANN and genetic algorithms (GA) for prediction of scour depth within channel contractions. Many researchers also verified the efficacy of techniques like adaptive neuro-fuzzy inference system (ANFIS) and support vector machines (SVM) for predicting bridge scour (Bateni *et al.* 2007; Muzzammil 2010; Ghazanfari-Hashemi *et al.* 2011; Pal *et al.* 2011; Hong *et al.* 2012; Akib *et al.* 2014; Khan *et al.* 2014; Najafzadeh *et al.* 2016; Chou & Pham 2017). More recently, several studies have reported the use of hybrid techniques for predicting bridge scour (Chou & Pham 2014; Jannaty *et al.* 2015; Dang *et al.* 2019). Najafzadeh & Barani (2011) introduced the group method of data handling (GMDH) in the prediction of scour depth around the pier. Two models of the GMDH network were developed using genetic programming (GP) and a back propagation (BP) algorithm. The results showed that, the GMDH-GP performed better than the GMDH-BP in both training and testing phases. Najafzadeh (2015) utilized neuro-fuzzy based GMDH as an adaptive learning network to predict the scour depth at the downstream of grade control structures.

The machine learning techniques like ANN, SVM, and ANFIS have shown some promising results in the modeling of bridge scour. However, ANN models are susceptible to problems like local minima wherein, the optimization process may stop locally. SVM models face the problem of overfitting, depending on the used kernel function. The multivariate adaptive regression spline (MARS) models have the limitation to handle the large data and are less accurate for sparse data. Models like GTB use ensemble learning technique that combines various small learners to build a better learner. Boosting is the most widely used ensemble method that repeatedly applies a weak learner to modified versions of training data, thereby improving the predictive capability of the final model. Here, it becomes necessary to test the model for estimating bridge scour. In the present work, the efficacy of the GTB model to predict bridge scour around various pier shapes under different streambed conditions is tested. The performance of the proposed GTB model was further compared to the GMDH model.

Downloaded from http://iwa.silverchair.com/jh/article-pdf/23/4/849/910421/jh0230849.pdf
by guest

THEORETICAL OVERVIEW

Gradient tree boosting (GTB)

The gradient boosting (GB) is a machine learning approach that generates the model in the form of ensemble for the prediction studies. GB simplifies the random differences of loss function for optimizing the different levels of stages. Subsequently, the gradient boosting algorithm was established to optimize the cost function and iteratively select the function points to the negative direction of the gradient (Friedman 2001). The GTB approach is based on an ensemble algorithm, which involves numerous base models (Muller *et al.* 2018). In each of the base models, a distinct tree model is generated through bootstrapping the sample from the data used for the training and then further segregation of the feature space is done to the region sets. For every region, a simple model is then fitted.

In gradient boosting, a fixed size of the decision tree is used as a base learner to improve the fitting quality of every base learner. At the nth step of conventional gradient boosting, pseudo residuals fit to a decision tree $h_n(x)$. The number of leaves in the decision tree is denoted as i_n. The input space of the decision tree is then broken into distinct regions like $R_{1n}, R_{2n} \ldots R_{in}$ and the constant rate is predicted for each region. For the given input (x), the output of the decision tree $h_n(x)$ is represented as given in Equation (1):

$$h_n(x) = \sum_{i=1}^{i_n} b_{in} 1_{Rin}(x) \tag{1}$$

where, the predicted value in the region of R_{in} is denoted as b_{in}. A linear search is then adopted to choose the value of γ_n, the coefficient b_{in} is multiplied with γ_n to reduce the loss function. Therefore, the mathematical representation of model is restructured as:

$$F_n(x) = F_{n-1}(x) + \gamma_n h_n(x; a_m) \tag{2}$$

where, $h_n(x; a_m)$ can be selected for the function of X with a set of parameters $a_m = \{a_1^m \ldots . . a_n^m\}$. $f_o(X)$ can be selected to be equal to zero. The gradient descent-like procedure is adopted to calculate a_m and γ_n as follows:

$$a_m = avg\ min_{a,p} \sum_{i=1}^{N} [\bar{y}_{im} - Pf(x_i; a)]^2 \tag{3}$$

where, P is designed as step size

$$\bar{y}_{im} = -\left[\frac{\delta\ LOSS\ (y_i, F(x_i))}{\delta\ F(x_i)}\right]_{F(x)=F_{n-1}(x)} \tag{4}$$

$$\gamma_n = avg_\gamma\ min \sum_{i=1}^{n} L(y_i, F_{n-1}(x_i) + \gamma h_n(x_i)) \tag{5}$$

Also, a modified gradient boosting algorithm for the selection of distinct optimal value γ_{in}, for every region of tree, as an alternative of one γ_{in}, for the complete tree was proposed. Therefore, the coefficient b_{in} is disconnected in the formula for fitting the tree. A tree can be regarded as an additive model of the form, $m(x) = \sum_{i=1}^{N} k_i \times I(x \epsilon D_i)$ where, k_i are constants; $I(\cdot)$ is an implicit indicator function returning 1 if its argument is true and 0 otherwise; and D_i represents disjoint partitions of the training data. The modified algorithm is named as gradient tree boosting and the model equation is as below:

$$F_n(x) = F_{n-1}(x) + \sum_{i=1}^{in} \gamma_{in} 1R_{in}(x) \tag{6}$$

$$\gamma_n = avg_\gamma\ min \sum_{x_i \in R_{in}} L(y_i, F_{n-1}(x_i) + \gamma) \tag{7}$$

Further, the basic gradient boosting algorithm is improvised based on the constraints of tree and regularization.

Downloaded from http://iwa.silverchair.com/jh/article-pdf/23/4/849/910421/jh0230849.pdf
by guest

Constraints of tree

The parameter i in the equations represents the number of terminal nodes of a tree; it can be used for fine-tuning the parameters. In the GTB algorithm, i regulates the communication between the variables for the maximum number of times. It has been observed that the value of i ($4 \leq i \leq 8$) works in a superior way to boost the algorithm (Natekin & Knoll 2013).

Regularization

This reduces the effect of overfitting that occurs through the restrictions of the fitting procedure. Here, n is the regularization parameter, which represents the number of trees in the model. On increasing n, errors in the training set reduce. However, if the value of n is too high, then overfitting may occur. The ideal value of n is obtained through observing the error in the prediction. The second regularization parameter is the depth of the trees; the greater the value, the more the algorithm overfits the training data. The regularization can also be executed using shrinkage. The mathematical form is represented in Equation (8):

$$F_n(x) = F_{n-1}(x) + \mathrm{v} \cdot \gamma_n h_n(x) \tag{8}$$

Learning rate is represented by v $(0 < v \leq 1)$ and the smaller the value of v, the better generalized is the model. For comprehensive details related to GTB, the authors recommend referring to Ke *et al.* (2017) and Biau *et al.* (2019).

Group method of data handling (GMDH)

The GMDH algorithm is modeled with multi-parametric data sets and is completely an automated structure. The procedure of GMDH is categorized as an inductive process. In this structure, complex polynomial models are sorted out progressively and the finest model is identified in the view of external criterion (Onwubolu 2016).

The GMDH model provides one output data pair after considering numerous input data pairs. The model is established with a fixed number of neurons. In each layer, neurons transmit different input pairs with a quadratic polynomial and produce the neuron for the next layer. The following is the base function of GMDH that produces the output pair from the subset of components:

$$x(y_1, y_2, y_3 \ldots, y_n) = b_0 + \sum_{i=1}^{n} b_i g_i \tag{9}$$

where, g is the fundamental function determined with the diverse class of inputs and the coefficients are denoted by b. Additionally, n is the quantity of the element of base function and i is $1, 2, 3 \ldots n$.

It is apparent that training the GMDH connected ANN is possible for the prediction of output value $\hat{x}_i$ for any certain input vector $x = (y_{i1}, y_{i2}, y_{i3} \ldots . y_{in})$.

To identify the ideal solution, the GMDH algorithm takes a different element from the subclass of the base function called partial models. It employs least squares method for the identification of model coefficients. The GMDH algorithm gradually progresses the number of components in the partial model and determines the finest structure with ideal complexity, represented by the smallest possible value of external criterion. Therefore, the procedure of the GMDH algorithm is known as self-organization of models.

To establish the neural network type of GMDH, the differences of experimental and predicted values are squared and the output forecasted is then minimized as given in Equation (10):

$$\sum_{i=1}^{n} [g_i(y_{i1}, y_{i2}, y_{i3} \ldots . y_{in}) - x_i] \rightarrow min \tag{10}$$

The complex discrete form of Volterra functional series represents the linking of input and output values. The primary base function of GMDH is complicated and numerous base models are generated with a maximum of second-degree functions.

$$X(y_i, \ldots y_n) = b_0 + \sum_{i=1}^{n} b_i g_i + \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=j}^{n} b_{ijk} y_i y_j y_k + \ldots \tag{11}$$

Downloaded from http://iwa.silverchair.com/jh/article-pdf/23/4/849/910421/jh0230849.pdf
by guest

The partial quadratic polynomials (Jamali *et al.* 2009) represent the detailed mathematical form of Equation (11) by comprising only with two variables as shown in Equation (12):

$$\hat{x} = g(y_i, y_j) = b_0 + b_1 y_i + b_2 y_j + b_3 y_i y_j + b_4 y_i^2 + b_5 y_j^2 \tag{12}$$

In the GDMH type of neural network, the partial quadratic is implemented recursively for the construction of the mathematical relationship between the input and output variables as shown in Equation (11). In Equation (12), b_0 are the coefficients, which are determined using the least squares for reducing the differences of calculated and observed outputs. Consequently, polynomial trees are constructed using the quadratic form (Müller *et al.* 1998).

DATA ANALYSIS

The data for the present study were taken from the experimental investigation conducted by Goswami (2013). The experiments were conducted under both clear-water and live-bed scour conditions. The scour depth was analyzed in connection with the following independent parameters: approach velocity, flow time, and median sediment size (d_{50}) in the case of clear-water scour and correspondingly in the case of live-bed scour assessment, sediment quantity (ppm) was considered as a parameter along with approach velocity and flow time. Additionally, the relative importance of each parameter with respect to the total variation of the scour depth was explored via experimental study carried out by Goswami (2013). In the clear-water scour case, the sediments used in the study were of average diameter, ranging between 0.42 and 4.2 mm. The trial was run for 6 hr duration and the data were recorded at every 1 hr interval. Different velocities considered for the clear-water scour were in the range of 0.184 to 0.351 m/sec. Likewise, in the live-bed scour case, the sediment quantity used ranged between 747.78 ppm and 1,066.67 ppm. The velocity was maintained within the range of 0.226 to 0.251 m/sec. The piers of circular, rectangular, round-nosed and sharp-nosed shapes were tested for both the conditions.

METHODOLOGY

Model development

The local transient scour depth was predicted using the three input parameters, namely, sediment size (d_{50}), velocity and flow time in the case of clear-water scour condition. Correspondingly, in the case of live-bed scour condition, the input parameters considered were sediment quantity, velocity, and flow time. The scour depth was predicted in both condition for all four types of pier shapes mentioned earlier. The descriptive statistics in terms of maximum, minimum, mean, standard deviation, and kurtosis for every input and output parameter of all four shapes for both training and testing under clear-water and live-bed scour conditions are tabulated in Tables 1 and 2.

The whole data set was divided randomly into training and testing data sets. The soft computing GTB and GMDH models were developed by optimizing the model parameters. Performance of the GTB model depends on the various parameters such

Table 1 | Descriptive statistics – clear-water scour

		Variables						
					Scour depth (mm)			
Data set	Statistical parameters	Sediment size, d_{50} (mm)	Velocity (m/s)	Flow time (hr)	Circular	Rectangular	Round nosed	Sharp nosed
Training	Max	4.2	0.351	6	118	122	113	120
	Min	0.42	0.184	0	55	55	53	53
	Mean	2.31	0.261	3	84.262	87.89	82.512	84.77
	SD	1.89	0.0515	2	14.166	14.99	13.06	13.799
	CV	0.818	0.197	0.667	0.168	0.171	0.158	0.163
Testing	Max	4.2	0.351	6	115	121	111	120
	Min	0.42	0.184	0	54	55	54	55
	Mean	2.31	0.261	3	84.512	88.34	83.32	84.14
	SD	1.89	0.0515	2	14.206	14.58	13.23	13.33
	CV	0.818	0.197	0.667	0.168	0.165	0.159	0.158

SD: standard deviation; CV: coefficient of variation.

Downloaded from http://iwa.silverchair.com/jh/article-pdf/23/4/849/910421/jh0230849.pdf
by guest

Table 2 | Descriptive statistics – live-bed scour

		Variables						
					Scour depth (mm)			
Data set	Statistical parameters	Sediment quantity (ppm)	Velocity (m/s)	Flow time (hr.)	Circular	Rectangular	Round-nosed	Sharp-nosed
Training	Max	1,066.67	0.251	4	98	108	98	99
	Min	747.78	0.226	0	71	71	70	68
	Mean	907.225	0.2385	2	83.575	89.213	83.513	84.48
	SD	159.445	0.0125	1.414	7.692	9.907	7.20	8.24
	CV	0.176	0.052	0.707	0.092	0.111	0.086	0.098
Testing	Max	1,066.67	0.251	4	99	106	97	98
	Min	747.78	0.226	0	70	73	68	68
	Mean	907.225	0.2385	2	83.825	89.363	83.35	85.24
	SD	159.445	0.0125	1.414	7.938	10.024	7.39	8.95
	CV	0.176	0.052	0.707	0.095	0.112	0.089	0.105

SD: standard deviation; CV: coefficient of variation.

as number of estimators, maximum depth, minimum sample split, learning rate, and loss function. The GMDH performance is based on the model parameters such as maximum number of layers, maximum number of layers, selection pressure, and number of loops. Table 3 lists the key GMDH and GTB parameters that have been tuned to build effective model architectures.

Performance evaluation

The predicted scour depths from the soft computing models are compared with measured scour depth values. The model prediction efficiency is evaluated using the following statistical indices, relative root mean square error (RRMSE), normalized Nash–Sutcliffe efficiency (NNSE), Wilmott index (WI), mean normalized error (MNE), scatter index (SI), and Kling-Gupta efficiency (KGE).

$$RRMSE = \frac{RMSE}{\sigma_{obs}}, \quad 0 \leq RRMSE \leq 1 \tag{13}$$

$$NNSE = \frac{1}{2 - NSE} \quad 0 \leq NNSE \leq 1 \tag{14}$$

$$WI = 1 - \frac{\sum_{i=1}^{N}(X_i - Y_i)^j}{\sum_{i=1}^{N}(|Y_i - \overline{X}_i| + |X_i - \overline{X}_i|)^j}, \quad 0 \leq WI \leq 1 \tag{15}$$

Table 3 | Parameters of the GMDH and GTB models

Parameters	Clear-water scour condition	Live-bed scour condition
GMDH		
Maximum number of layers	5	5
Maximum number of neurons	25	20
Selection pressure (α)	0.05	0.05
Number of loops	50	50
GTB		
n_estimators	100	100
max_depth	2	2
min_samples_split	6	4
learning_rate	0.1	0.1
loss	ls[a]	ls[a]

[a]ls: least squares.

Downloaded from http://iwa.silverchair.com/jh/article-pdf/23/4/849/910421/jh0230849.pdf
by guest

$$MNE = \frac{1}{N}\frac{\sum_{i=1}^{N}(X_i - Y_i)}{X_i} X\ 100\ (\%), \quad 0 \le MNE \le +\infty \tag{16}$$

$$SI = \frac{RMSE}{\overline{X_i}}, \quad 0 \le SI \le 1 \tag{17}$$

$$KGE = 1 - \sqrt{(R-1)^2 + (\beta-1)^2 + (\gamma-1)^2}, \quad 0 \le KGE \le +1 \tag{18}$$

where,

$$Correlation\ coefficient,\ R = \left[\frac{\sum_{i=1}^{N}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{N}(X_i - \bar{X})^2 \sum_{i=1}^{N}(Y_i - \bar{Y})^2}}\right]$$

$$Bias\ Ratio,\ \beta = \frac{\bar{Y}}{\bar{X}}$$

$$Variability,\ \gamma = \frac{CV_Y}{CV_X} = \frac{\frac{\sigma_y}{\bar{Y}}}{\frac{\sigma_x}{\bar{X}}}$$

$$Root\ Mean\ Square\ Error,\ RMSE = \sqrt{\frac{\sum_{i=1}^{N}(X_i - Y_i)^2}{N}}$$

$$Nash\text{–}Sutcliffe\ efficiency,\ NSE = 1 - \left(\frac{\sum_{i=1}^{N}(X_i - Y_i)^2}{\sum_{i=1}^{N}(X_i - \bar{X})^2}\right)$$

Table 4 | Performance of the GTB and GMDH models in prediction of scour depth under clear-water scour and live-bed scour conditions in the test phase

		Clear-water scour		Live-bed scour	
Pier shapes	**Statistical indices**	**GTB**	**GMDH**	**GTB**	**GMDH**
Circular	RRMSE	0.290	0.360	0.397	0.390
	NNSE	0.922	0.890	0.862	0.869
	WI	0.977	0.965	0.955	0.957
	MNE	3.39	4.40	2.97	2.91
	SI	0.049	0.060	0.038	0.037
	KGE	0.936	0.891	0.882	0.875
Rectangular	RRMSE	0.410	0.530	0.320	0.340
	NNSE	0.859	0.780	0.910	0.896
	WI	0.955	0.922	0.974	0.968
	MNE	5.008	6.88	2.59	2.76
	SI	0.067	0.088	0.036	0.038
	KGE	0.892	0.845	0.938	0.908
Round-nosed	RRMSE	0.310	0.480	0.390	0.401
	NNSE	0.912	0.810	0.862	0.860
	WI	0.974	0.937	0.956	0.957
	MNE	3.54	6.12	2.66	2.87
	SI	0.050	0.077	0.035	0.036
	KGE	0.911	0.882	0.904	0.915
Sharp-nosed	RRMSE	0.410	0.510	0.350	0.420
	NNSE	0.852	0.791	0.890	0.848
	WI	0.952	0.918	0.963	0.942
	MNE	4.750	5.670	2.980	3.590
	SI	0.066	0.081	0.037	0.044
	KGE	0.885	0.800	0.865	0.780

Downloaded from http://iwa.silverchair.com/jh/article-pdf/23/4/849/910421/jh0230849.pdf
by guest

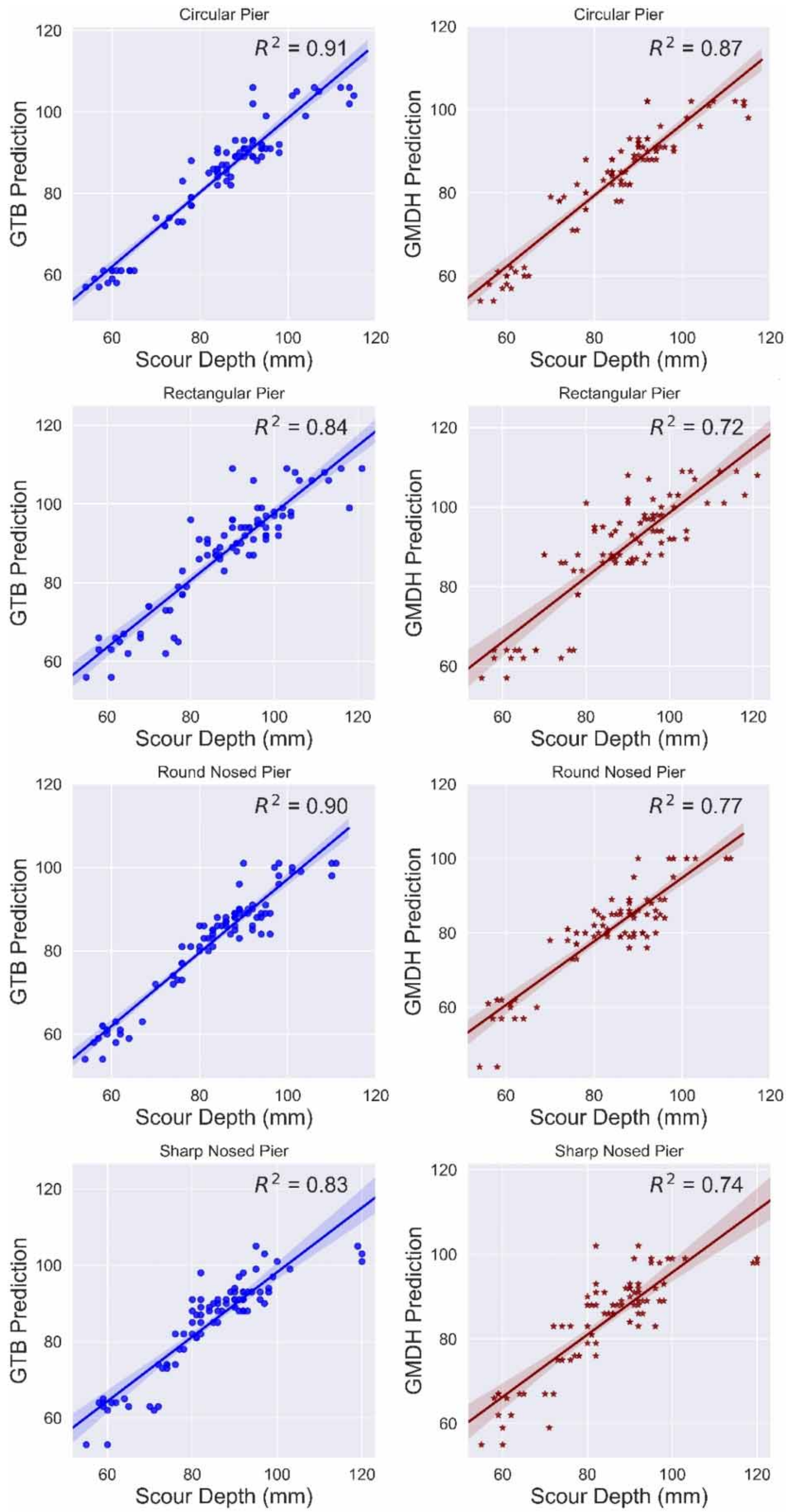

Figure 2 | Scatter plots of the GTB and GMDH models with respect to scour depth prediction under clear-water scour condition.

Downloaded from http://iwa.silverchair.com/jh/article-pdf/23/4/849/910421/jh0230849.pdf
by guest

where, σ_{obs} is standard deviation of measured data; X is observed/measured values and Y is predicted values; N is number of total data set points; $\overline{X_i}$ is mean of actual data and $\overline{Y_i}$ is mean of predicted data; and j is exponent term.

RESULTS AND DISCUSSION

The simulation results obtained from the GTB and GMDH models exhibit their suitability for the prediction of scour depth around various pier shapes under different streambed conditions. The GMDH architecture with networks of various size (1–5 layers) have been trained for determining the best-fit (optimal) GMDH models. Configuration of GTB involves tuning of parameters such as number of trees, tree depth, step size (learning rate), etc.

Prediction of scour depth under clear-water scour condition

Fairly accurate scour depth predictions (around all the pier shapes considered) under clear-water condition were offered by the GMDH and GTB models. The performance of the models evaluated in terms of error and accuracy measures are presented in Table 4. The scour depth predictions around circular and round-nosed pier shapes were relatively accurate and precise compared to those of rectangular and sharp-nosed pier shapes. On comparison of RRMSE values of GTB model predictions obtained for each of the pier shapes considered, the model calibrated for circular pier shape data had the least RRMSE (0.290) indicating a robust performance during testing. Additionally, the GTB model predictions were relatively superior to those of the GMDH model in each of the pier shapes considered. The GTB model calibrated for circular pier shape data ranked first with superior test phase NNSE index followed by the GTB models of round-nosed, rectangular and sharp-nosed pier shapes. On comparison between GMDH and GTB models, the differences in the extent of goodness-of-fit (WI) measure were 1.2%, 3.7%, 3.3%, and 3.4%, with reference to circular, round-nosed, rectangular and sharp-

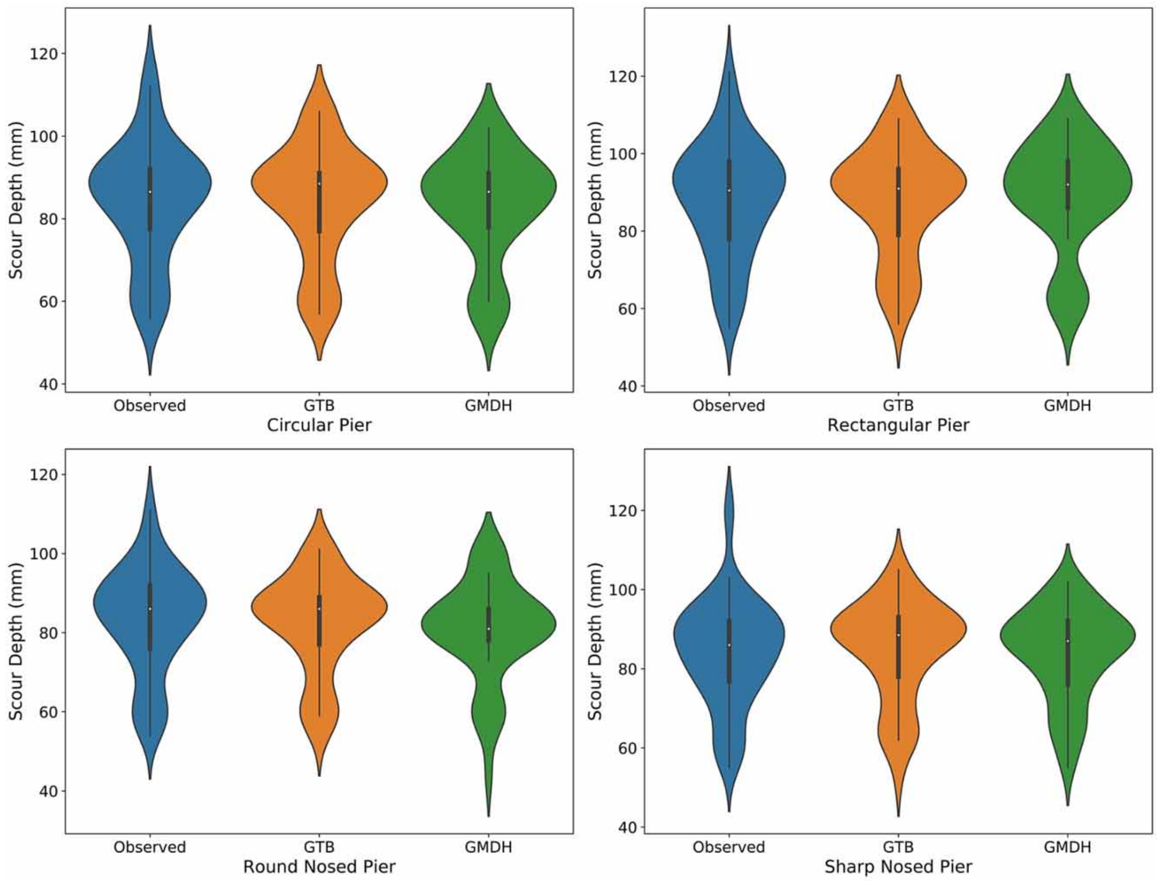

Figure 3 | Violin plots for comparative evaluation of the GTB and GMDH models used for the prediction of scour depth under clear-water scour condition. Violin plots are very simple to understand. The white dot in the middle represents the median. The thick black box depicts the interquartile range and the whiskers show 95% confidence interval. Finally, the shape of the violin displays frequencies of values.

Downloaded from http://iwa.silverchair.com/jh/article-pdf/23/4/849/910421/jh0230849.pdf
by guest

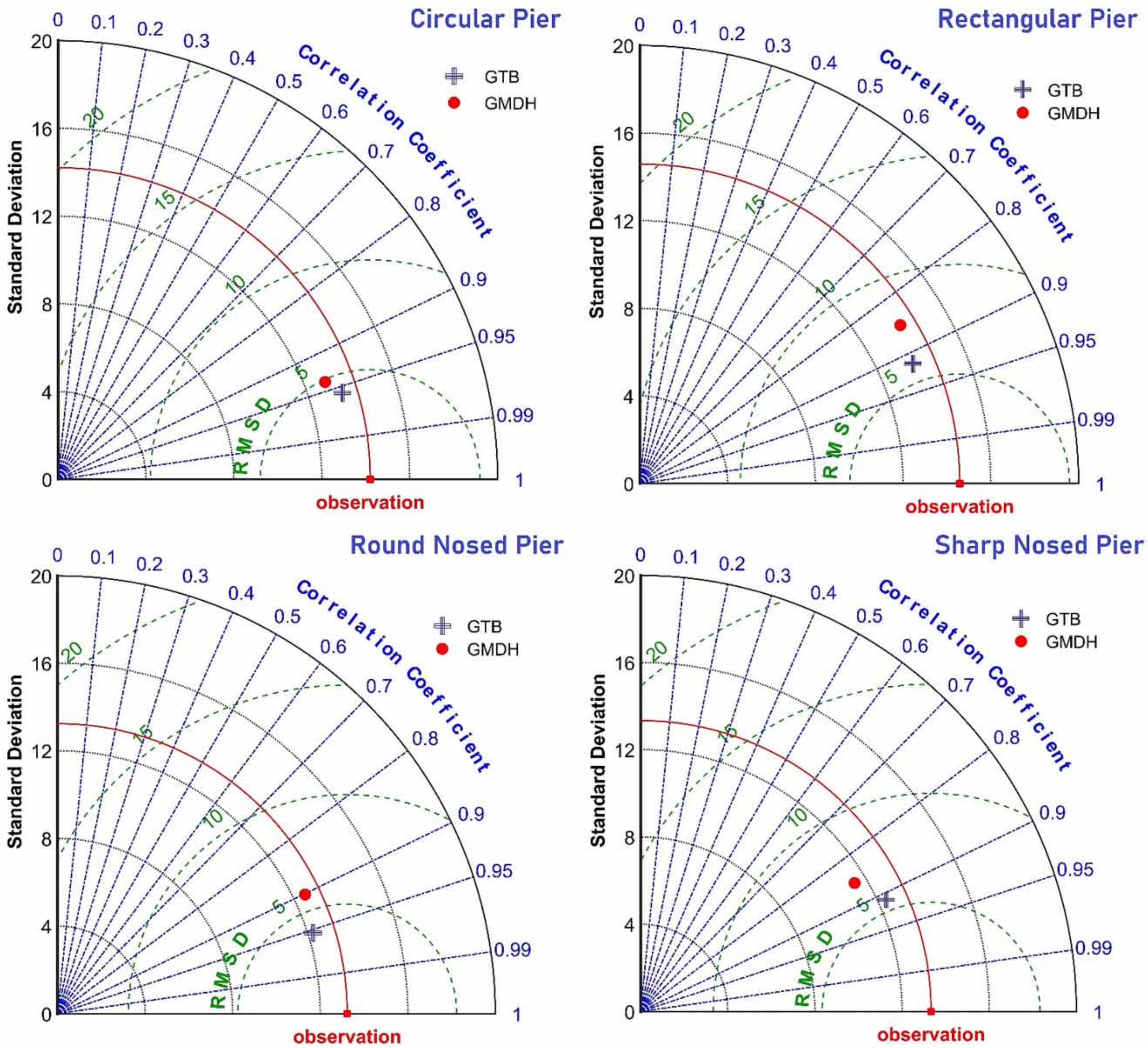

Figure 4 | Taylor diagrams for comparative evaluation of the GTB and GMDH models used for the prediction of scour depth under clear-water scour condition.

nosed pier shapes, respectively. The GTB predictions were highly precise with reference to circular pier shape, with performance statistics of MNE = 3.39, KGE = 0.936 and the scatter index of 0.049. However, with reference to sharp-nosed pier shape, the GTB offered relatively less prediction skill with MNE = 4.75, the KGE = 0.885, and the scatter index of 0.066.

Table 5 | Performance of the GTB and GMDH models against SVM, ANFIS, and PSO-SVM models in prediction of scour depth under clear-water scour condition

		Clear-water scour				
Pier shapes	**Statistical index**	**GTB**	**GMDH**	**SVM[a]**	**ANFIS[a]**	**PSO-SVM[a]**
Circular	R^2	0.91	0.87	0.83	0.91	0.90
Rectangular	R^2	0.84	0.72	0.66	0.82	0.83
Round-nosed	R^2	0.90	0.77	0.72	0.90	0.90
Sharp-nosed	R^2	0.83	0.74	0.71	0.83	0.85

[a]Sreedhara *et al.* (2019).

Downloaded from http://iwa.silverchair.com/jh/article-pdf/23/4/849/910421/jh0230849.pdf
by guest

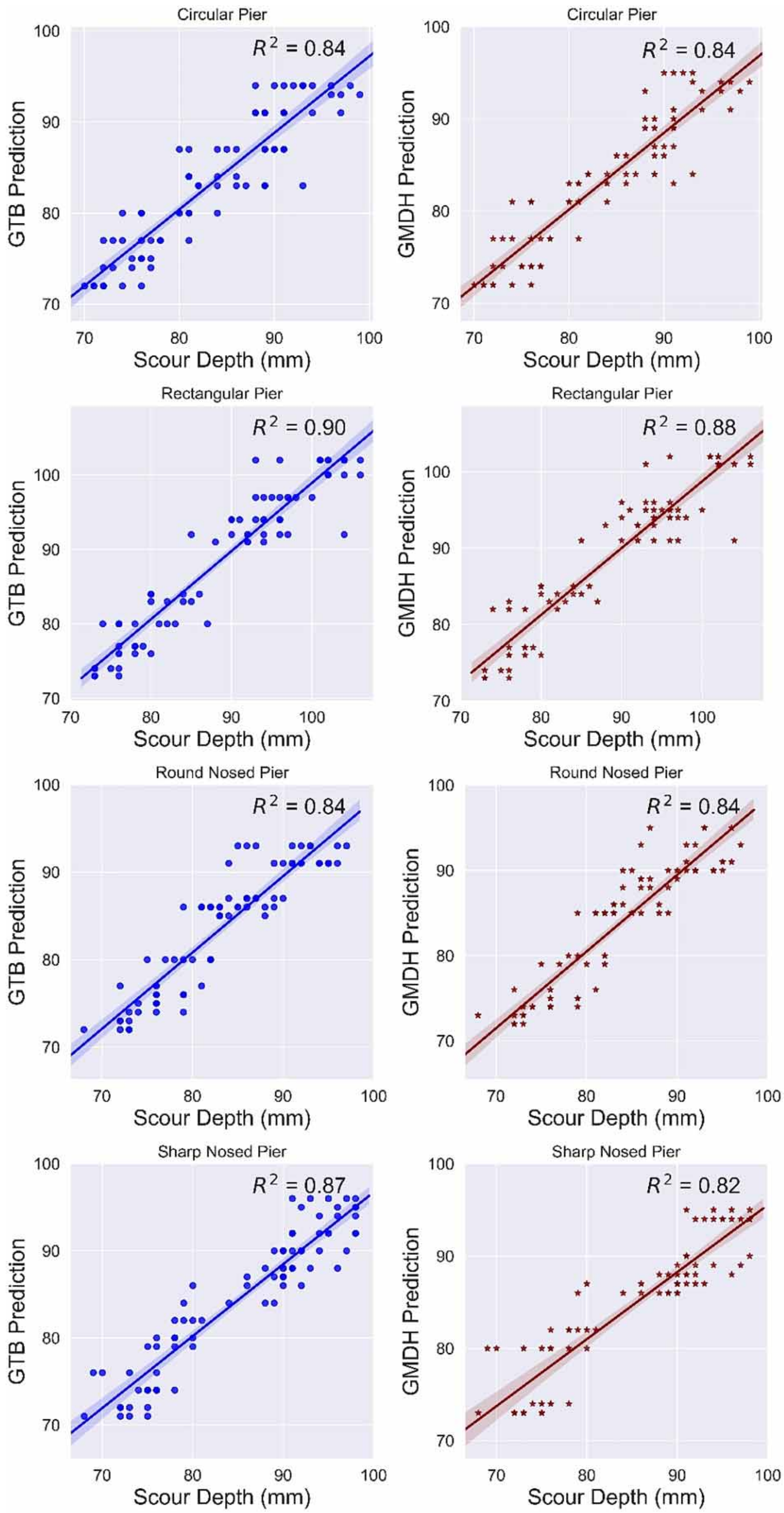

Figure 5 | Scatter plots of the GTB and GMDH models with respect to scour depth prediction under live-bed scour condition.

Downloaded from http://iwa.silverchair.com/jh/article-pdf/23/4/849/910421/jh0230849.pdf
by guest

The GTB modeled scour depths were in close agreement with the observed (scour depth) measurements with the coefficient of determination (R^2) above 0.90 in the case of circular and round-nosed pier shapes (refer to Figure 2). The scatter plots of GMDH models showed some lack of association between expected and modeled data along with many outlier points. The violin plots (Figure 3) depict the results of scour depth predictions from the GTB model with that of the observed and GMDH model predictions. The density curves of observed and GTB predictions complement each other in terms of symmetry, skew, and variability characteristics. The Taylor diagrams (Figure 4) plotted to comparatively evaluate the individual model performances based on multiple statistical indicators (RMSD, R, and standard deviation) portray the superiority and predictive power of the GTB model against the GMDH model.

Since the models developed in the current study used the same data and model (input–output) structure of the earlier research by Sreedhara *et al.* (2019), the performance of GTB and GMDH models were further compared against the SVM, (ANFIS), and hybrid particle swarm optimization-based SVM (PSO-SVM) models. Table 5 presents the performance of GTB and GMDH models and other AI models (SVM, ANFIS, PSO-SVM), evaluated in terms of coefficient of determination (R^2) statistic. It is evident from Table 5 that the newly developed GTB model revealed similar capabilities in the prediction process compared to the hybrid PSO-SVM model.

Prediction of scour depth under live-bed scour condition

Similar to clear-water scour condition, GMDH and GTB models offered relatively reliable predictions of scour depth under live-bed scour condition around all the pier shapes considered. The evaluation metrics tabulated in Table 4 render the test phase performance of the GMDH and GTB models constructed for different pier shape data. The efficiency index (NNSE) of the GTB model shows that there is good agreement between the observed and predicted scour depths around

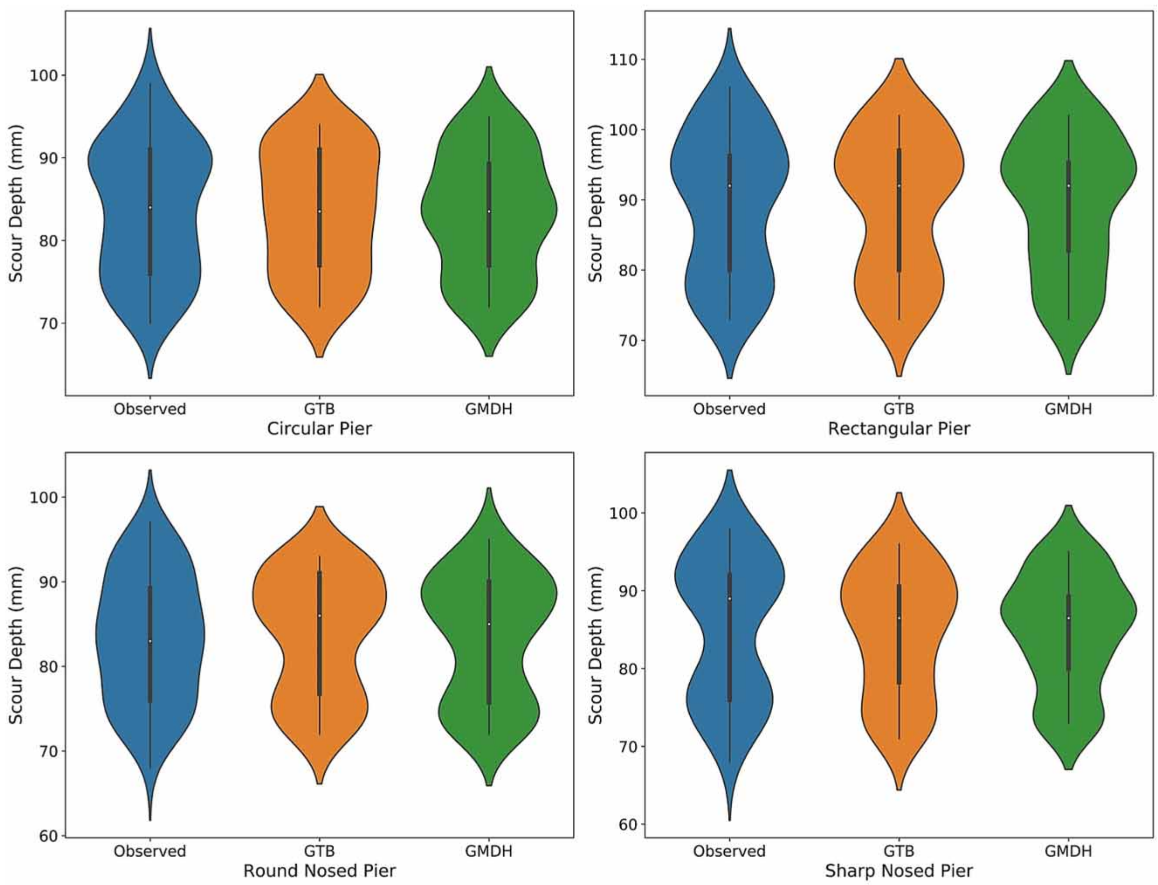

Figure 6 | Violin plots for comparative evaluation of the GTB and GMDH models used for the prediction of scour depth under live-bed scour condition. Please refer to the online version of this paper to see this figure in colour: http://dx.doi.org/10.2166/hydro.2021.011.

Downloaded from http://iwa.silverchair.com/jh/article-pdf/23/4/849/910421/jh0230849.pdf
by guest

each of the pier shapes considered. For sharp-nosed pier data, the RRMSE of the GTB model was 0.35, which was 0.833 times less than that of the GMDH model (0.42). Among the GTB models, the highest WI of 0.974 was observed for simulations of rectangular pier shape and the minimum WI of 0.955 was witnessed for circular pier shape. It is noteworthy that the WI and the NNSE metrics convey similar skill information. Additionally, the differences in scatter index of GTB and GMDH models were relatively less in the case of circular, rectangular and round-nosed pier shapes portraying a relatively equal performance.

Analysis of scatter plots (Figure 5) suggests that the scour depth predictions from the GTB model had an improved R^2 measure compared to that of GMDH model predictions. In scatterplots, corresponding to high values of R^2 lesser variability is reflected in the point cloud. The scour depth predictions around rectangular and sharp-nosed pier shapes were relatively accurate and precise compared to that of circular and round-nosed pier shapes. There is good evidence of this in the violin plots presented in Figure 6, which clearly depict large differences in the magnitude and shape of the distribution between the observed and modeled (GTB and GMDH) scour depths. From Taylor diagrams (Figure 7), it is evident that both GTB and GMDH models tend to underestimate the variability of scour depth observations as they both lie within the red observation line. However, the GTB models do best in terms of high correlation with the observed scour depth and lowest RMS deviations and, quantitatively, the GTB model achieved the highest KGE with reference to circular, rectangular and sharp-nosed pier shapes.

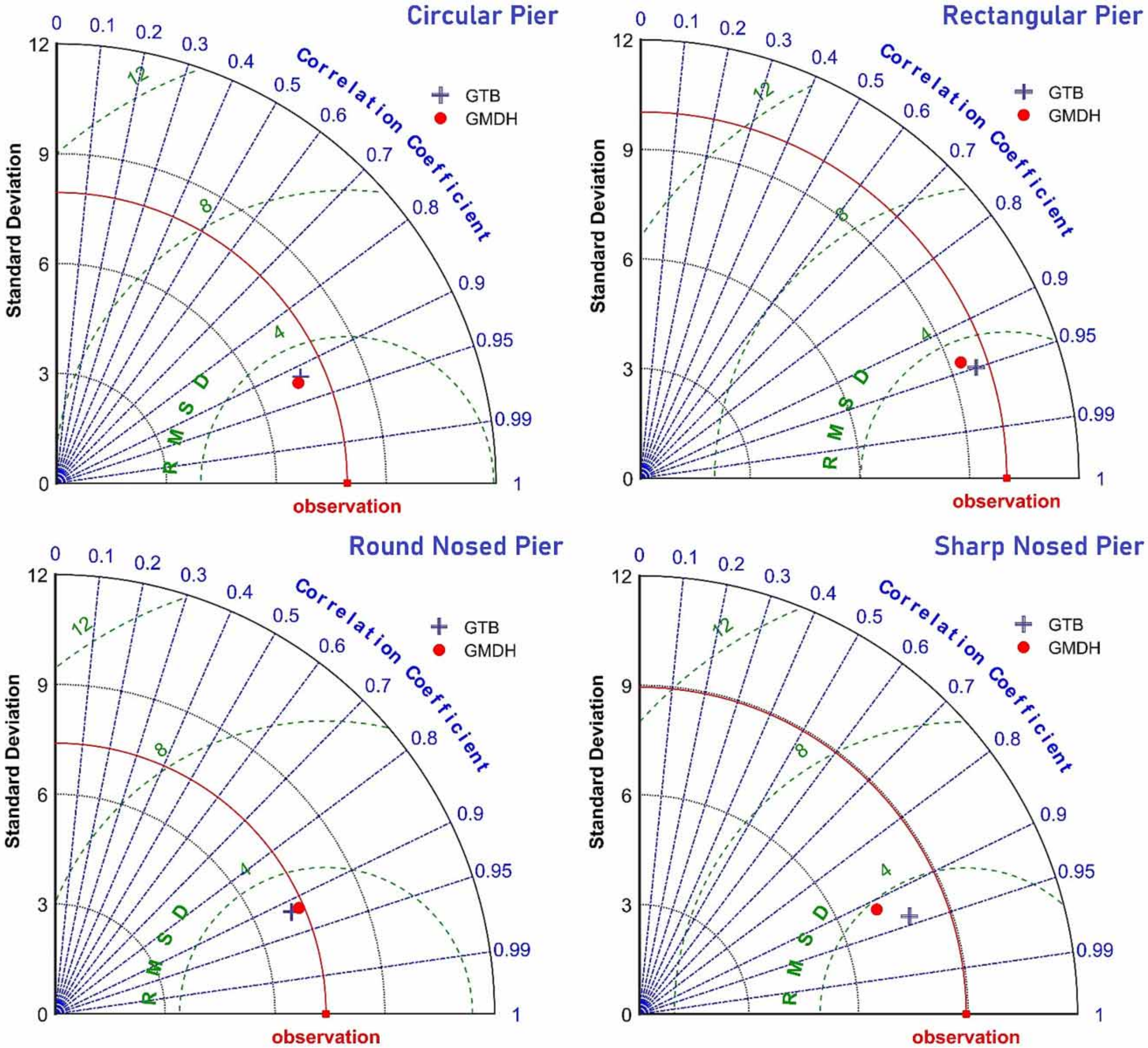

Figure 7 | Taylor diagrams for comparative evaluation of the GTB and GMDH models used for the prediction of scour depth under live-bed scour condition.

Downloaded from http://iwa.silverchair.com/jh/article-pdf/23/4/849/910421/jh0230849.pdf
by guest

CONCLUSIONS

In this study, GTB and GMDH models were developed to predict the scour depth under clear-water and live-bed scour conditions for different pier shapes. The results of the developed models were then analyzed using statistical parameters and compared with measured values. The following conclusions were drawn from the study:

1. The GTB model performed with a relatively better accuracy and efficiency when compared to the GMDH model under both the scour conditions and for all pier shapes.
2. In clear-water scour condition, the GTB model performed well for circular shape with higher NSE and WI values compared to the other three pier shapes.
3. For live-bed scour condition, the GTB model performed better for rectangular shape compared to the other three pier shapes.
4. The results obtained in terms of all the statistical indices show that the GTB model better predicted the scour and can be used as an alternative tool for scour depth prediction under clear-water and live-bed scour conditions.

ACKNOWLEDGEMENTS

The authors would like to express their sincere thanks to Dr. Goswami Pankaj and his supervisor Dr. Saikia Bibha Das, Gauhati University for providing the data.

DATA AVAILABILITY STATEMENT

All relevant data are available from an online repository or repositories (https://shodhganga.inflibnet.ac.in/handle/10603/74698).

REFERENCES

Akib, S., Mohammadhassani, M. & Jahangirzadeh, A. 2014 Application of ANFIS and LR in prediction of scour depth in bridges. *Computers and Fluids* **91**, 77–86. http://dx.doi.org/10.1016/j.compfluid.2013.12.004.

Azmathullah, H. M., Deo, M. C. & Deolalikar, P. B. 2005 Neural networks for estimation of scour downstream of a ski-jump bucket. *Journal of Hydraulic Engineering* **131** (10), 898–908.

Azamathulla, H. M. 2012 Gene-expression programming to predict scour at a bridge abutment. *Journal of Hydroinformatics* **14** (2), 324–331.

Babovic, V. 2009 Introducing knowledge into learning based on genetic programming. *Journal of Hydroinformatics* **11** (3–4), 181–193.

Bateni, S. M., Borghei, S. M. & Jeng, D.-S. 2007 Neural network and neuro-fuzzy assessments for scour depth around bridge piers. *Engineering Applications of Artificial Intelligence* **20** (3), 401–414. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0952197606001230.

Biau, G., Cadre, B. & Rouvière, L. 2019 Accelerated gradient boosting. *Machine Learning* **108** (6), 971–992. https://doi.org/10.1007/s10994-019-05787-1.

Chadalawada, J., Herath, H. M. V. V. & Babovic, V. 2020 Hydrologically informed machine learning for rainfall-runoff modeling: a genetic programming-based toolkit for automatic model induction. *Water Resources Research* **56** (4), 1–23.

Choi, S. U., Choi, B. & Choi, S. 2015 Improving predictions made by ANN model using data quality assessment: an application to local scour around bridge piers. *Journal of Hydroinformatics* **17** (6), 977–989.

Chou, J.-S. & Pham, A.-D. 2014 Hybrid computational model for predicting bridge scour depth near piers and abutments. *Automation in Construction* **48**, 88–96. http://dx.doi.org/10.1016/j.autcon.2014.08.006.

Chou, J.-S. & Pham, A.-D. 2017 Nature-inspired metaheuristic optimization in least squares support vector regression for obtaining bridge scour information. *Information Sciences* **399**, 64–80. http://dx.doi.org/10.1016/j.ins.2017.02.051.

Dang, N. M., Duong, T. A. & Dang, T. D. 2019 ANN optimized by PSO and firefly algorithms for predicting scour depths around bridge piers. *Engineering with Computers* **37**, 293–303. https://doi.org/10.1007/s00366-019-00824-y.

Farooq, R. & Ghumman, A. R. 2019 Impact assessment of pier shape and modifications on scouring around bridge pier. *Water (Switzerland)* **11** (9), 14–20.

Friedman, J. H. 2001 Greedy function approximation: a gradient boosting machine. *The Annals of Statistics* **29** (5), 1189–1232. Available from: https://projecteuclid.org/journals/annals-of-statistics/volume-29/issue-5/Greedy-function-approximation-A-gradient-boosting-machine/10.1214/aos/1013203451.full.

Ghazanfari-Hashemi, S., Etemad-Shahidi, A., Kazeminezhad, M. H. & Mansoori, A. R. 2011 Prediction of pile group scour in waves using support vector machines and ANN. *Journal of Hydroinformatics* **13** (4), 609–620. Available from: https://iwaponline.com/jh/article/13/4/609/3105/Prediction-of-pile-group-scour-in-waves-using.

Goswami, P. 2013 *Evaluation of Scour Depth Around Bridge Piers*. PhD Thesis, Department of Civil engineering, Gauhati University, India. Available from: https://shodhganga.inflibnet.ac.in/handle/10603/74698 (accessed 1 January 2021).

Downloaded from http://iwa.silverchair.com/jh/article-pdf/23/4/849/910421/jh0230849.pdf
by guest

Hong, J.-H., Goyal, M. K., Chiew, Y.-M. & Chua, L. H. C. 2012 Predicting time-dependent pier scour depth with support vector regression. *Journal of Hydrology* **468–469**, 241–248. http://dx.doi.org/10.1016/j.jhydrol.2012.08.038.

Jamali, A., Nariman-zadeh, N., Darvizeh, A., Masoumi, A. & Hamrang, S. 2009 Multi-objective evolutionary optimization of polynomial neural networks for modelling and prediction of explosive cutting process. *Engineering Applications of Artificial Intelligence* **22** (4–5), 676–687. http://dx.doi.org/10.1016/j.engappai.2008.11.005.

Jannaty, M. H., Eghbalzadeh, A. & Hosseini, S. A. 2015 Hybrid ANFIS model for predicting scour depth using particle swarm optimization. *Indian Journal of Science and Technology* **8** (22), 83–89. Available from: https://indjst.org/articles/hybrid-anfis-model-for-predicting-scour-depth-using-particle-swarm-optimization.

Johnson, P. A., Clopper, P. E., Zevenbergen, L. W. & Lagasse, P. F. 2015 Quantifying uncertainty and reliability in bridge scour estimations. *Journal of Hydraulic Engineering* **141** (7), 04015013. Available from: http://ascelibrary.org/doi/10.1061/%28ASCE%29HY.1943-7900.0001017.

Karpatne, A., Atluri, G., Faghmous, J. H., Steinbach, M., Banerjee, A., Ganguly, A., Shekhar, S., Samatova, N. & Kumar, V. 2017 Theory-guided data science: a new paradigm for scientific discovery from data. *IEEE Transactions on Knowledge and Data Engineering* **29** (10), 2318–2331.

Kaya, A. 2010 Artificial neural network study of observed pattern of scour depth around bridge piers. *Computers and Geotechnics* **37** (3), 413–418. http://dx.doi.org/10.1016/j.compgeo.2009.10.003.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q. & Liu, T.-Y. 2017 LightGBM: a highly efficient gradient boosting decision tree. In: *NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems*. December 2017, Long Beach, CA, USA, pp. 3149–3157.

Khan, M., Andres, D., Neill, C. & Ab Ghani, A. 2014 Discussion: bridge pier scour prediction by gene expression programming. *Proceedings of the Institution of Civil Engineers: Water Management* **167** (6), 368–369.

Kızılöz, B., Çevik, E. & Aydoğan, B. 2015 Estimation of scour around submarine pipelines with artificial neural network. *Applied Ocean Research* **51**, 241–251. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0141118715000516.

Müller, J. A., Ivachnenko, A. G. & Lemke, F. 1998 GMDH algorithms for complex systems modelling. *Mathematical and Computer Modelling of Dynamical Systems* **4** (4), 275–316. Available from: http://www.tandfonline.com/doi/abs/10.1080/13873959808837083.

Muller, F., Schug, D., Hallen, P., Grahe, J. & Schulz, V. 2018 Gradient tree boosting-based positioning method for monolithic scintillator crystals in positron emission tomography. *IEEE Transactions on Radiation and Plasma Medical Sciences* **2** (5), 411–421. Available from: https://ieeexplore.ieee.org/document/8360486/.

Muzzammil, M. 2008 Application of neural networks to scour depth prediction at the bridge abutments. *Engineering Applications of Computational Fluid Mechanics* **2** (1), 30–40. Available from: http://www.tandfonline.com/doi/abs/10.1080/19942060.2008.11015209.

Muzzammil, M. 2010 ANFIS approach to the scour depth prediction at a bridge abutment. *Journal of Hydroinformatics* **12** (4), 474–485.

Najafzadeh, M. & Barani, G.-A. 2011 Comparison of group method of data handling based genetic programming and back propagation systems to predict scour depth around bridge piers. *Scientia Iranica* **18** (6), 1207–1213. http://dx.doi.org/10.1016/j.scient.2011.11.017.

Najafzadeh, M. 2015 Neuro-fuzzy GMDH based particle swarm optimization for prediction of scour depth at downstream of grade control structures. *Engineering Science and Technology* **18** (1), 42–51. http://dx.doi.org/10.1016/j.jestch.2014.09.002.

Najafzadeh, M., Etemad-Shahidi, A. & Lim, S. Y. 2016 Scour prediction in long contractions using ANFIS and SVM. *Ocean Engineering* **111**, 128–135. http://dx.doi.org/10.1016/j.oceaneng.2015.10.053.

Natekin, A. & Knoll, A. 2013 Gradient boosting machines, a tutorial. *Frontiers in Neurorobotics* **7**, 21. Available from: http://journal.frontiersin.org/article/10.3389/fnbot.2013.00021/abstract.

Nearing, G. S., Kratzert, F., Sampson, A. K., Pelissier, C. S., Frame, J. M., Prieto, C. & Gupta, H. V. 2020 What role does hydrological science play in the age of machine learning? *Water Resources Research* **57** (3), e2020WR028091. https://doi.org/10.1029/2020WR028091.

Onwubolu, G. 2016 *GMDH-Methodology and Implementation in MATLAB*. Imperial College Press, London. https://doi.org/10.1142/p982.

Pal, M., Singh, N. K. & Tiwari, N. K. 2011 Support vector regression based modeling of pier scour using field data. *Engineering Applications of Artificial Intelligence* **24** (5), 911–916. http://dx.doi.org/10.1016/j.engappai.2010.11.002.

Raikar, R. V., Wang, C. Y., Shih, H. P. & Hong, J. H. 2016 Prediction of contraction scour using ANN and GA. *Flow Measurement and Instrumentation* **50**, 26–34. http://dx.doi.org/10.1016/j.flowmeasinst.2016.06.006.

Sreedhara, B. M., Rao, M. & Mandal, S. 2019 Application of an evolutionary technique (PSO–SVM) and ANFIS in clear-water scour depth prediction around bridge piers. *Neural Computing and Applications* **31** (11), 7335–7349. https://doi.org/10.1007/s00521-018-3570-6.

Toth, E. & Brandimarte, L. 2011 Prediction of local scour depth at bridge piers under clear-water and live-bed conditions: comparison of literature formulae and artificial neural networks. *Journal of Hydroinformatics* **13** (4), 812–824.

Wang, C., Yu, X. & Liang, F. 2017 A review of bridge scour: mechanism, estimation, monitoring and countermeasures. *Natural Hazards* **87** (3), 1881–1906.

First received 14 January 2021; accepted in revised form 6 May 2021. Available online 3 June 2021

Downloaded from http://iwa.silverchair.com/jh/article-pdf/23/4/849/910421/jh0230849.pdf
by guest