


Prediction of Manning's coefficient of roughness for high-gradient streams using M5P

Parveen Sihag^a, Balraj Singh^b, Md. Azlin Bin Md. Said^c and H. Md. Azamathulla ^{d,*}

^a Department of Civil Engineering, Chandigarh University, Mohali, Punjab, India

^b Department of Civil Engineering, Panipat Institute of Engineering & Technology, Samalkha, India

^c Water Resources Engineering, School of Civil Engineering, Universiti Sains Malaysia, Engineering Campus, 14300 Nibong Tebal, Pulau Pinang, Malaysia

^d Civil & Environmental Engineering, The University of the West Indies, St. Augustine Campus, Port of Spain, Trinidad

*Corresponding author. E-mail: azmatheditor@gmail.com

 HMA, 0000-0002-5436-4147

ABSTRACT

The coefficient of Manning's roughness (n) has been generally implemented in the determination of depth and discharge in open channels and canals. This study unravels the novel idea and potential of Random Forest (RF), M5P, and Random Tree (RT) approaches to evaluate and predict the coefficient of Manning's roughness for hydraulic designing. To achieve this purpose, 42 observations were collected for high-gradient streams in Colorado, USA. All the observations were from boulder-bed, cobble and high gradient ($S > 0.002$ m/m) streams within bank flows. In order to ascertain the best model, the above-mentioned approaches were evaluated and compared using performance evaluation indices such as mean absolute error (MAE), coefficient of correlation (CC), and root mean square error (RMSE). Outcomes of performance evaluation indices revealed that the proposed pruned M5P approach outperformed other applied models for predicting the coefficient of Manning's roughness for hydraulic designing with $CC = 0.7858, 0.7910$, $RMSE = 0.0195, 0.0195$, and $MAE = 0.0157, 0.0165$ for model development and validation period, correspondingly. Furthermore, Taylor diagram and Box plot also suggest that the M5P based approach works better than RF and RT based approaches for predicting the coefficient of Manning's roughness for high-gradient streams using the given data set.

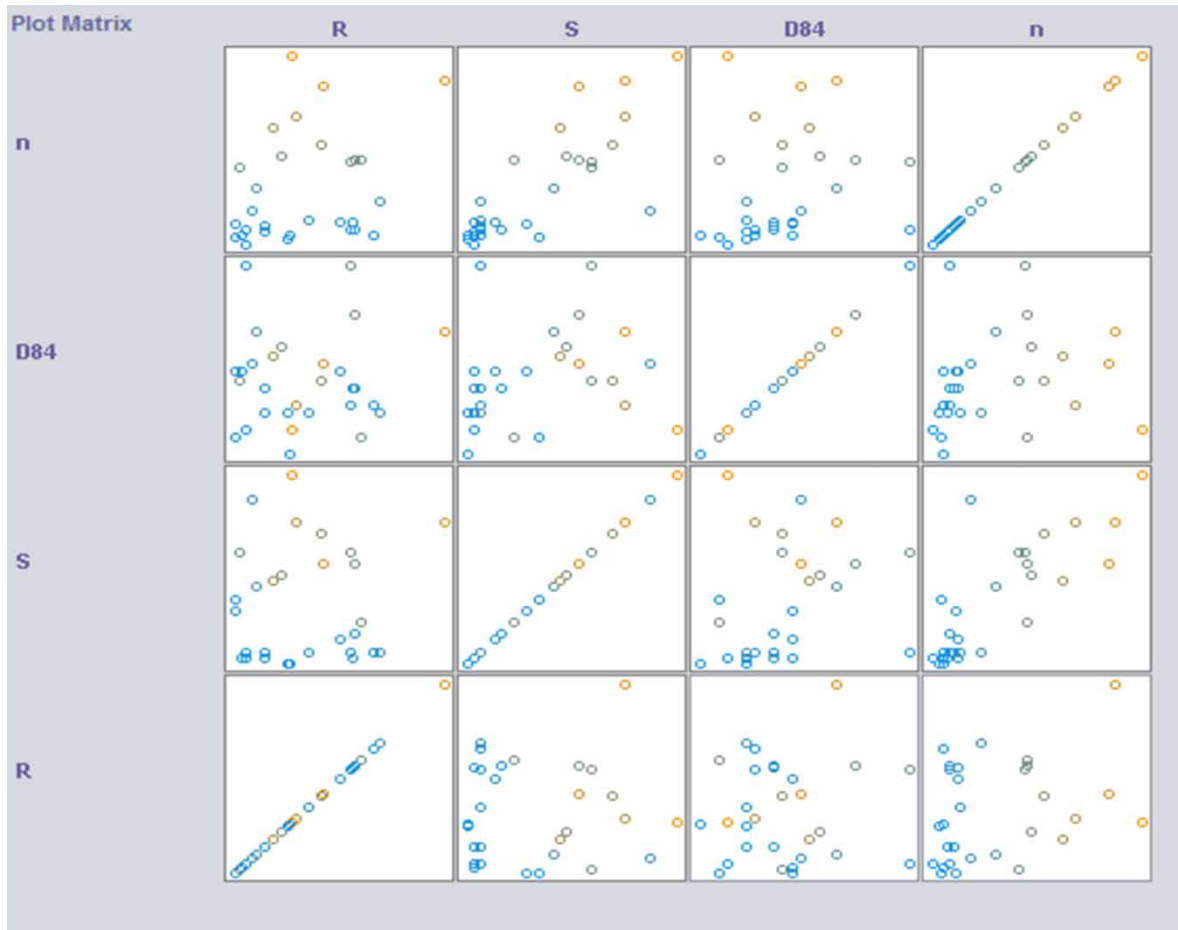
Key words: M5P, Manning's roughness coefficient, random forest, random tree, Taylor diagram

HIGHLIGHTS

- Three soft computing-based modelling approaches (M5P, RF and RT) were developed in the prediction of Manning's roughness coefficient.
- The performance of modelling approaches was compared by mean absolute error (MAE), coefficient of correlation (CC), and root mean square error (RMSE).
- The total dataset was divided into training and testing subset in the ratio of 70:30 to perform the modelling approaches.
- M5P modelling approach is the best approach in the prediction of the Manning's roughness coefficient.

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/>).

GRAPHICAL ABSTRACT



INTRODUCTION

Evaluation of irregular features is essential for hydraulic designing of overbank areas of floodplains and the flow in channels (Chow 1959) which is usually done by utilizing the coefficient of Manning's roughness where n denotes the flow resistance or relative roughness of floodplains and channels. The symbol n is taken from Manning's equation for OPC (open channel flow) which is as follows:

$$V = (R^{0.67} S^{0.5})/n \quad (1)$$

wherein, V stands for the velocity (average cross sectional) in m/s, R for hydraulic mean depth in m, S for friction slope or energy gradient in m/m, and n for the coefficient of Manning's roughness.

In the MKS system (SI unit), Equation (1) is divided by 1, which carries the unit of $m^{0.5}/s$. For consistency and simplicity purpose, the unit of the n is ignored. After substituting the values in the continuity equation, the parent equation becomes:

$$Q = A.V \quad (2)$$

$$Q = (AR^{0.67} S^{0.5})/n \quad (3)$$

where Q stands for discharge in m^3/s ; A is the area of cross-section in m^2 .

For the conditions of relative uniform flow, when Equations (1) and (3) were formulated, the energy gradient, friction slope, and water-surface slope are in the direction to the stream bed, and the hydraulic radius, area, and slope remain relatively

constant. However, the energy slopes, friction, and water may be different in many natural channels, especially in high-gradient streams. Therefore, several authors select such cases which are relatively uniform with gradually varied flow for the investigation of resistance in the channels of natural flow (e.g. Barnes 1967; Limerinos 1970; Jarrett 1984). More so, velocity distribution is assumed to be logarithmic in the channels of natural flow (Chow 1959). Jarrett (1992) conducted a study which revealed that in higher gradient channels vertical-velocity profiles were found to be S-shaped which can affect sediment transport and hydraulic interpretations (assuming the profile is logarithmic).

Despite the availability of equations, guidelines and data to help in the estimation of n , the most extensively used approach for the computing depths or discharges of flow in natural channels, for instance morpho dynamics, sediment transport, flood inundation mapping, and the riverine ecosystem studies is Manning's equation, thus one n value is chosen for the entire range of flow depth. Costa & Jarrett (2008), Jiang & Li (2010), and Wohl (2000) have succinctly remarked that there is still no exact method for determining the n values in natural rivers, therefore, the research based on physical observations with greater reliance on the data obtained from the field is required to decrease the uncertainties for estimating the n -values.

There are certain modelling approaches used in an effective manner in the field of scour estimation and n -values estimation, i.e. multilayer perceptron-induced with the firefly algorithm (MLP_FFA) (Diop *et al.* 2020; Roushangar *et al.* 2020), terrestrial-remote sensing technique and flood modeling using FLO-2D (Demir & Keskin 2020), ANFIS (adaptive neuro-fuzzy inference systems) (Azamathulla *et al.* 2009; Bahramifar *et al.* 2013; Moharana & Khatua 2014; Singh *et al.* 2020) and ANN (Azamathulla *et al.* 2008; Singh *et al.* 2020; Zounemat-Kermani *et al.* 2020). Also, it has been found that ANNs give reasonably better solutions for the issues of hydraulic and hydrology engineering, especially in cases of complex and nonlinear relationships between the pairs of output-input in corresponding data (Güven 2009; Azamathulla *et al.* 2010; Güven & Talu 2010; Azamathulla & Ghani 2011; Traore & Güven 2012; Singh *et al.* 2020; Zounemat-Kermani *et al.* 2020).

Several researchers have resorted usage of modelling approaches for analyzing the field and laboratory data, and have found significantly better results as compared to conventional statistical methods during the last few decades (Giustolisi 2004; Azmathullah *et al.* 2005; Azamathulla *et al.* 2010; Singh *et al.* 2018; Mohanty *et al.* 2019; Singh Nain *et al.* 2019; Sihag *et al.* 2020; Rani *et al.* 2021; Sihag *et al.* 2021). Several researchers have analyzed the data obtained downstream of hydraulic structures and scour around using ANN (Azmathullah *et al.* 2005; Azamathulla *et al.* 2008). Recently, M5P has attracted the attention of researchers in the prediction of hydraulic characteristics. The present study presents RF, RT and M5P as alternative tools for estimating Manning's n value(s).

M5 tree model

Model trees have been derived from the regression trees which have linear functions on their leaves (Witten *et al.* 2005). The M5P tree model was propounded by Quinlan (1992), which is a binary decision tree having a linear function (leaf) at the terminal nodes that can estimate continuous mathematical features. The M5 tree model is thus an amalgamation of the linear regression and tree regression method. The M5 tree model employs division of data into several categories through linear regression which is named leaf. This separation method introduced at each node enables better knowledge to be obtained with less deviation in the intra-subdivision class values in each branch. The separation criteria help in estimating the standard deviations of the class values extending to nodes which leads to the generation of the basic tree model. This approach generates linear functions at each node and utilizes the standard deviation approach by measuring the estimated error at the terminal node. The construction of the M5P model employs three stages: tree growth, pruning and smoothing. The chances of overfitting can be reduced through pruning. This method of constructing the model tree divides parameter space into subspaces and structure and includes a linear regression model in each of them (Quinlan 1992). Equation (4) represents the criteria for the standard deviation in each of the leaves (Sihag *et al.* 2018; Kumar & Sihag 2019; Sihag *et al.* 2019):

$$SDR = sd(T) - \sum \frac{|T_i|}{|T|} sd(T_i) \quad (4)$$

where SDR is the standard deviation reduction, T is the data set inputs into the tree branches, T_i is the data set in leaf, sd is the standard deviation. Figure 1 shows a schematic shape of the M5 tree model development. In the Figure 1(a), the X_1 and X_2 are the input variables (independent parameters) and Y is the output data (dependent parameter). Figure 1(b) shows the tree model development for mapping the input and output data. Four linear models were developed at the leaf nodes in Figure 1. These linear models are based on input variables conditions such as greater than or equal to or less than.

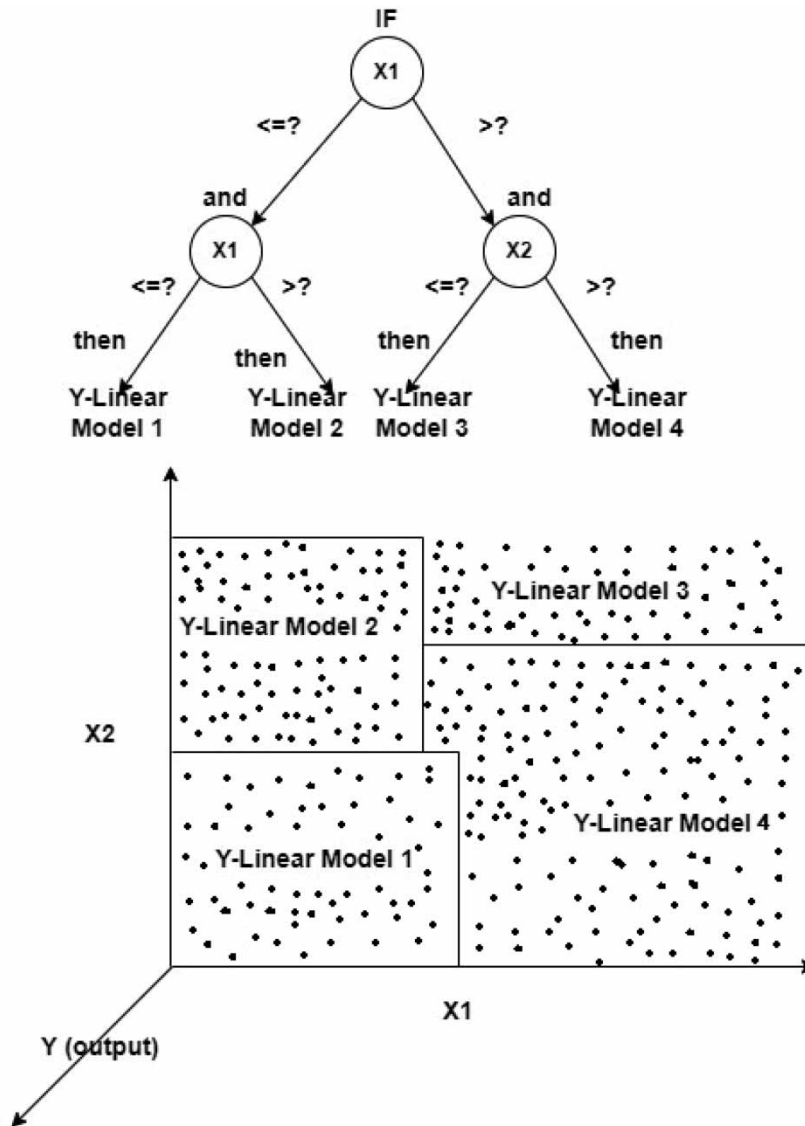


Figure 1 | The visualization of the M5 tree model developed by Etemad-Shahidi & Mahjoobi (2009).

Random forest (RF)

The random forest was introduced by Breiman (1996), which is an adaptable assembly of decision trees that seamlessly manages variance and bias for linear and nonlinear predictions. This approach is versatile and has been preferred to solve various nonlinear or complex engineering issues. This technique employs creation of a large number of trees with the root node achieving a different bootstrap (bagging) sample of the original data set (Breiman 1996, 2001). Using a randomly chosen subset of the parameters of the estimator at each node, division is performed that results in a single generic vote (Liaw & Wiener 2002) in which each tree undergoes a distinct prediction cycle. This model has two different standard frameworks: the number of input variables (q) selected to create a tree at each node, and the number of trees to be generated (k). For the best heterogeneity, only specific variables represent 'whole' at each node. So, the regression of the random forest includes k trees. The random forest predictor considers a form by taking over k trees joint of general error. The random forest algorithm is simple, comparatively oblivious to the training set features, and capable of achieving high precision in prediction (Breiman 1999, 2001). For the development of the model, a trial-and-error process is employed. The WEKA 3.9 software was used to develop the random forest-based model in this current investigation.

Random tree (RT)

The random tree algorithm is used to assess a particular number of random characteristics at every node without pruning. Random tree has little to do with machine learning however it uses arbitrary knowledge, particularly bagging (Hamoud *et al.* 2018). Every node in a random forest is finely divided into the arbitrarily chosen forerunner subsets of that node. It deals with both problems with classification and regression. Random trees are a set of forests known as tree estimators. The classification of RT is as follows: the vector input property is taken by the classifier of random trees, it classifies each tree in the forest, and extracts the category mark which receives maximum votes. Average response is the response of the classifier of all the trees in the forest in the denial period (Cutler *et al.* 2012). RTs are basically an amalgamation of two algorithms existing in machine learning: RF principles and single model trees. Model trees are decision trees which represent the linear pattern on the basis of which each leaf is designed for the local subdomain that this leaf represents. The performance of single stable trees has been shown to significantly improve RFs. Basically there are two methods that create tree diversity: first, by removing each tree, the training data is sampled, as in bagging; second, instead of always calculating the best possible division for each node when growing a tree, only one random subset of all attributes is considered for each node, and the best part of that subset is determined. For the first time, Random Model trees merge random forests and model trees. RTs employ this result for dividing criteria and thus encourage considerable balanced trees where a spherical ridge environment runs on all leaves, thus simplifying the optimization method (Barddal *et al.* 2019).

Performance evaluation indices

To check and compare the performance of modelling approaches, three performance evaluation indices, including CC, MAE and RMSE, were chosen to assess the correctness of the developed approaches. The equations of these performance evaluation indices are as follows:

$$CC = \frac{\sum_{i=1}^N (P_i - \bar{P})(R_i - \bar{R})}{\sqrt{\sum_{i=1}^N (P_i - \bar{P})^2 \sum_{i=1}^N (R_i - \bar{R})^2}} \quad (5)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (P_i - R_i)^2} \quad (6)$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |P_i - R_i| \quad (7)$$

where P_i are the actual values and R_i are the predicted values of manning coefficient, while \bar{P} and \bar{R} are the mean of actual and predicted values, and N are the number of observations.

Data set

A total of 42 observations were used for this investigation. The complete data set was separated into two portions; the first portion is 70% of total data and the remainder is the second portion (Bhoria *et al.* 2021). The first portion (29 observations) was used for developing the model and the second portion (13 observations) was for validation. The range, features and statistical description of both portions are shown in Table 1. R, S and D84 are independent variables so these are selected as input variable whereas n is selected as output in the model development and validation stages. Figure 2 shows the correlation plot of all input and output variables.

RESULTS AND DISCUSSION

For the accurate prediction of the coefficient of Manning's roughness for features for hydraulic designing, soft computing and regression-based modelling approaches are used in the current investigation. The test performance of all the implemented models was carried out by utilizing three standard statistical parameters, CC, RMSE and MAE. The lower values of MAE and RMSE values show higher model accuracy, and higher CC values show higher model accuracy. The range of CC is -1 to 1. WEKA 3.9 software was implemented for the model development and validation in this study. A trial and error

Table 1 | Range and features of the observed data

Statistic	First portion				Second portion			
	R	S	D84	n	R	S	D84	n
Minimum	0.1500	0.0020	0.0910	0.0280	0.1800	0.0020	0.0910	0.0300
Maximum	1.6800	0.0340	0.7920	0.1590	1.2300	0.0310	0.6100	0.1030
Mean	0.6507	0.0122	0.3647	0.0667	0.6015	0.0102	0.3040	0.0528
Standard deviation	0.3948	0.0097	0.1737	0.0378	0.3711	0.0096	0.1498	0.0208
Kurtosis	-0.1845	-0.7745	0.8270	0.0211	-0.8385	0.6385	-0.1525	1.5271
Skewness	0.5913	0.6759	0.9192	1.0281	0.7582	1.3583	0.6203	1.2408
Confidence level (95%)	0.1502	0.0037	0.0661	0.0144	0.2242	0.0058	0.0905	0.0126



Figure 2 | Correlation plot using observation.

method was employed for model preparation. The ideal values of the first parameters was achieved after a number of tests. There are well-defined statistical criteria for selecting and defining first parameters that are unique to the model.

Assessment of M5P tree based modelling approach

The M5P modelling approach is developed by a trial and error process. This model utilises linear regression models for defining an input-output relationship which is based on division of the parameter space of the data into several subspaces. In this investigation pruned and unpruned both models were developed for prediction of the coefficient of Manning’s roughness for

hydraulic designing. Developed linear equations using pruned and unpruned M5P based models are listed in Tables 2 and 3 respectively.

Results of the M5P model to predict n is shown in Figure 3 for model development as well as the validation stage. Results of the performance evaluation indices concludes that the pruned M5P based modelling approach is more accurate than the unpruned M5P based modelling approach for predicting n for hydraulic designing with $CC = 0.7858, 0.7910, RMSE = 0.0195, 0.0195, \text{ and } MAE = 0.0157, 0.0165$ for model development and validation period, correspondingly. Overall, assessing Figure 3 and Table 4 (performance evaluation parameters) suggests that both pruned and unpruned M5P based modeling approaches are suitable for predicting n for hydraulic designing.

Assessment of RF based modelling approach

Figure 4 provide plots of agreement between actual and predicted Manning’s values of n for hydraulic designing through an RF based modelling approach for model development and validation stages, respectively. Predicted values from the RF based modelling approach lies very close to the agreement line. Table 4 shows the results of model development and validation stages in terms of $CC, RMSE$ and MAE which indicates that the performance of the RF approach is suitable for the prediction of n for hydraulic designing with $CC, RMSE$ and MAE values are $0.9706, 0.0096, \text{ and } 0.0065$ for model development stage and $0.7581, 0.0203 \text{ and } 0.0165$ for validation stage respectively.

Assessment of RT based modelling approach

The random tree based model development process is similar to M5P and RF based models. This model is also developed using WEKA 3.9. Figure 5 provide plots of agreement between predicted and actual values of n for hydraulic designing through RT based modeling approach for model development and validation stages, respectively. Predicted values from

Table 2 | Linear equation for M5 pruned modelling approach

LM num	Equation
1	$n = 0.0258 * R + 3.1029 * S + 0.012$

Table 3 | Linear equation for M5 unpruned modeling approach

M5 unpruned model tree: (using smoothed linear models)	LM num	Equations
S <= 0.008 :	1	$n = 0.0197 * R + 2.3019 * S + 0.0188$
R <= 0.615 :	2	$n = 0.0197 * R + 2.3019 * S + 0.0188$
S <= 0.004 :	3	$n = 0.0202 * R + 2.4249 * S + 0.0184$
R <= 0.45 : LM1 (3/10.087%)	4	$n = 0.017 * R + 1.6049 * S + 0.0234$
R > 0.45 : LM2 (2/2.696%)	5	$n = 0.0157 * R + 1.6049 * S + 0.0244$
S > 0.004 : LM3 (2/4.044%)	6	$n = 0.0156 * R + 1.6049 * S + 0.0245$
R > 0.615 :	7	$n = 0.0286 * R + 2.4454 * S + 0.0196$
D84 <= 0.259 : LM4 (2/17.522%)	8	$n = 0.0286 * R + 2.4454 * S + 0.0201$
D84 > 0.259 :	9	$n = 0.0254 * R + 2.6391 * S + 0.0253$
R <= 1.01 : LM5 (3/6.354%)	10	$n = 0.0254 * R + 2.6391 * S + 0.0253$
R > 1.01 : LM6 (2/5.391%) S > 0.008 :	11	$n = 0.0254 * R + 2.6391 * S + 0.025$
R <= 0.37 :	12	$n = 0.0254 * R + 2.7219 * S + 0.0251$
R <= 0.165 : LM7 (2/12.131%)		
R > 0.165 : LM8 (3/33.016%)		
R > 0.37 :		
S <= 0.025 :		
R <= 0.89 :		
R <= 0.625 : LM9 (2/26.957%)		
R > 0.625 : LM10 (2/53.915%)		
R > 0.89 : LM11 (3/1.271%)		
S > 0.025 : LM12 (3/46.501%)		

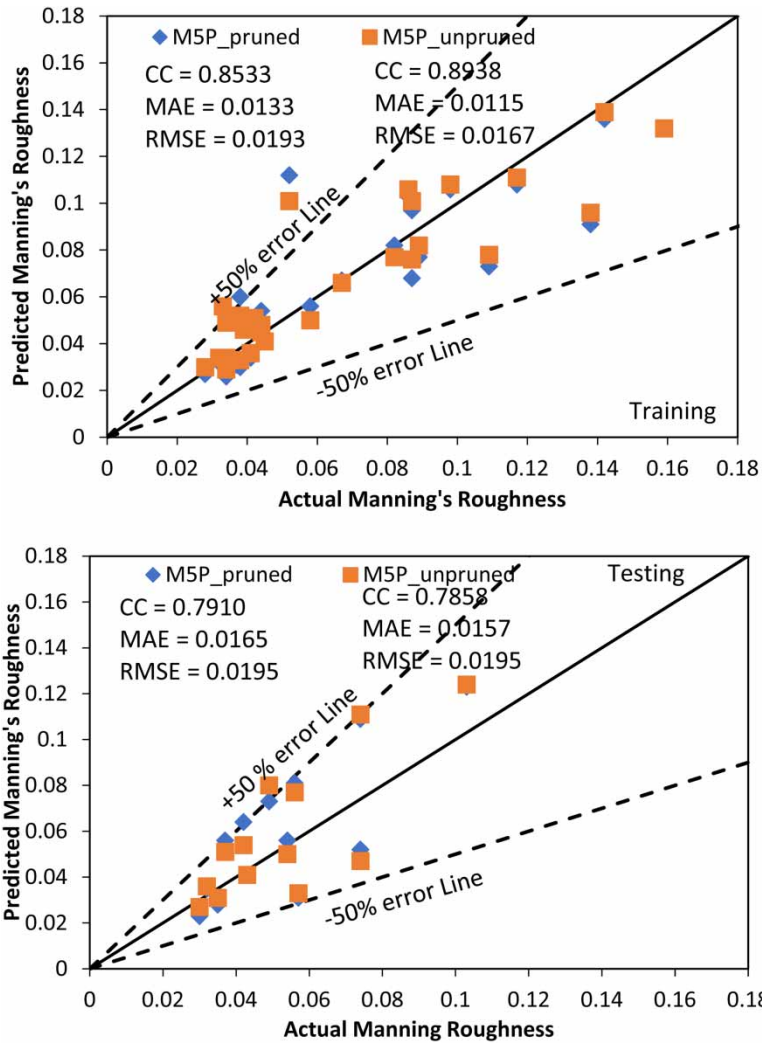


Figure 3 | Agreement plot among actual and predicted values of n using M5P based modelling approaches for model development and validation stages.

Table 4 | Calculated performance evaluation indices for soft computing based applied modelling approaches using model development and validation stages

Models	First portion (Model development)			Second portion (Validation stage)		
	CC	MAE	RMSE	CC	MAE	RMSE
M5P_pruned	0.8533	0.0133	0.0193	0.7910	0.0165	0.0195
M5P_unpruned	0.8938	0.0115	0.0167	0.7858	0.0157	0.0195
RF	0.9706	0.0064	0.0096	0.7581	0.0165	0.0203
RT	1.0000	0.0001	0.0004	0.7755	0.0221	0.0281

RF based modelling approach lies very nearer to the agreement line. Performance evaluation indices values for the test data set indicates that the performance of the RT approach is suitable for the prediction of values of n for hydraulic designing with CC, RMSE, and MAE values are 1.000, 0.0004, 0.0001 for the training stage and 0.7755, 0.0281, and 0.221 for the testing stage respectively.

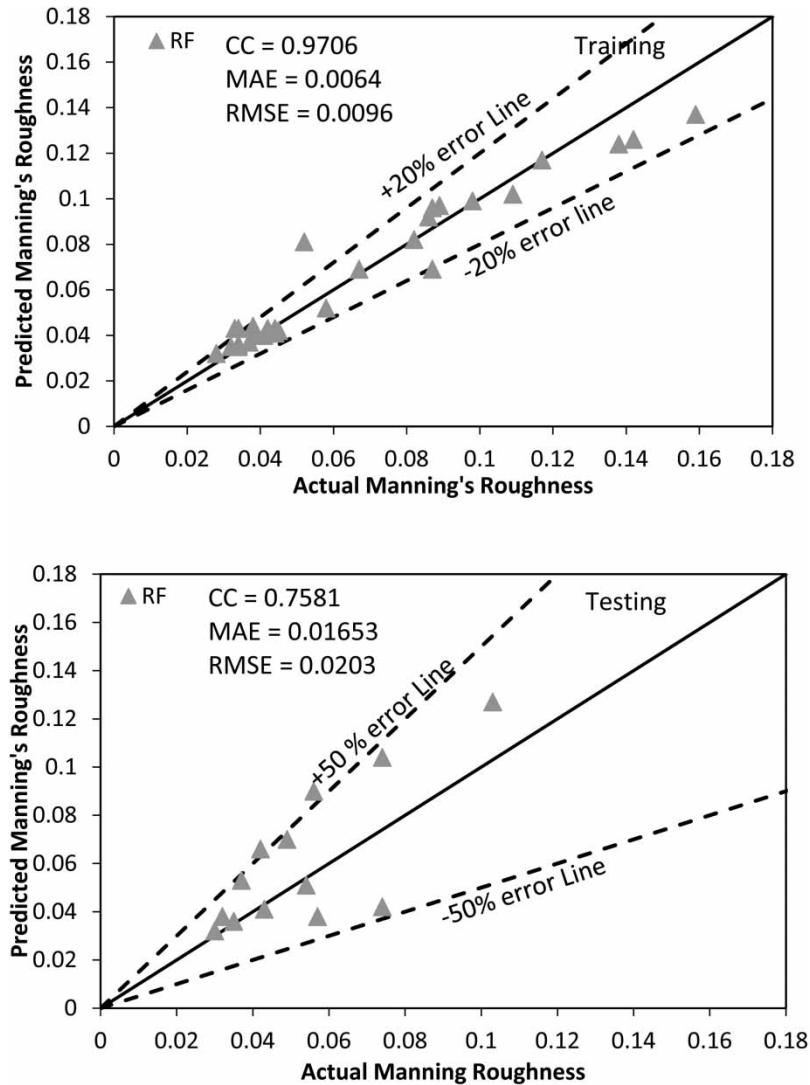


Figure 4 | Agreement plot among actual and predicted Manning's roughness coefficient values using RF based modelling approach for model development and validation stages.

Comparative assessment of soft computing based applied modelling approaches

Comparison of soft computing-based modelling approaches (Table 4 and Figure 6) shows that M5P based modelling approaches work better than other applied modelling approaches. To assess the potential of soft computing-based modeling approaches for predicting the values of n for hydraulic designing, agreement and performance graphs are plotted in Figure 6 for model development and validation stages. It is incidental from the plots that the predicted values produced by M5P based modelling approaches were in extremely near closeness to the actual values of n for hydraulic designing and predicted n for hydraulic designing are found to follow a similar pattern to that of actual values. The pruned M5P based modelling approach works better than unpruned M5P. The RF based modelling approach works better than the RT based modelling approach using this data set for the prediction the values of n for hydraulic designing. A box plot is also plotted in Figure 7 for the comparison of actual and predicted values using various applied designs for validation stage. Descriptive statistics of actual and applied models using validation stage are enlisted in Table 5. Figure 7 and Table 5 suggest that the pruned M5P modelling approach is outperforming in comparison to other applied modelling approaches. Minimum and maximum values of actual and predicted values using pruned M5P model is very close. In Figure 7, widths of the lower and upper quartile are almost the same which recommends that the pruned M5P modelling approach is most suitable for the prediction the values of n for hydraulic designing.

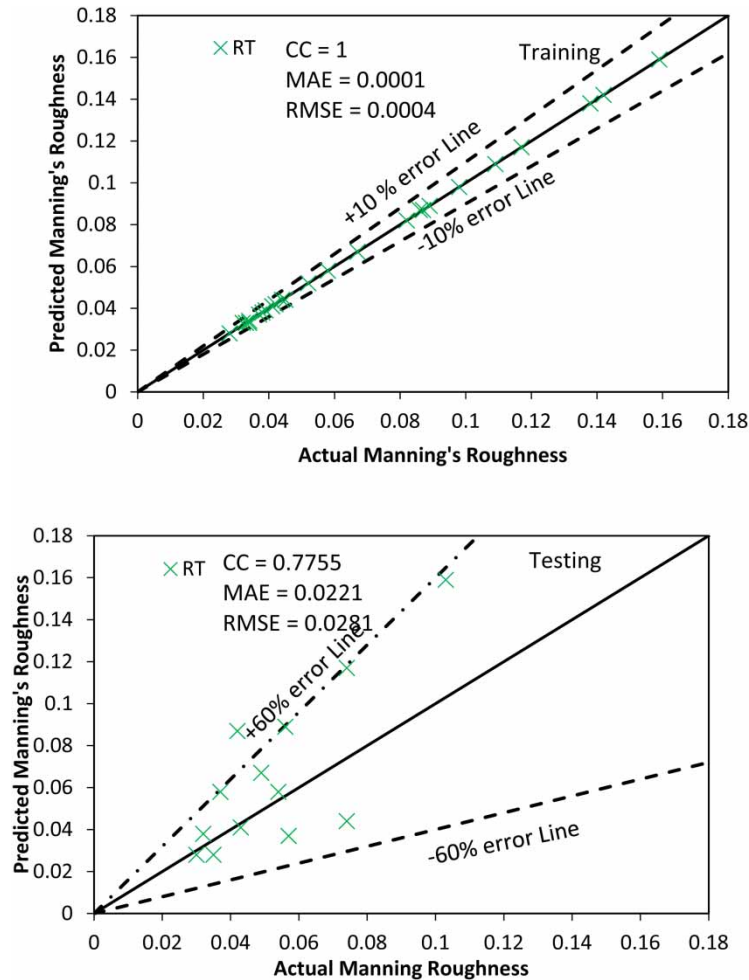


Figure 5 | Agreement plot among actual and predicted Manning's roughness coefficient values using the RT based modelling approach for model development and validation stages.

The Taylor diagram is a graphical illustration of the performance of developed models in terms of correlation, RMSE and standard deviation is shown in [Figure 8](#). [Figure 8](#) indicates that pruned M5P is the best performing model and the performance of the RT model is least in the prediction of the coefficient of Manning's roughness for hydraulic designing. Furthermore, a comparison of the obtained results was carried out with the previous published studies ([Jarrett 1984](#); [Azamathulla & Jarrett 2012](#)). This comparison is done on the basis of the obtained values of CC and RMSE. The details of the comparison are given in [Figure 9](#) which suggests that the results obtained from the M5P are much better than [Jarrett \(1984\)](#) and [Azamathulla & Jarrett \(2012\)](#). The values of CC and RMSE for the M5P_purned are 0.791 and 0.0195 which is better than M5P_unpurned (0.785 and 0.0195); [Azamathulla & Jarrett \(2012\)](#) (0.745 and 0.978) and [Jarrett \(1984\)](#) (0.58 and 1.205). Thus, the M5P model is a general model for the study area and gives closer values to the observed values as compared to other models as well as the previous studies. The M5P model is used to represent the coefficient of Manning's roughness of the study area and it saves time and effort in comparison with experimentation and conventional models.

CONCLUSION

In this study, M5P, RT and RF based modeling approaches have been developed for the prediction of the coefficient of Manning's roughness for hydraulic designing. The comparison analysis using performance evaluation indices conclude that the pruned M5P approach out-performed the RF and RT based modelling approaches using given data set with $CC = 0.7858$,

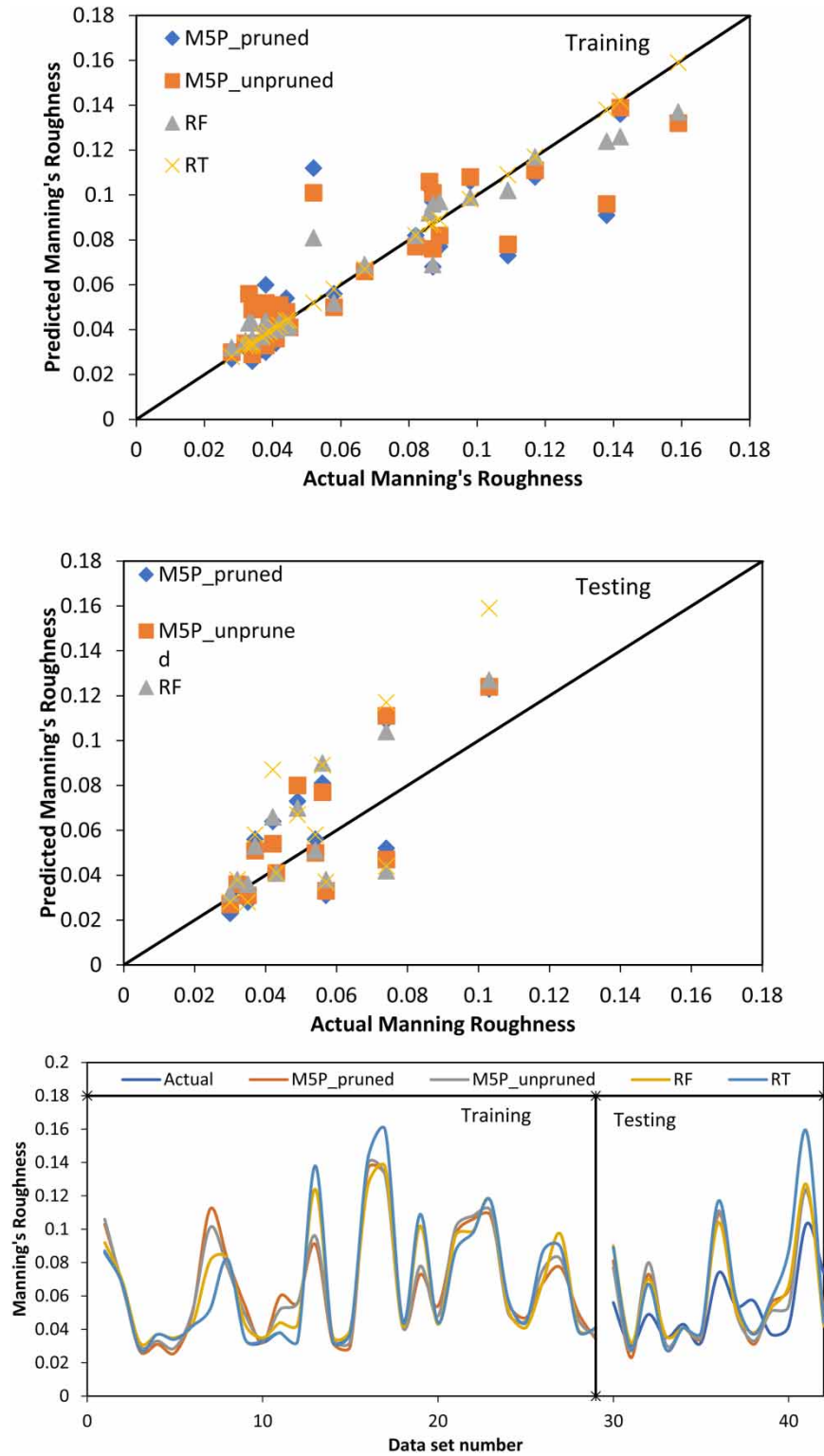


Figure 6 | Agreement and performance plot for actual and predicted Manning's roughness coefficient values using various soft computing based applied modelling approaches for model development and validation stages.

0.7910, RMSE = 0.0195, 0.0195, and MAE = 0.0157, 0.0165 for model development and validation period, correspondingly. The training and testing results of the M5P model encourage the utility of this method relative to the other tested methods and show a good potential in representing the Manning coefficient of the study area. In M5P based models, the pruned model

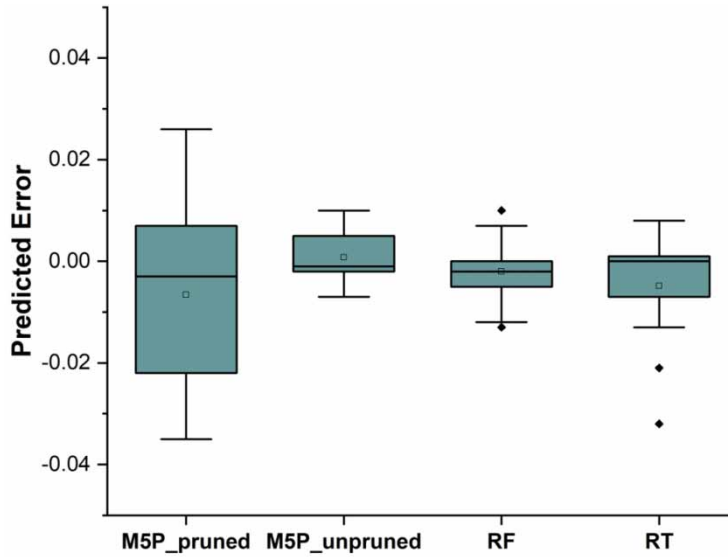


Figure 7 | Box plot of errors distribution using various soft computing based applied modelling approaches for validation stage.

Table 5 | Descriptive statistics of error distribution using various soft computing based applied modelling approaches for validation stage

Statistic	M5P_pruned	M5P_unpruned	RF	RT
Minimum	-0.0350	-0.0370	-0.0340	-0.0560
Maximum	0.0260	0.0270	0.0320	0.0300
1st quartile	-0.0220	-0.0210	-0.0240	-0.0330
Median	-0.0030	-0.0040	-0.0060	-0.0060
3rd quartile	0.0070	0.0040	0.0020	0.0020
Mean	-0.0066	-0.0058	-0.0078	-0.0127

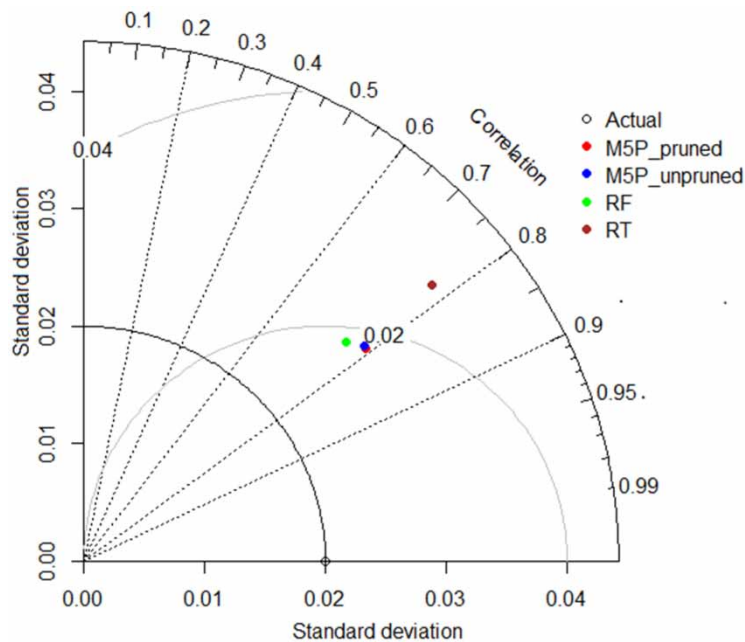


Figure 8 | Taylor diagram for n values using various soft computing based applied modeling approaches for validation stage.

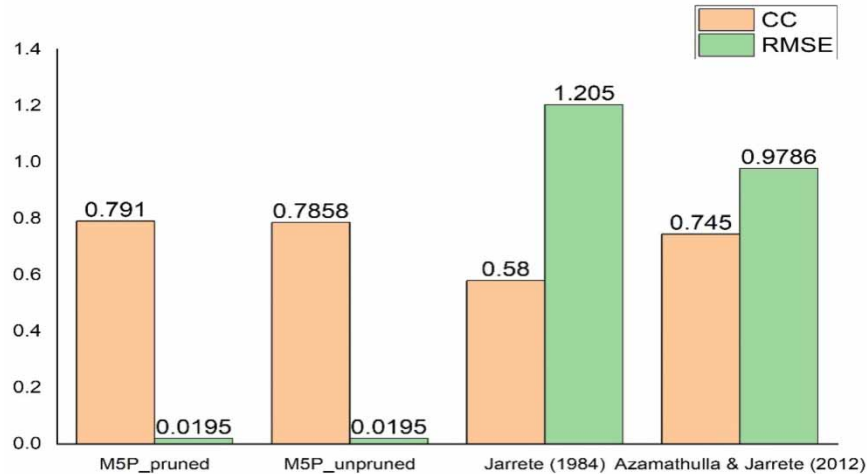


Figure 9 | Comparison of the results with previous studies.

works better than the unpruned model. The Taylor diagram and box plot also suggest that M5P based modelling approaches work better than RT and RF based modelling approaches for predicting the coefficient of Manning's roughness for hydraulic designing using this data set.

DATA AVAILABILITY STATEMENT

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

REFERENCES

- Azamathulla, H. M. & Ghani, A. A. 2011 Genetic programming for predicting longitudinal dispersion coefficients in streams. *Water Resources Management* **25** (6), 1537–1544.
- Azamathulla, H. M. & Jarrett, R. D. 2012 Use of gene-expression programming to estimate Manning's roughness coefficient for high gradient streams. *Water Resources Management* **27** (3), 715–729.
- Azamathullah, 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., Deo, M. C. & Deolalikar, P. B. 2008 Alternative neural networks to estimate the scour below spillways. *Advances in Engineering Software* **39** (8), 689–698.
- Azamathulla, H. M., Chang, C. K., Ghani, A. A., Ariffin, J., Zakaria, N. A. & Hasan, Z. A. 2009 An ANFIS-based approach for predicting the bed load for moderately sized rivers. *Journal of Hydro-Environment Research* **3** (1), 35–44.
- Azamathulla, H. M., Ab Ghani, A., Zakaria, N. A. & Guven, A. 2010 Genetic programming to predict bridge pier scour. *Journal of Hydraulic Engineering* **136** (3), 165–169.
- Bahramifar, A., Shirkhani, R. & Mohammadi, M. 2013 An anfis-based approach for predicting the manning roughness coefficient in alluvial channels at the bank-full stage. *International Journal of Engineering* **26** (2), 177–186.
- Barddal, J. P., Enembreck, F., Gomes, H. M., Bifet, A. & Pfahringer, B. 2019 Boosting decision stumps for dynamic feature selection on data streams. *Information Systems* **85**, 13–29.
- Barnes, H. H. 1967 *Roughness Characteristics of Natural Channels*. No. 1849. US Government Printing Office.
- Bhoria, S., Sihag, P., Singh, B., Ebtehaj, I. & Bonakdari, H. 2021 Evaluating Parshall flume aeration with experimental observations and advance soft computing techniques. *Neural Computing and Applications* **33** (24), 17257–17271.
- Breiman, L. 1996 Bagging predictors. *Machine Learning* **24** (2), 123–140.
- Breiman, L. 1999 Random forests. UC Berkeley TR567.
- Breiman, L. 2001 Random forests. *Machine Learning* **45** (1), 5–32.
- Chow, V. T. 1959 *Open-channel Hydraulics*. McGraw-Hill, New York.
- Costa, J. E. & Jarrett, R. D. 2008 *An Evaluation of Selected Extraordinary Floods in the United States Reported by the US Geological Survey and Implications for Future Advancement of Flood Science*. Scientific Investigations Report. U. S. Geological Survey.
- Cutler, A., Richard Cutler, D. & Stevens, J. R. 2012 Random forests. In: Cutler, A., Cutler, D. R., Stevens, J. R. & Zhang, C. (eds). *Ensemble Machine Learning*. Springer, Boston, MA, pp. 157–175.

- Demir, V. & Keskin, A. 2020 Obtaining the Manning roughness with terrestrial-remote sensing technique and flood modeling using FLO-2D: a case study Samsun from Turkey. *G Eofizika* **37** (2), 131–156.
- Etemad-Shahidi, A. & Mahjoobi, J. 2009 Comparison between M5' model tree and neural networks for prediction of significant wave height in Lake Superior. *Ocean Engineering* **36** (15–16), 1175–1181.
- Diop, L., Samadianfard, S., Bodian, A., Yaseen, Z. M., Ghorbani, M. A. & Salimi, H. 2020 Annual rainfall forecasting using hybrid artificial intelligence model: integration of multilayer perceptron with whale optimization algorithm. *Water Resources Management* **34** (2), 733–746.
- Giustolisi, O. 2004 Using genetic programming to determine Chezy resistance coefficient in corrugated channels. *Journal of Hydroinformatics* **6** (3), 157–173.
- Güven, A. 2009 Linear genetic programming for time-series modelling of daily flow rate. *Journal of Earth System Science* **118** (2), 137–146.
- Güven, A. & Talu, N. E. 2010 Gene expression programming for estimating suspended sediment yield in Middle Euphrates Basin, Turkey. *CLEAN–Soil, Air, Water* **38** (12), 1159–1168.
- Hamoud, A., Hashim, A. S. & Awadh, W. A. 2018 Predicting student performance in higher education institutions using decision tree analysis. *International Journal of Interactive Multimedia and Artificial Intelligence* **2018** (5), 26–31. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3243704
- Jarrett, R. D. 1984 Hydraulics of high-gradient streams. *Journal of Hydraulic Engineering* **110** (11), 1519–1539.
- Jarrett, R. D. 1992 Hydraulics of mountain rivers. In: Yen, B. C. (ed.). *Channel Flow Resistance: Centennial of Manning's Formula*. Water Resource Publications, CO, USA, pp. 287–298.
- Jiang, M. & Li, L.-X. 2010 An improved two-point velocity method for estimating the roughness coefficient of natural channels. *Physics and Chemistry of the Earth. Parts A/B/C* **35** (3–5), 182–186.
- Kumar, M. & Sihag, P. 2019 Assessment of infiltration rate of soil using empirical and machine learning based models. *Irrigation and Drainage* **68** (3), 588–601.
- Liaw, A. & Wiener, M. 2002 Classification and regression by randomForest. *R News* **2** (3), 18–22.
- Limerinos, J. T. 1970 *Determination of the Manning Coefficient from Measured Bed Roughness in Natural Channels*. Government Printing Office, Washington, DC, USA.
- Mohanty, S., Roy, N., Singh, S. P. & Sihag, P. 2019 Estimating the strength of stabilized dispersive soil with cement clinker and fly ash. *Geotechnical and Geological Engineering* **37** (4), 2915–2926.
- Moharana, S. & Khatua, K. K. 2014 Prediction of roughness coefficient of a meandering open channel flow using Neuro-Fuzzy Inference System. *Measurement* **51**, 112–123.
- Quinlan, J. R. 1992 Learning with continuous classes. In: *5th Australian Joint Conference on Artificial Intelligence*, Vol. 92. World Scientific Publishing Co Pvt Ltd., Hobart, pp. 343–348.
- Rani, K., Suthar, M., Sihag, P. & Boora, A. 2021 Experimental investigation and prediction of strength development of GGBFS, LFS- and SCBA-based Green concrete using soft computing techniques. *Arabian Journal of Geosciences* **14**, 2612. <https://doi.org/10.1007/s12517-021-08869-4>.
- Roushangar, K., Saghebani, S. M., Kirca, V. O. & Ghasempour, R. 2020 Prediction of form roughness coefficient in alluvial channels using efficient hybrid approaches. *Soft Computing* **24** (24), 18531–18543.
- Sihag, P., Jain, P. & Kumar, M. 2018 Modelling of impact of water quality on recharging rate of storm water filter system using various kernel function based regression. *Modeling Earth Systems and Environment* **4** (1), 61–68.
- Sihag, P., Tiwari, N. K. & Ranjan, S. 2019 Prediction of unsaturated hydraulic conductivity using adaptive neuro-fuzzy inference system (ANFIS). *ISH Journal of Hydraulic Engineering* **25** (2), 132–142.
- Sihag, P., Kumar, M. & Singh, B. 2020 Assessment of infiltration models developed using soft computing techniques. *Geology, Ecology, and Landscapes* **5** (4), 241–251.
- Sihag, P., Pandhiani, S. M., Sangwan, V., Kumar, M. & Angelaki, A. 2021 Estimation of ground-level O₃ using soft computing techniques: case study of Amritsar, Punjab State, India. *International Journal of Environmental Science and Technology* **18** (6), 1–8.
- Singh, B., Sihag, P. & Singh, K. 2018 Comparison of infiltration models in NIT Kurukshetra campus. *Applied Water Science* **8** (2), 1–8.
- Singh, N. K., Singh, Y., Kumar, S. & Sharma, A. 2020 Predictive analysis of surface roughness in EDM using semi-empirical, ANN and ANFIS techniques: a comparative study. *Materials Today: Proceedings* **25**, 735–741.
- Singh Nain, S., Sai, R., Sihag, P., Vambol, S. & Vambol, V. 2019 Use of machine learning algorithm for the better prediction of SR peculiarities of WEDM of Nimonic-90 superalloy. *Archives of Materials Science and Engineering* **1** (95), 12–19.
- Traore, S. & Güven, A. 2012 Regional-specific numerical models of evapotranspiration using gene-expression programming interface in Sahel. *Water Resources Management* **26** (15), 4367–4380.
- Witten, I. H., Frank, E., Hall, M. A. & Pal, C. J. 2005 Practical machine learning tools and techniques. *Morgan Kaufmann* **578**, 1.
- Wohl, E. E. 2000 Mountain rivers. *American Geophysical Union* **14** (1), 1–320.
- Zounemat-Kermani, M., Fadaee, M., Adarsh, S. & Hinkelmann, R. 2020 Predicting Sediment transport in sewers using integrative harmony search-ANN model and factor analysis. *IOP Conference Series: Earth and Environmental Science* **491** (1), 012004. IOP Publishing.

First received 30 May 2021; accepted in revised form 8 December 2021. Available online 20 December 2021