ACS OMEGA

Open Access

This article is licensed under CC-BY-NC-ND 4.0

http://pubs.acs.org/journal/acsodf Article

Decision Tree-Based Modeling of the Aeration Effectiveness of Circular Plunging Jets

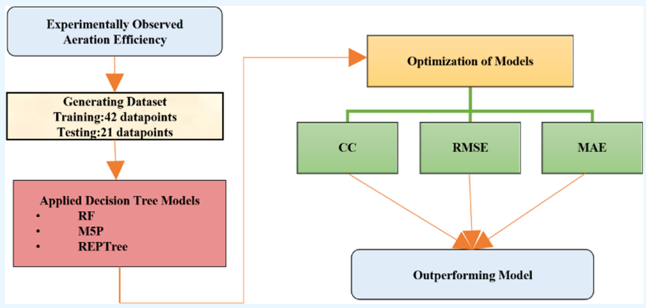
Diksha Puri, Daeho Lee, Dhananjay Vasant khankal, Mohindra Singh Thakur, Faisal M. Alfaisal, Shamshad Alam, Raj Kumar,* and Mohammad Amir Khan*

Cite This: *ACS Omega* 2023, 8, 38950−38960

Read Online

ACCESS | Metrics & More | Article Recommendations

ABSTRACT: Since soft computing has gained a lot of attention in hydrological studies, this study focuses on predicting aeration efficiency (E_{20}) using circular plunging jets employing soft computing techniques such as reduced error pruning tree (REPTree), random forest (RF), and M5P. The study undertaken required the development and validation of models, which were achieved using 63 experimental data values with input variables, such as angle of inclination of tilt channel (α), number of plunging jets (J_N), discharge of each jet (Q), hydraulic radius of each jet (HR), and Froude number (Fr. No), to evaluate the aeration efficiency (E_{20}), which served as the output variable. To evaluate the effectiveness of the developed models, three different statistical indices were used such as the coefficient of correlation (CC), root-mean-square error (RMSE), and mean absolute error (MAE), and it was found that all of the applied techniques possessed good forecasting ability since their correlation coefficient values were greater than 0.8. Upon testing, it was discovered that the M5P model outperformed other soft computing-based models in its ability to predict E_{20}, as demonstrated by its correlation coefficient value of 0.9564 and notably low values of MAE (0.0143) and RMSE (0.0193).



1. INTRODUCTION

The use of plunging jets as aeration devices is highly preferred due to their minimal energy requirements. Plunging jets do not need compressed air stirring machinery or consideration regarding obstruction or physical installation restrictions, in contrast to commercially utilized aerators, which makes them highly advantageous for many applications. For instance, the utilization of plunging jet aerators has become prevalent in industries such as chemical manufacturing, fermentation, and wastewater treatment owing to their numerous benefits. Therefore, it is evident that plunging jet aerators offer a highly effective and efficient method for aeration, which is why they have gained widespread adoption in various fields.

Numerous studies have been carried out on the subject of air−water−oxygen transfer through the utilization of plunging jets. Experimental studies were thoroughly examined and summarized in a detailed review by Bin and Smith.[1] Various experiments have been conducted on the oxygen transfer achieved through plunging water jets.[2−5] It is clear that there is considerable interest in the topic of air−water−oxygen transfer through the employment of plunging jets, and as such, it is a topic that continues to get a lot of interest from academics in the field.

The impact of artificial intelligence on various scientific disciplines, including civil engineering, has been noteworthy. Researchers have been able to approach complex systems in innovative and diverse ways with the advent of new algorithms and models, especially those based on soft computing. These advances have unlocked fresh possibilities and approaches in civil engineering applications, thereby making it a stimulating and continuously evolving field to explore. The model tree approach such as M5P has been thoroughly studied and applied in different hydraulic applications, with the effectiveness of the method demonstrated in several studies.[6−8] Additionally, RF, another tree-based regression approach, has been found effective in various fields.[9−13] These studies provide compelling evidence of the efficacy of tree-based regression approaches in a range of practical applications. Thus, it is important to consider these approaches as potential solutions in various fields, where prediction and forecasting are essential. Through the use of LS-SVM, researchers successfully predicted the air entrainment rate and E_{20} of diving over fall jets from weirs.[14] Prediction abilities of SVM regression and GP regression were compared while modeling oxygen transport by several plunging jets.[15] The

Received: May 15, 2023
Accepted: September 13, 2023
Published: October 9, 2023

© 2023 The Authors. Published by American Chemical Society

https://doi.org/10.1021/acsomega.3c03375
ACS Omega 2023, 8, 38950−38960

Table 1. Review of the Literature on Soft Computing Methods/Techniques for Forecasting Jet Aeration

authors	jet device and configuration	machine learning/soft computing technique (s) used	review
Kumar et al.[17]	hollow jet aerator with jet angles of 30°, 45°, and 60°.	SVM, M5P, MNLR.	SVM is found to be better than M5P and MNLR.
Kumar et al.[18]	plunging hollow jet aerator.	MNLR, SVM, GP.	the study suggests SVM with RBF kernel performs well and predicts the experimental K_La_{20} nearly within a 20% scatter.
Singh et al.[19]	plunging jets from nozzles with circular, square, rectangular, and rectangular with rounded edge.	NN, SVM, GP.	SVM was found to be the outperforming model.
Bodana et al.[20]	various geometries of hollow jets	ANN, GP, MNLR.	GP_RBF and GP_PUK were found to be the exceeding models for forecasting the B_{PD} entrained by the falling hollow jets.
Onen[21]	hollow plunging jets from nozzle.	ANN, GEP, MLR, MNLR.	GEP model gives better prediction results.
Kumar et al.[22]	plunging hollow jet aerator having configuration plain, curved, and plain-modified shape.	NLR and GP (PUK and RBF kernels).	GP_RBF performed better.
Kumar et al.[23]	plunging hollow jets with jet angle 60°.	MNLR, ANN, ANFIS, MARS, and GRNN.	the models' testing indicates that gbellmf-based ANFIS and ANN are effective for forecasting the oxygen transfer characteristics of descending hollow jets. MARS is more capable of making predictions than the GRNN model.
Kumar et al.[24]	various configurations of plunging hollow jet aerators.	ANFIS, SVM, ANN, M5P, and RF.	the regression equation developed in the present study's nondimensional version reasonably matches the data.
Kumar et al.[25]	conical-shaped plunging hollow jet aerator.	ANN, MNLR, MLR.	the ANN approach is more accurate and outperforms both MNLR and MLR techniques when it comes to forecasting the K_La_{20} of aerators with hollow jets.
Kramer et al.[26]	circular nozzles with diameters ranging from 13.0 to 18.1 mm.	ANN, NLR.	using the data from performed experiments and published research, the ANN and NLR approaches enabled estimation of B_{PD} over a broad range of flow (at plunge point) and jet lengths. The anticipated penetration depths match up with the investigation's observed results.
Bagatur et al.[27]	circular nozzle.	GEP, MLR, MNLR.	the suggested GEP formulations' findings are compared to those of proposed regression-based equations, and it is discovered that they are more accurate.
present Study	acrylic sheets with circular hollow jets.	RF, REPTree, and M5P.	M5P is found to be the outperforming model.

https://doi.org/10.1021/acsomega.3c03375
ACS Omega 2023, 8, 38950−38960

SVM approach was found to achieve a higher level of predictive accuracy. GEP and ANN modeling techniques were employed to predict the K_La_{20}, B_{PD}, and oxygenation efficiency of plunging water jets. The study revealed that GEP achieved improved performance when compared with ANN.[16] The present study aims to study the efficacy of decision tree techniques in predicting E_{20} in an open-channel flow system. Literature in Table 1 shows the numerous machine learning algorithms, including decision trees, which have been studied in the prediction of E_{20} in closed system jet aeration using a variety of input parameters, including jet length, thickness, diameter, and fall height. However, the coagulation of the present study's input variables, including α, J_N, Q, HR, and Fr. No, to evaluate E_{20} in an open channel is yet to be explored with machine learning/soft computing such as REPTree RF and M5P decision trees due to their efficacy in prediction. Keeping in mind the gap in research, the objectives of the current study are listed as follows.

1. To determine the potential of decision tree techniques for prediction of E_{20} in an open-channel flow system.
2. Optimization of applied decision tree techniques for outperformance, based on statistical parameters.

2. DECISION TREE APPROACH

A DT is considered to be supervised machine learning, which means the model is developed and validated on a data set that comprises the desired categorization. It consists of a tree-based structure where the information on the data set is shown by internal nodes, the choice is represented by the branches, and the outcome is each leaf node.

The DT models used in this study are discussed below.

2.1. M5P. The M5P tree is a DT that incorporates a linear regression function at the leaf nodes to make predictions regarding continuous numerical attributes.[28] Developing a tree-based model employs a partition-and-overcome strategy that consists of two stages. During the first stage, a DT is made by using separating criteria. The standard deviation of the class data determines how the M5P tree model method splits the data and entails assessment of each characteristic at each node. This culminates in a reduction of the standard deviation of the information present in the child nodes in place of the parent node. Once the all-feasible splits have been examined, the M5P algorithm selects the model that reduces the error to the utmost degree. Notably, the resulting tree-like structure may be excessively intricate, leading to overfitting. Pruning the tree back is necessary to solve this problem, for example, by replacing a leaf with a subtree. In the following stage of model tree creation, the overgrown tree is trimmed, and subsets are substituted by linear regression functions. By dividing the parameter space into subspaces, this technique gives each one a linear regression model structure. Ultimately, the M5P model tree approach serves as an effective method for predicting numerical attributes, particularly in the realm of machine learning.

2.2. Random Forest. In the field of machine learning research, ensemble learning methods have gained significant attention due to their ability to produce multiple classifiers and combine their outputs. Popular among these methods are boosting, bagging, and Random Forests.[29−31] Using input vectors and random vector samples, RF is a potent classification and regression algorithm that produces a collection of tree predictors that is organized.[32] The RF algorithm utilizes a hit-or-miss strategy to select parameters based on the best split and produces a set of random trees through the combination of bagging and random subspace techniques. Weak classification trees are integrated into the RF model, and a majority vote determines the final decision. When forest trees are created, it is important to consider the quantity of DTs (Ntree) to be produced and the number of characteristics to be checked to determine the best division. Due to the efficacy of the RF classifier and the absence of overfitting, Ntree may be set at any value.[33] Each tree is developed using two-thirds of the training data, with the remaining one-third being utilized as "out of bag" (OOB) data to assess performance. The user may select any value for this parameter, and the RF consists of k trees, where k is the total number of trees to be produced. The forest trees are developed using the CART algorithm without pruning. RF regression enables trees to grow to the depth of all of the fresh training data by using a number of factors. A training set of values is selected at random to create specific trees, and the amount of purity in the variables relative to the output is determined by the Gini index. In order to create the best RF model, variables are categorized according to their significance.

2.3. Reduced Error Pruning Tree. REPT is a machine learning approach that streamlines the modeling processes. It was introduced as the decision tree technique that is utilized initially on the training data set to identify errors, which are then minimized using REP, a variance, and the information gain reduction technique.[34] The REPT method aims to reduce the intricacy of the tree structure to simplify the modeling process when dealing with vast input data. This method is highly valued by researchers as it yields a good subtree based on postpruning outcomes. In REPT, two methods that employ the information gain ratio are employed to split and prune decision trees. The first step is the prepruning process that curtails tree expansion when data building is in progress. The second stage is the postpruning process, in which the tree develops until there are no issues in the training set process. When prepruning and postpruning phases are compared, the latter produces a more accurate tree but takes more time, whereas prepruning produces trees faster.

3. METHODOLOGY

With the goal of predicting E_{20} by circular plunging jets, the models used in this study were developed and tested based on five input variables: α, Q, J_N, HR, and Fr. No. The model development and validation were done using software called WEKA 3.9. Each model needs to have a few user-defined variables set in order to be optimized. Table 4 contains a list of each model's user-defined parameters for the research at hand. The studies were carried out to produce data sets for E_{20} from each machine learning algorithm. The data set and the method for conducting experiments are included in the sections that follow.

3.1. Data Set. The input variables, including α (varying from 0 to 3°), Q (varying from 3.41 to 4.75 l/s), and J_N (varying from 1 to 64), were utilized for predicting E_{20}. The three DT models adopted for the current investigation are RF, M5P, and REPTree. Utilizing the seven acrylic sheets, each having J_N of 1, 2, 4, 8, 16, 32, and 64 (Figure 2), comprising 63 experimental readings for E_{20} (eqs 1−3) were obtained. Out of the complete data set, the training data set was randomly chosen comprising 42 readings, which was used for model development. The remaining data set was considered for model validation with 21 data points. Table 2 shows the characteristics of training and testing data sets.

Table 2. Statistics of Training and Testing Data Sets

Training Data set						
	α (⁰)	Q (l/s)	J_N (number)	HR (cm)	Fr. No	E_{20}
mean	1.50	3.94	18.14	0.69	5.84	0.18
median	1.50	3.84	8.00	0.55	5.46	0.20
std. dev.	1.23	0.54	21.49	0.46	2.11	0.05
kurtosis	−1.53	−1.25	0.46	−0.66	−1.20	−0.59
skewness	1.14	0.61	1.33	0.76	0.26	−0.49
minimum	0.00	3.41	1.00	0.19	2.83	0.07
maximum	3.00	4.75	64.00	1.56	9.37	0.29
Testing Data set						
	α (⁰)	Q (l/s)	J_N (number)	HR (cm)	Fr. No	E_{20}
mean	1.50	4.10	18.14	0.69	6.09	0.19
median	1.50	3.84	8.00	0.55	6.37	0.18
std. dev.	1.25	0.55	21.49	0.49	2.47	0.55
kurtosis	−1.57	−1.25	0.47	−0.73	0.44	0.36
skewness	7.36	0.64	1.38	0.79	1.09	−0.07
minimum	0.00	3.41	1.00	0.19	3.79	0.09
maximum	3.00	4.75	64.00	1.56	11.15	0.32

3.2. Experimental Procedure. A tilting flume was employed for the present study to conduct experimental procedures (Figure 1) having dimensions of 0.45 m × 0.25 m × 5 m. The aeration device (Figure 2) was equipped in the channel such that water could transfer purely via hollow jets, which were impinged onto the screens. To begin the tests, tap water was used to fill the water tank, and the water was allowed to circulate in the channel with the help of a "2 HP" motor. In order to decrease the concentration of DO in the water tank, sodium sulfite (Na_2SO_3) and a catalyst, cobalt chloride ($CoCl_2$), were added in the water collected in the tank. For each test run, it was required to collect two samples of water, one being the deoxygenated water that was collected from the upstream location of the screen. The other sample (oxygenated water) was collected from the downstream location after a time of 2 min had elapsed. The concentration of dissolved oxygen in each sample of water collected was calculated using the Winkler test.[35] The water temperature of the two samples was measured by using a lab thermometer. Equations 1– 3 were used to calculate $E_{20.}$

To determine E_{20}, the following equations were used.[36]

$$E = \frac{C_y - C_x}{C_{S*} - C_x} = 1 - \frac{C_{S*} - C_y}{C_{S*} - C_x} = 1 - \frac{1}{r} \tag{1}$$

where C_x, C_y, C_{S*}, and r are the oxygen levels in the upstream and downstream points, the concentration of saturated oxygen, and the oxygen aeration deficient ratio, respectively.

The E_{20} is 1 when all of the oxygen is delivered to the water, whereas a 0 value indicates that no amount of dissolved oxygen can be transported. Using the following formula, the results from multiple temperature tests are normalized at 20 °C to maintain uniformity in measured experiments.[37]

$$1 - E_{20} = (1 - E)^{1/f} \tag{2}$$

The following time-independent formula can be used to determine the oxygen aeration efficiency (E) at water temperatures aside from 20 °C, where E_{20} denotes the transfer efficiency of oxygen at 20 °C, f denotes the exponent of aeration, and T denotes the water temperature.

$$f = 1 + 2.1 \times 10^{-2}(T - 20) + 8.25 \times 10^{-5}(T - 20)^2 \tag{3}$$

4. RESULTS

4.1. Experimental Result. The correlation matrix among governing variables was calculated using Pearson's method and is shown in Figure 3. Figure 3 reveals that Fr. No has a higher correlation of 0.859 with E_{20} followed by α, J_N, Q, HR, and Fr. No. The parameters Q and HR show a negative correlation as they are based on the property of each jet section. The cumulative effect of these parameters culminates in a positive correlation.

The following equations (eqs 4 and 5) are used to calculate the Fr. No as a function of the discharge rate and jet area

$$\text{Fr. No} = \frac{v}{\sqrt{g^{*}\text{HR}}} \tag{4}$$

Here, v is the discharge velocity and g is the accelerated gravity (g/cm^3).

$$\text{HR} = \frac{\text{flow area}}{\pi^{*}D^{*}J_N} \tag{5}$$

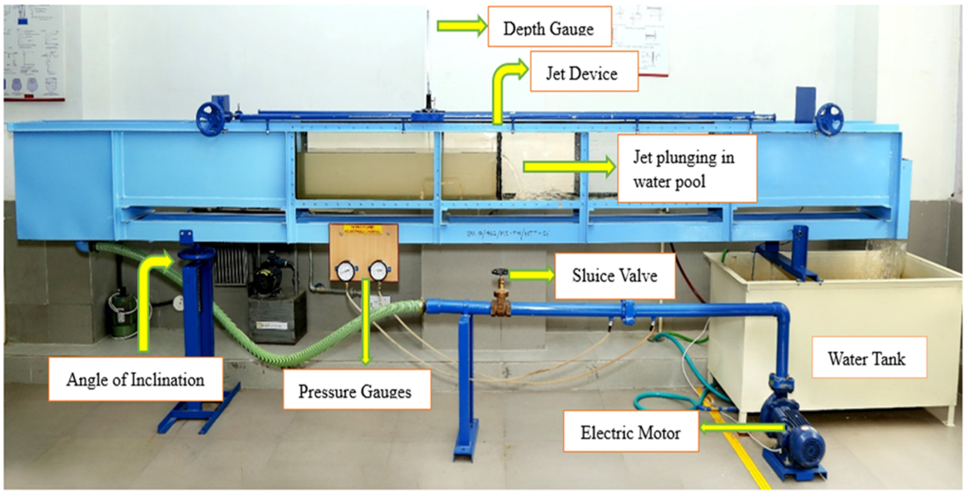


Figure 1. Experimental setup.

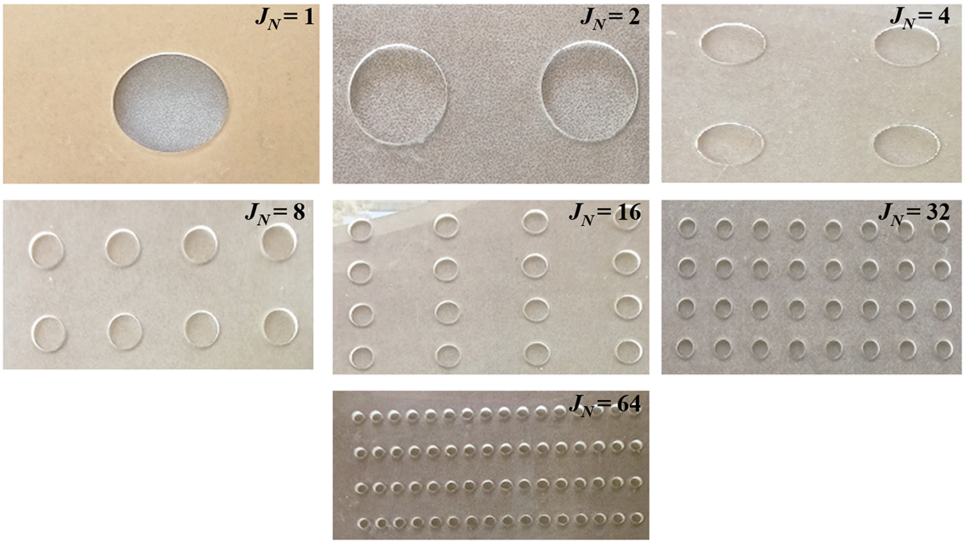


Figure 2. Circular jets fabricated on acrylic jets with different numbers.

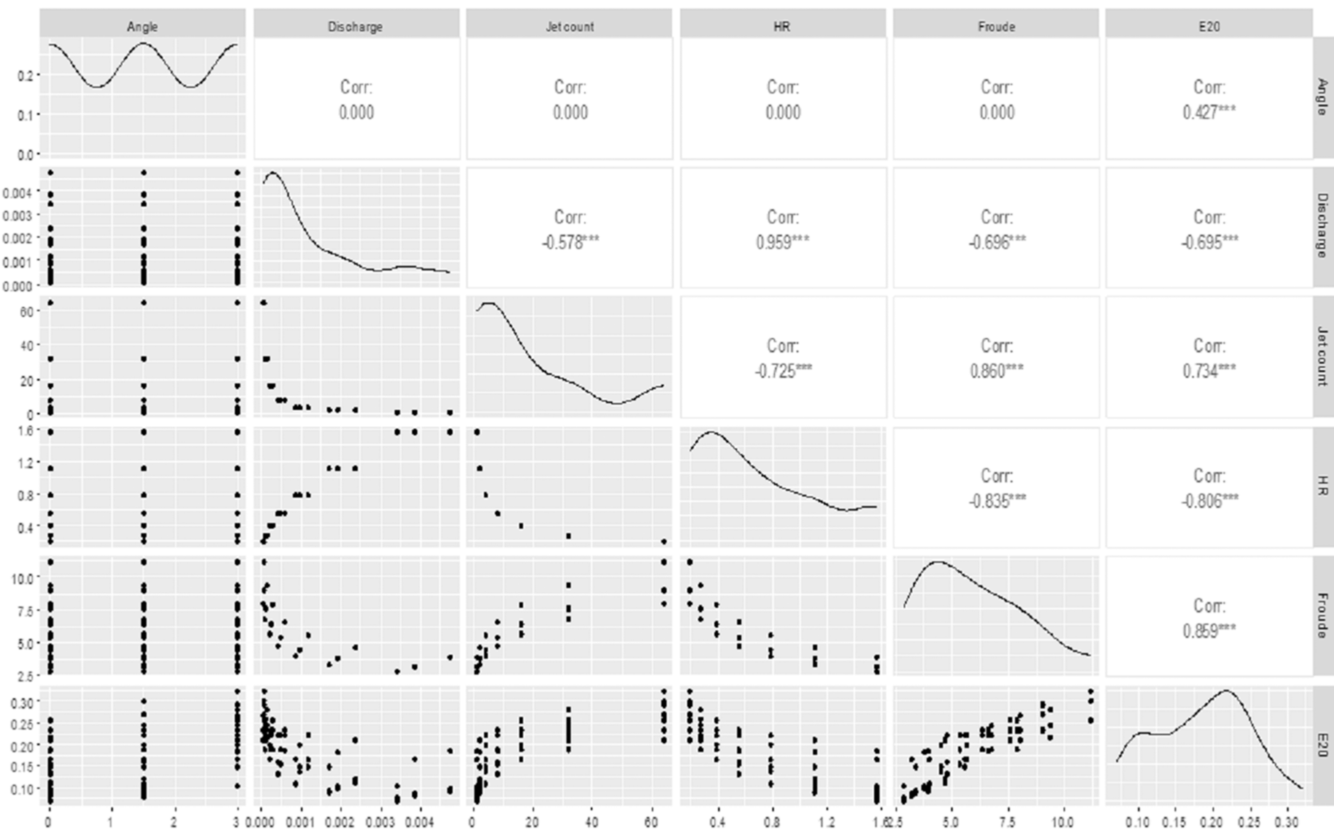


Figure 3. Pearson's correlation matrix of input parameters.

Here, the flow area is 30.75 cm^2, J_N = jet count, and D is the diameter of each jet section.

4.2. Decision Tree Modeling-Based Results. *4.2.1. Assessment of REPTree and RF Models.* The study generated predictive models, namely, REPTree and RF, and determined their performance evaluation parameters, CC, MAE, and RMSE, using eqs 6−8 as shown in Table 3, with training and testing data sets. To get the optimized model, certain user-defined values are given. The user-defined values for this research are listed in Table 4. The RF model displayed superior predictive capabilities for both data sets as compared to the REPTree model, as evidenced by the higher CC values. Specifically, the model development and validation data sets had CC values of 0.9922 and 0.9461, respectively, while the testing data CC value for the REPTree model was 0.8998, which is only marginally satisfactory (Table 5). The results for both models, obtained from Figures 4(a,b) and 5(a,b), showed that the RF model outperformed the REPTree model in predicting E_{20}, with the points in Figure 5(a,b) being close to the line of perfect agreement, signifying the accuracy of the RF model. In contrast, the points in Figure 4(a,b) were more scattered, indicating less accuracy of the REPTree model. In conclusion, the RF model

https://doi.org/10.1021/acsomega.3c03375
ACS Omega 2023, 8, 38950−38960

Table 3. Model Evaluation Parameters

S.no	parameter	range	equation
1	$\text{correlation coefficient(CC)} = \frac{\sum_{i=1}^{N}(k_i - \bar{k})(l_i - \bar{l})}{\sqrt{\sum_{i=1}^{N}(k_i - \bar{k})^2 \sum_{i=1}^{N}(l_i - \bar{l})^2}} \quad (6)$	−1 to +1	(6)
2	$\text{mean absolute error(MAE)} = \frac{1}{N}\sum_{i=1}^{N}\lvert k_i - l_i\rvert \quad (7)$	$0 < \text{MAE} < \infty$	(7)
3	$\text{root mean squre error(RMSE)} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(k_i - l_i)^2} \quad (8)$	$0 < \text{RMSE} < \infty$	(8)

Table 4. Parameters Defined by the User for the Present Study to Optimize the Models

DT models	parameters defined by the user
RF	• batchSize: 100 • numIterations: 100 • seed: 1
REPTree	• batchSize: 100 • numFolds: 3 • seed: 4
M5P	• batchSize: 100

Table 5. Statistical Indices Used for Model Evaluation

DT models	CC	RMSE	MAE
	training data set		
REPTree	0.9405	0.0178	0.0126
random forest	0.9922	0.0083	0.0068
M5P	0.9728	0.014	0.0112
	testing data set		
REPTree	0.8998	0.0293	0.0258
random forest	0.9461	0.0226	0.0185
M5P	0.9564	0.0193	0.0143

proved to be more effective than the REPTree model in predicting E_{20}.

4.2.2. Assessment of the M5P Model. The study presented here utilized the M5P model to predict E_{20}. To train and validate the M5P model, both training and testing data sets were used. Table 4 lists the parameters defined by a user in this research to optimize the M5P model. The accuracy of the model is higher when the projected value of the slope of the line of agreement is closer to unity when compared to the measured data, as demonstrated in Figure 6(a,b). The statistical results obtained from the M5P model are presented in Table 5, which show fair results with agreeable CC values of 0.9728 and 0.9564 for the model development and validating stages, respectively. Additionally, it was found that the observations of RMSE and MAE were lower during the training stage but slightly increased during the testing stage.

4.3. Comparison of Models. This section pertains to the evaluation of various models employed in the present study. REPTree, RF, and M5P models were utilized to predict E_{20}, using α, Q, J_N, HR, and Fr. No. as input parameters. To analyze the outperforming model, statistical parameters were used as shown in Table 3, of which results of evaluating each model against three statistical indices are presented in Table 5. The consistency of each model with experimental data is depicted in Figure 7, which indicates that the prediction of E_{20} can be done effectively with the DT models employed in this study. In order to obtain conclusive results, it is also important to observe the errors associated with models, the results of which are illustrated in Figure 8. It is observed that the REPTree model has higher errors when compared to the other models, in both model development and validation stages. The M5P model exhibited better consistency both before and after training. Figure 9 displays the box plot of model outcomes for the testing stage. The actual and M5P model median and maximum values are quite close. The IQR range of the actual data is 0.122, while

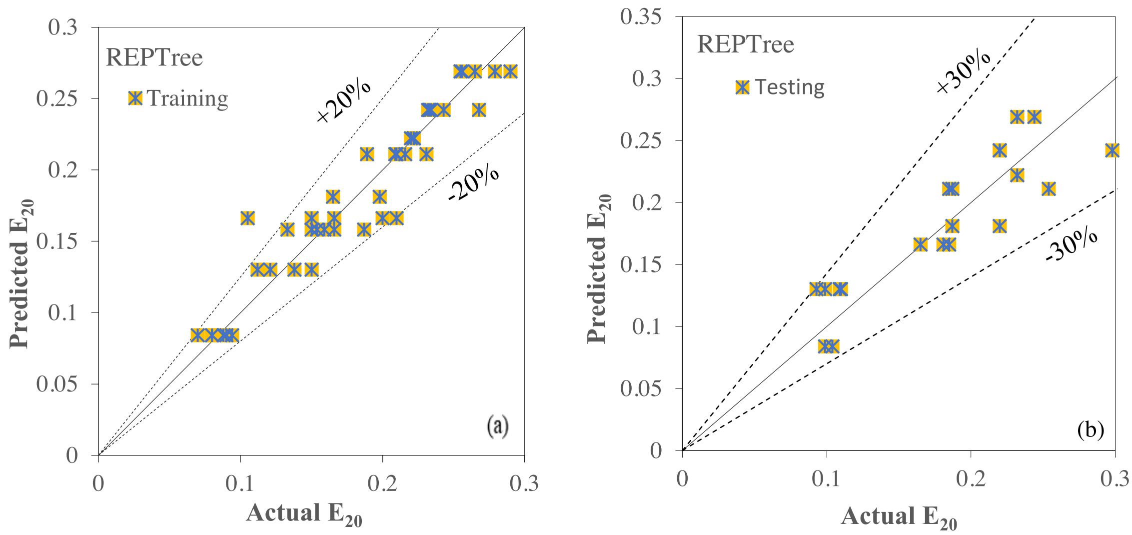


Figure 4. Observed and predicted value of E_{20} using REPTree during (a) training and (b) testing.

https://doi.org/10.1021/acsomega.3c03375
ACS Omega 2023, 8, 38950−38960

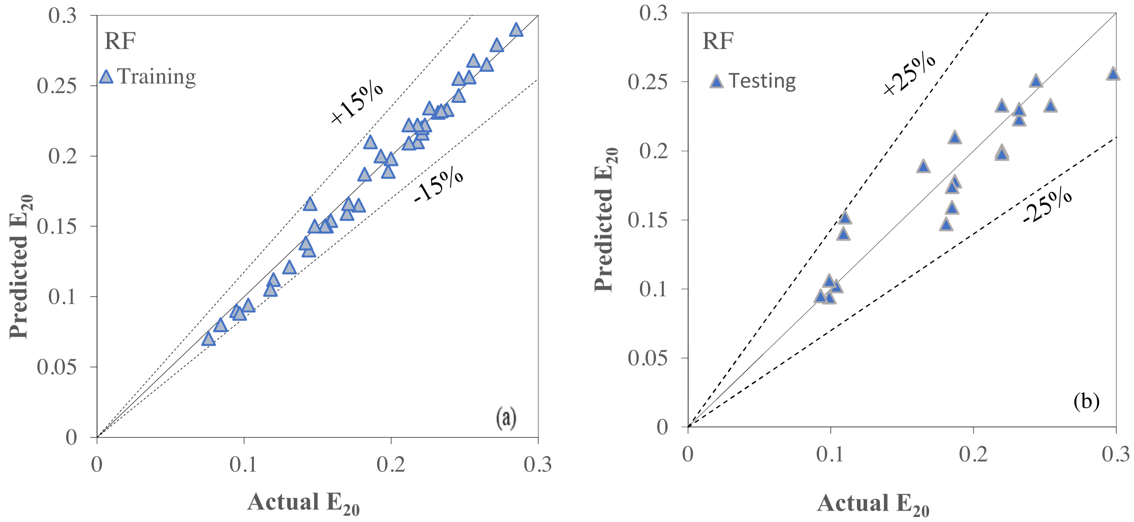


Figure 5. Actual and predicted values of E_{20} using the RF model during (a) training and (b) testing.

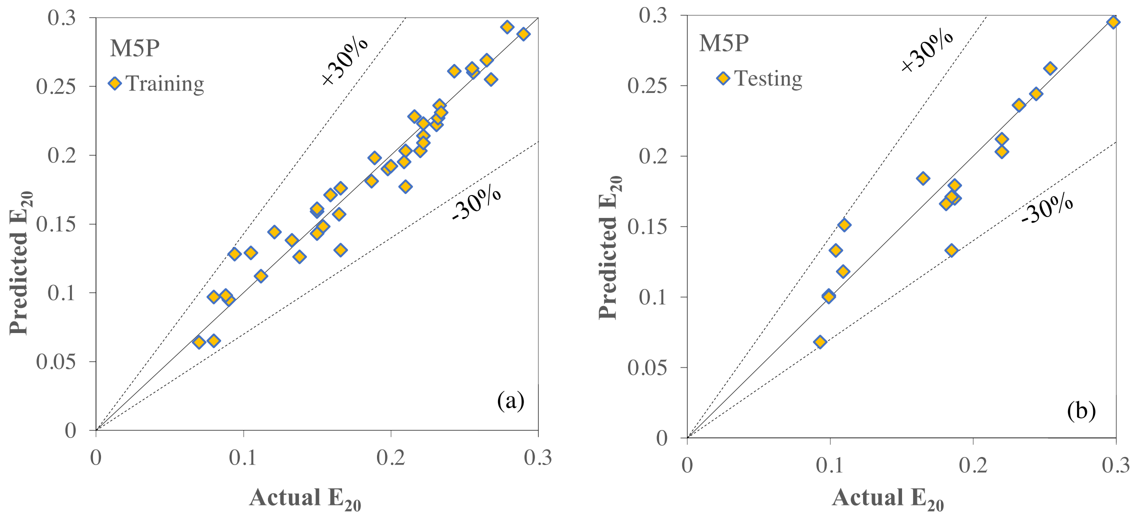


Figure 6. Actual and predicted values of E_{20} using the M5P model (a) training and (b) testing.

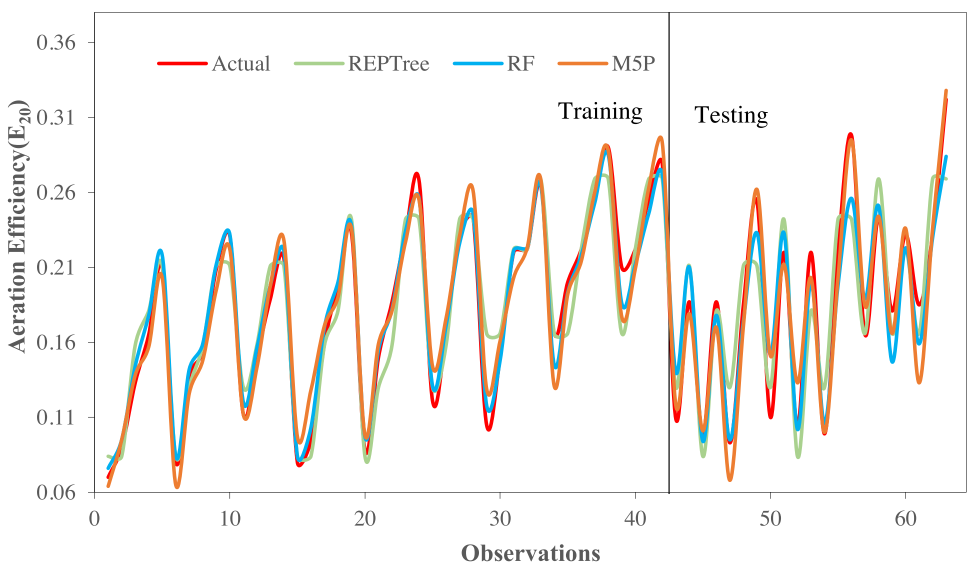


Figure 7. Comparison of decision trees with actual data.

https://doi.org/10.1021/acsomega.3c03375
ACS Omega 2023, 8, 38950−38960

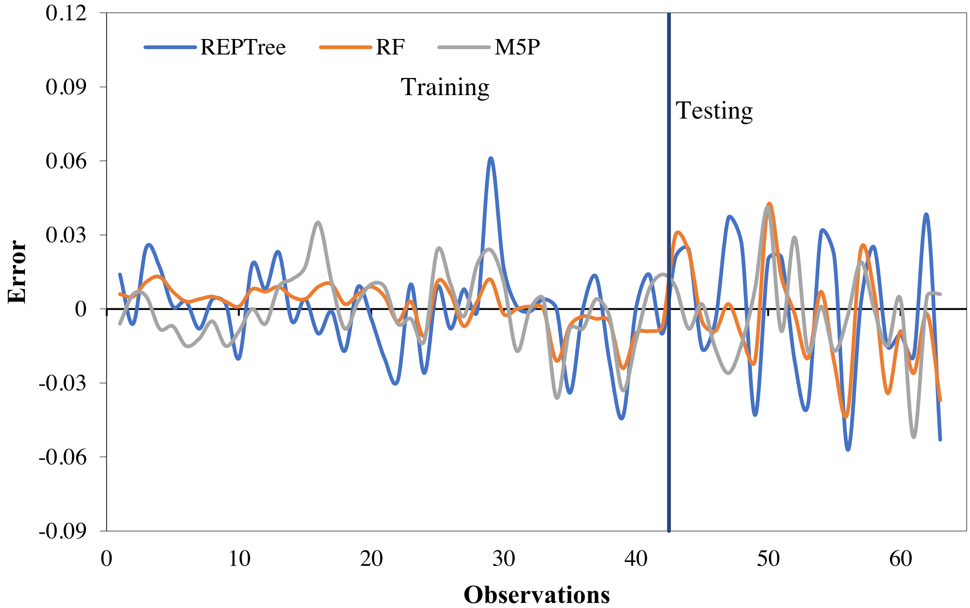


Figure 8. Error values of REPTree, RF, and M5P during the training and testing stages.

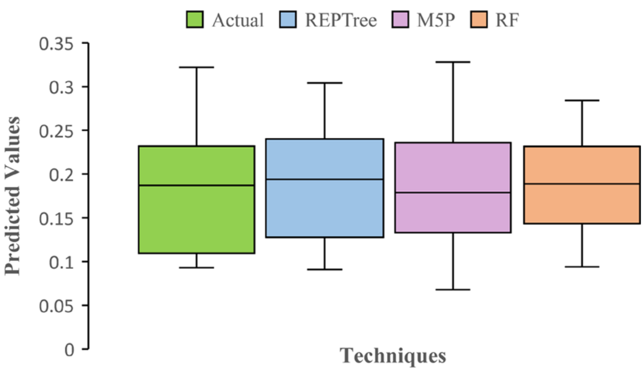


Figure 9. Box plot with actual and applied decision tree techniques.

those of REPTree, M5P, and RF are 0.112, 0.103, and 0.083, respectively. The difference in mean between the actual and observed values is the minimum for the M5P model, at 0.000476. The Q2 values obtained from the box plot of the actual data set is 0.187. The Q2 values for the models REPTree, M5P, and RF are 0.181, 0.179, and 0.189, respectively. The mean values of M5P (0.185381) and RF (0.183524) are very close to the observed data set (0.187905).

As observed from Table 5 and Figure 10, it is concluded that M5P is the outperforming model followed by RF and REPTree, respectively.

5. DISCUSSION

A supervised machine learning method called a DT is capable of performing tasks such as regression and classification. The DT method pulls key information from databases and arranges them into a tree-like structure with internal and terminal nodes exhibiting splits and leaves, respectively. By combining different trees, each of which has its own unique set of rules, a collection of rules that would be used during the regression stage is beginning to take shape. As DT combines fundamental inquiries regarding the availability of the data, occasionally it might be simpler to understand than classifiers like NN and SVM. Additionally, there have been successful techniques for eliminating decision rules from decision trees. Due to their versatility, DTs can handle items with a mix of real-valued and category characteristics as well as those with certain features missing.[38] They are expressive enough to represent a range of data partitions, as opposed to classifiers that depend on only one limit, such as logistic regression or SVM. Due to how closely these models resemble human thinking and how simple they are to comprehend, DT approaches have been frequently used to create categorization models.[39]

In this study, the focus is on predicting E_{20}, which is essential for the effective management of densely polluted water resources. The study aims to evaluate the effectiveness of decision trees, including M5P, RF, and REPTree, to estimate circular jet aeration in an open-channel flow. To analyze E_{20}, five input parameters are used, which include α, Q, J_N, HR, and Fr. No. The models' effectiveness is evaluated using several statistical criteria, and it is found that during the training stage, the CC value of RF was 0.9922, but it declined in the testing stage to 0.9461. The results demonstrate that M5P is an accurate model for predicting E_{20} based on the current data set with the highest CC of 0.9564 and the least errors during the testing stage as compared to the other models. On the other hand, REPTree acquired the least CC values among other applied DT models during testing, making it the least-performing model for the current data set. The correlation matrix (Figure 3) results showed that the parameter Fr. No has a significant impact on output. It has acquired the highest correlation among other parameters with E_{20}.

Decision trees such as RF, M5P, and REPTree have been used globally.[40−42] M5P and ANFIS were used to predict E_{20} at modified small Parshall Flume. The comparison among these two models showed that M5P was the outperforming model.[43] To predict the aeration performance of weirs, machine learning techniques such as M5P (pruned and unpruned) and GP (PUK and RBF) were used. The study's findings demonstrated that the M5P pruned tree had similar observations as experimented data, and hence, the M5P pruned tree can be very well used to predict the E_{20}.[44] The efficacy of artificial intelligence models, namely, M5P and EPR, was tested to predict E_{20} at stepped weirs. It was observed that both the EPR and M5P methods provided

https://doi.org/10.1021/acsomega.3c03375
ACS Omega 2023, 8, 38950−38960

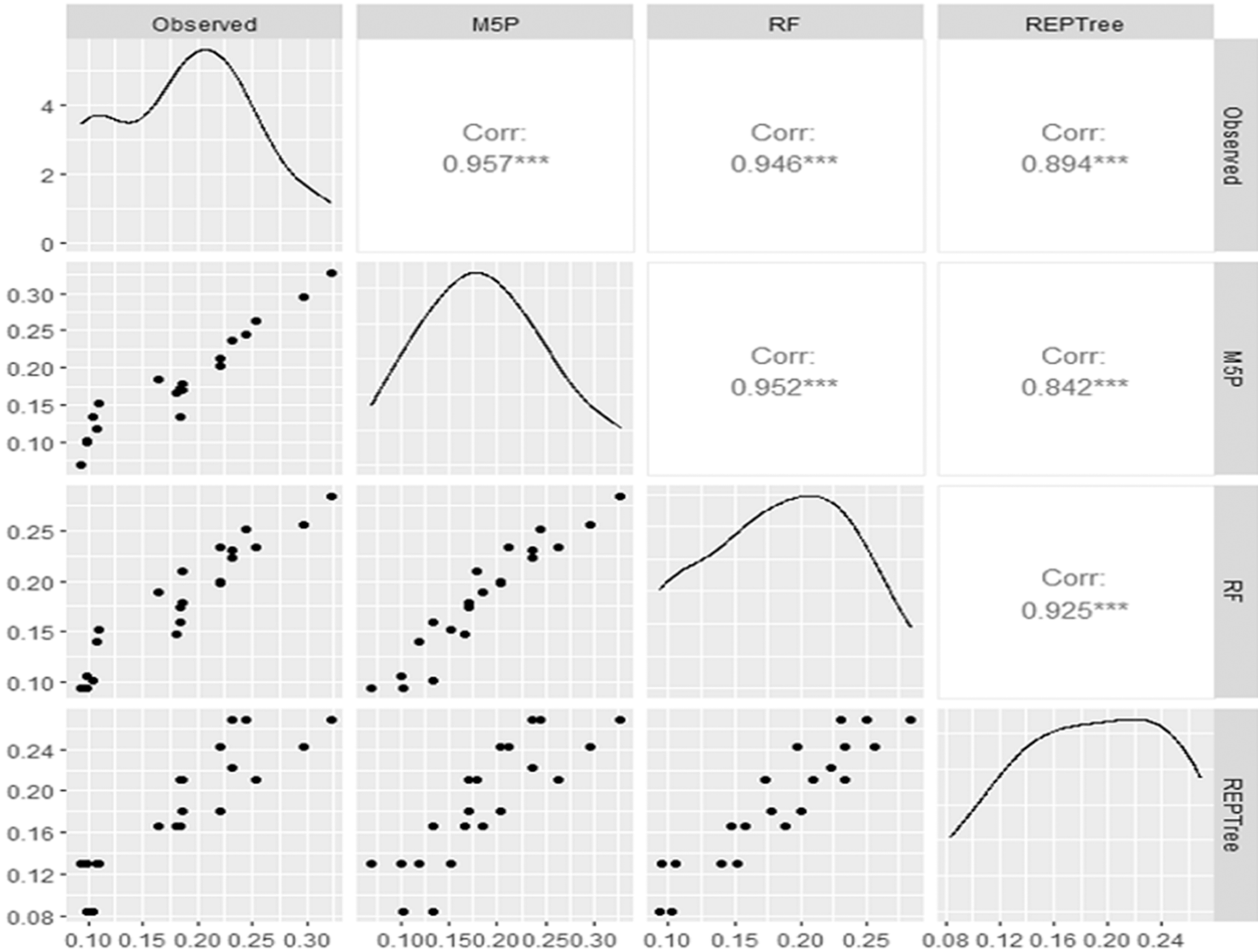


Figure 10. Pearson's correlation among decision tree techniques.

satisfactory predictions for the aeration efficiency. However, it was also noticed that for each regime the M5 model provided multiple equations, whereas the EPR method gave a single equation for each turn.[45] The potential of soft computing techniques such as RF, M5P, MARS, and GMDH was observed in an aeration study of Parshall and modified venturi flume. Results of the analysis showed that MARS had better predictions when compared to other models.[46] Another study investigated the modeling performances of soft computing techniques such as RF and ANN to predict the E_{20} of gabion stepped weirs. The comparison of performance of these two techniques suggests that ANN outperformed RF.[47] In an open-channel flow, RF, reduced error pruning tree, ANN, GP, and support vector machine were tested to predict circular and square jet aeration efficiency (E_{20}).[48]

Due to the simplicity of DT models, many researchers have been using them in various other fields other than predicting E_{20}. One such study was to predict the daily water level of the Zrebar lake in Iran. The authors compared the efficacy of applied models, i.e., M5P (pruned tree), RF, RT, and REPTree. Their findings showed that each constructed model had a high level of accuracy. M5P exceeded other models, followed by RT and RF equally and then REPTree.[49] M5P and ANN were used in another study to predict SPI and meteorological draft. The models were compared, and M5P was found to be a better-predicting model than ANN with a higher R value and fewer errors.[50] The predictive models for the dynamic modulus of asphalt concretes were also created by using the M5P model tree approach. The findings showed that models created using the M5P method performed better than models created before.[51]

Therefore, from the above discussion, it can be concluded that the DT plays a vital role in predicting the output parameters, for example, E_{20}, etc. In this study, the M5P model outperformed all of the other applied models.

6. CONCLUSIONS

The findings of the current study may be summed up as follows.

1. The focus of the present study was to analyze the efficacy of decision tree techniques to predict E_{20}. It was found that all of the applied techniques (M5P, RF, and REPTree) showed considerably good results as they all had CC values higher than 0.8; this indicates that decision tree techniques can be used to predict the dissolution of dissolved oxygen, i.e., aeration efficiency (E_{20}) in water in an open channel.

2. When comparing the techniques, it was observed that M5P achieved the highest CC value in the testing stage, which showed the compatibility of M5P with the current data set. It was also found that REPTree has the least CC value.

https://doi.org/10.1021/acsomega.3c03375
ACS Omega 2023, 8, 38950−38960

7. FUTURE SCOPE

This study is based on circular jet aeration in an open streamflow system. The authors recommend increasing the number of jets and exploring the effect of different jet configurations on oxygenation in an open channel. Also, the study only focuses on one output parameter, i.e., E_{20}. Other parameters can be used such as the OTE and kinetic energy of the system pertaining to an open channel.

■ AUTHOR INFORMATION

Corresponding Authors

Raj Kumar − *Department of Mechanical Engineering, Gachon University, Seongnam 13120, South Korea;* orcid.org/0000-0002-6507-7231; Email: errajap@gmail.com

Mohammad Amir Khan − *Department of Civil Engineering, Galgotia College of Engineering, Greater Noida 201310, India;* orcid.org/0009-0006-9847-2197; Email: amirmdamu@gmail.com

Authors

Diksha Puri − *School of Environmental Science, Shoolini University, Solan 173229 Himachal Pradesh, India*

Daeho Lee − *Department of Mechanical Engineering, Gachon University, Seongnam 13120, South Korea;* orcid.org/0000-0002-8119-9677

Dhananjay Vasant khankal − *Department of Mechanical Engineering, Sinhgad College of Engineering, Pune 411041 Maharashtra, India*

Mohindra Singh Thakur − *Department of Civil Engineering, Shoolini University, Solan 173229 Himachal Pradesh, India*

Faisal M. Alfaisal − *Department of Civil Engineering, College of Engineering, King Saud University, Riyadh 11421, Saudi Arabia*

Shamshad Alam − *Department of Civil Engineering, College of Engineering, King Saud University, Riyadh 11421, Saudi Arabia*

Complete contact information is available at:
https://pubs.acs.org/10.1021/acsomega.3c03375

Author Contributions

D.P.: Writing and preparing the manuscript. D.L.: Reviewing, editing, revision, supervision. D.V.K.: Reviewing, editing. M.S.T.: Reviewing, editing, supervision. F.M.A.: Reviewing, editing, supervision. S.A.: Reviewing, editing, supervision. R.K.: Writing and preparing the manuscript, methodology, supervision, revision. M.A.K.: Reviewing, editing, supervision.

Notes

The authors declare no competing financial interest.

The data was gathered from published works that are in the public domain and properly cited.

■ ACKNOWLEDGMENTS

The authors would like to acknowledge the support provided by Researchers Supporting Project Number RSP2023R297, King Saud University, Riyadh, Saudi Arabia.

■ ABBREVIATIONS

ANFIS	adaptive network-based fuzzy inference
ANN	artificial neural network
B_{PD}	bubble penetration depth
CART	classification and regression tree
DT	decision tree
DO	dissolved oxygen
EPR	evolutionary polynomial regression
GEP	genetic expression programming
GP	Gaussian process
GMDH	group method of data handling
GRNN	general regression neural network
IQR	interquartile range
K_La_{20}	volumetric oxygen transfer coefficient
LS-SVM	least-square support vector machine
MARS	multivariate adaptive regression spline
MLR	multiple linear regression
MNLR	multiple nonlinear regression
NLR	nonlinear regression
NN	neural network
OTE	oxygen transfer efficiency
PUK	Pearson's VII universal kernel
Q2	median value
R	Pearson's correlation coefficient
RBF	radial basis function
RT	random tree
Std. dev.	standard deviation
SPI	standardized precipitation index

■ REFERENCES

(1) Bin, A. K.; SMITH, J. M. Mass transfer in a plunging liquid jet absorber. *Chem. Eng. Commun.* **1982**, *15* (5−6), 367−383.

(2) Deswal, S. Oxygen transfer by multiple inclined plunging water jets. *Int. J. Civil Environ. Eng.* **2008**, *2* (3), 57−63.

(3) Shukla, B. K.; Kumar, V. R.; Goel, A. A comprehensive review of surface jet aerators. *Pollut. Res.* **2018**, *37*, 20−25.

(4) Kumar, M.; Tiwari, N. K.; Ranjan, S. Experimental study on oxygen mass transfer characteristics by plunging hollow jets. *Arabian J. Sci. Eng.* **2021**, *46*, 4521−4532.

(5) Puri, D.; Sihag, P.; Kumar, R. A Comparison of Single and Multiple Jets in Terms of Aeration Efficiency. In *IOP Conference Series: Earth and Environmental Science*; IOP Publishing, 2023012014 DOI: 10.1088/1755-1315/1110/1/012014.

(6) Bhattacharya, B.; Solomatine, D. P. Neural networks and M5 model trees in modelling water level−discharge relationship. *Neurocomputing* **2005**, *63*, 381−396.

(7) Solomatine, D. P.; Siek, M. B. L.Flexible and optimal M5 model trees with applications to flow predictions. In *Hydroinformatics: (In 2 Volumes, with CD-ROM)* 2004; pp 1719−1726 DOI: 10.1142/9789812702838_0212.

(8) Pal, M.; Singh, N. K.; Tiwari, N. K. M5 model tree for pier scour prediction using field dataset. *KSCE J. Civil Eng.* **2012**, *16*, 1079−1084, DOI: 10.1007/s12205-012-1472-1.

(9) Herrera, M.; Torgo, L.; Izquierdo, J.; Pérez-García, R. Predictive models for forecasting hourly urban water demand. *J. Hydrol.* **2010**, *387* (1−2), 141−150, DOI: 10.1016/j.jhydrol.2010.04.005.

(10) Kusiak, A.; Zheng, H.; Zhang, Z. Virtual wind speed sensor for wind turbines. *J. Energy Eng.* **2011**, *137* (2), 59−69.

(11) Polishchuk, P. G.; Muratov, E. N.; Artemenko, A. G.; Kolumbin, O. G.; Muratov, N. N.; Kuz'min, V. E. Application of random forest approach to QSAR prediction of aquatic toxicity. *J. Chem. Inf. Model.* **2009**, *49* (11), 2481−2488, DOI: 10.1021/ci900203n.

(12) Leshem, G.; Ritov, Y. A. Traffic flow prediction using adaboost algorithm with random forests as a weak learner. *Int. J. Math. Comput. Sci.* **2007**, *1* (1), 1−6.

(13) Larivière, B.; Van den Poel, D. Predicting customer retention and profitability by using random forests and regression forests techniques. *Expert Syst. Appl.* **2005**, *29* (2), 472−484, DOI: 10.1016/j.eswa.2005.04.043.

(14) Baylar, A.; Hanbay, D.; Batan, M. Application of least square support vector machines in the prediction of aeration performance of plunging overfall jets from weirs. *Expert Syst. Appl.* **2009**, *36* (4), 8368−8374.

(15) Deswal, S. Modeling oxygen-transfer by multiple plunging jets using support vector machines and Gaussian process regression techniques. *Int. J. Civil Environ. Eng.* **2011**, *5* (1), 1−6.

(16) Bagatur, T.; Onen, F. A predictive model on air entrainment by plunging water jets using GEP and ANN. *KSCE J. Civil Eng.* **2014**, *18*, 304−314.

(17) Kumar, M.; Tiwari, N. K.; Ranjan, S. Prediction of oxygen mass transfer of plunging hollow jets using regression models. *ISH J. Hydraul. Eng.* **2020**, *26* (1), 23−30.

(18) Kumar, M.; Tiwari, N. K.; Ranjan, S. Kernel function-based regression approaches for estimating the oxygen transfer performance of plunging hollow jet aerator. *J. Achiev. Mater. Manuf. Eng.* **2019**, *2* (95), 74−84, DOI: 10.5604/01.3001.0013.7917.

(19) Singh, S.; Deswal, S.; Pal, M. Modeling of overall volumetric oxygen transfer by plunging jets of different geometries. *Int. J. Civ. Struct. Eng.* **2010**, *1* (3), 591−605, DOI: 10.6088/ijcser.00202010049.

(20) Bodana, D.; Tiwari, N. M.; Ranjan, S.; Ghanekar, U. Estimation of the depth of penetration in a plunging hollow jet using artificial intelligence techniques. *Arch. Mater. Sci. Eng.* **2020**, *2*, 49−61, DOI: 10.5604/01.3001.0014.3354.

(21) Onen, F. Prediction of penetration depth in a plunging water jet using soft computing approaches. *Neural Comput. Appl.* **2014**, *25*, 217−227, DOI: 10.1007/s00521-013-1475-y.

(22) Kumar, M.; Ranjan, S.; Tiwari, N. K.; Gupta, R. Plunging hollow jet aerators-oxygen transfer and modelling. *ISH J. Hydraul. Eng.* **2018**, *24* (1), 61−67.

(23) Kumar, M.; Tiwari, N. K.; Ranjan, S. Soft computing based predictive modelling of oxygen transfer performance of plunging hollow jets. *ISH J. Hydraul. Eng.* **2022**, *28* (sup1), 223−233.

(24) Kumar, M.; Tiwari, N. K.; Ranjan, S. Application of machine learning methods in estimating the oxygenation performance of various configurations of plunging hollow jet aerators. *J. Environ. Eng.* **2022**, *148* (11), No. 04022070.

(25) Kumar, M.; Ranjan, S.; Tiwari, N. K. Oxygen transfer study and modeling of plunging hollow jets. *Appl. Water Sci.* **2018**, *8*, No. 121, DOI: 10.1007/s13201-018-0740-8.

(26) Kramer, M.; Wieprecht, S.; Terheiden, K. Penetration depth of plunging liquid jets−A data driven modelling approach. *Exp. Therm. Fluid Sci.* **2016**, *76*, 109−117, DOI: 10.1016/j.expthermflusci.2016.03.007.

(27) Bagatur, T.; Onen, F. Prediction of flow and oxygen transfer by a plunging water jets with genetic expression programming (GEP) models. *Arabian J. Sci. Eng.* **2014**, *39*, 4421−4432.

(28) Quinlan, J. R. Learning with Continuous Classes. In *5th Australian Joint Conference on Artificial Intelligence* Singapore, 1992; pp 343−348.

(29) Freund, Y.; Schapire, R. E.Experiments with a new boosting algorithm. In *icml* 1996; Vol. *96*, pp 148−156.

(30) Breiman, L. Bagging predictors. *Mach. Learn.* **1996**, *24* (2), 123−140.

(31) Upadhya, A.; Thakur, M. S.; Sihag, P.; Kumar, R.; Kumar, S.; Afeeza, A.; Afzal, A.; Saleel, C. A. Modelling and prediction of binder content using latest intelligent machine learning algorithms in carbon fiber reinforced asphalt concrete. *Alexandria Eng. J.* **2023**, *65*, 131−149, DOI: 10.1016/j.aej.2022.09.055.

(32) Scornet, E.; Biau, G.; Vert, J. P. Consistency of random forests. *Ann. Statist.* **2015**, *43* (4), 1716−1741.

(33) Guan, H.; Li, J.; Chapman, M.; Deng, F.; Ji, Z.; Yang, X. Integration of orthoimagery and lidar data for object-based urban thematic mapping using random forests. *Int. J. Remote Sens.* **2013**, *34* (14), 5166−5186, DOI: 10.1080/01431161.2013.788261.

(34) Quinlan, J. R. Simplifying decision trees. *Int. J. Man-Mach. Stud.* **1987**, *27* (3), 221−234, DOI: 10.1016/S0020-7373(87)80053-6.

(35) Apha, A. WEF Standard Methods for the Examination of Waters and Wastewaters *Washington, DC* 2005.

(36) Gulliver, J. S.; Wilhelms, S. C.; Parkhill, K. L. Predictive capabilities in oxygen transfer at hydraulic structures. *J. Hydraul. Eng.* **1998**, *124* (7), 664−671.

(37) Novak, P.; Gulliver, J. S.; Wilhelms, S. C.; Parkhill, K. L. Predictive Capabilities in Oxygen Transfer at Hydraulic Structures. *J. Hydraul. Eng.* **1999**, *125* (9), 989.

(38) Quinlan, J. R. *C4. 5: Programs for Machine Learning*, Elsevier 2014.

(39) Kotsiantis, S. B. Decision trees: a recent overview. *Artif. Intell. Rev.* **2013**, *39*, 261−283.

(40) Thai Pham, B.; Tien Bui, D.; Prakash, I. Landslide susceptibility modelling using different advanced decision trees methods. *Civil Eng. Environ. Syst.* **2018**, *35* (1−4), 139−157.

(41) Sekhar, C. R.; Madhu, E. Mode choice analysis using random forrest decision trees. *Trans. Res. Procedia* **2016**, *17*, 644−652, DOI: 10.1016/j.trpro.2016.11.119.

(42) Sihag, P.; Mohsenzadeh Karimi, S.; Angelaki, A. Random forest, M5P and regression analysis to estimate the field unsaturated hydraulic conductivity. *Appl. Water Sci.* **2019**, *9*, 1−9.

(43) Sangeeta; Ranjan, S.; Tiwari, N. K. Aeration Efficiency Evaluation of Modified Small Parshall Flume Using M5P and Adaptive Neuro-Fuzzy Inference System. In *Sustainable Engineering: Proceedings of EGRWSE 2018*; Springer: Singapore, 2019; pp 243−252.

(44) Jaiswal, A.; Goel, A. Evaluation of Aeration Efficiency of Triangular Weirs by Using Gaussian Process and M5P Approaches. In *Advanced Engineering Optimization Through Intelligent Techniques: Select Proceedings of AEOTIT 2018*; Springer: Singapore, 2020; pp 749−756.

(45) Sattar, A. A.; Elhakeem, M.; Rezaie-Balf, M.; Gharabaghi, B.; Bonakdari, H. Artificial intelligence models for prediction of the aeration efficiency of the stepped weir. *Flow Meas. Instrum.* **2019**, *65*, 78−89, DOI: 10.1016/j.flowmeasinst.2018.11.017.

(46) Sihag, P.; Dursun, O. F.; Sammen, S. S.; Malik, A.; Chauhan, A. Prediction of aeration efficiency of parshall and modified venturi flumes: application of soft computing versus regression models. *Water Supply* **2021**, *21* (8), 4068−4085.

(47) Verma, A.; Ranjan, S.; Ghanekar, U.; Tiwari, N. K. Soft Computing Techniques for Predicting Aeration Efficiency of Gabion Stepped Weir. In *Proceedings of the International Conference on Industrial and Manufacturing Systems (CIMS-2020) Optimization in Industrial and Manufacturing Systems and Applications*; Springer International Publishing, 2022; pp 117−122.

(48) Puri, D.; Kumar, R.; Sihag, P.; Thakur, M. S.; Perveen, K.; Alafia, F. M.; Lee, D. Analytical investigation of the impact of jet geometry on aeration effectiveness using soft computing techniques. *ACS Omega* **2023**, *8*, 31811−31825.

(49) Nhu, V. H.; Shahabi, H.; Nohani, E.; Shirzadi, A.; Al-Ansari, N.; Bahrami, S.; Nguyen, H.; et al. Daily water level prediction of Zrebar Lake (Iran): a comparison between M5P, random forest, random tree and reduced error pruning trees algorithms. *ISPRS Int. J. Geo-Inf.* **2020**, *9* (8), 479.

(50) Pande, C. B.; Al-Ansari, N.; Kushwaha, N. L.; Srivastava, A.; Noor, R.; Kumar, M.; Moharir, K. N.; Elbeltagi, A. Forecasting of SPI and meteorological drought based on the artificial neural network and M5P model tree. *Land* **2022**, *11* (11), 2040.

(51) Behnood, A.; Daneshvar, D. A machine learning study of the dynamic modulus of asphalt concretes: An application of M5P model tree algorithm. *Constr. Build. Mater.* **2020**, *262*, No. 120544.