Received: Added at production | Revised: Added at production | Accepted: Added at production

DOI: xxx/xxxx

ARTICLE TYPE

Radial basis neural tree model for improving waste recovery process in a paper industry

Tanujit Chakraborty*[1] | Swarup Chattopadhyay[2] | Ashis Kumar Chakraborty[1]

[1]SQC and OR Unit, Indian Statistical Institute, Kolkata, Indian
[2]Machine Intelligence Unit, Indian Statistical Institute, Kolkata, Indian

Correspondence
*Tanujit Chakraborty, ISI Email: tanujit_r@isical.ac.in

Present Address
Present address

Abstract

In this article, we propose a novel hybridization of regression trees (RT) and radial basis function networks (RBFN), namely, radial basis neural tree (RBNT) model, for waste recovery process improvement in the paper industry. As a by-product of the paper manufacturing process, a lot of waste along with valuable fibers and fillers come out from the paper machine. The waste recovery process (WRP) involves separating the unwanted materials from the valuable ones so that the recovered fibers and fillers can be further reused in the production process. This job is done by fiber-filler recovery equipment (FFRE). The efficiency of FFRE depends on several crucial process parameters and monitoring them is a difficult proposition. The proposed model can be useful to find the essential parameters from the set of available data and perform prediction task to improve waste recovery process efficiency. An idea of parameter optimization along with regularity conditions for the universal consistency of the proposed model are given. The proposed model has the advantages of easy interpretability and excellent performance when applied to the FFRE efficiency improvement problem. Improved waste recovery will help the industry to become environmentally friendly with less ecological damage apart from being cost-effective.

KEYWORDS:
Waste recovery; Regression tree; Radial basis function network; Radial basis neural tree

1 | INTRODUCTION

This work is motivated by a specific problem in a modern paper industry which produces papers for multiple uses. Paper machine produces papers by using pulp, fibers, fillers, chemical lubricants, and huge water resources. As a by-product of the paper manufacturing process, a lot of waste materials comes out from the paper machine. These waste materials contain lean water, garbages and valuable fibers, fillers. To save these useful materials from the waste, an equipment called FFRE (also popularly known as Krofta supercell) is used by many paper manufacturing industries [1,2]. Dissolved air flotation cum sedimentation process (DAFSP) is used in FFRE to save valuable materials from waste. If FFRE becomes efficient, it will help the company to reuse the relevant material which will be economical for the company. However, the efficiency of FFRE is not always satisfactory, and that causes a monetary loss for the company. A lot of preliminary analysis was done to find out a set of possible causal variables that affects FFRE efficiency which in other words means high recovery percentage of valuable materials. The objective of this paper is to correctly find out the most important process parameters from the set of all available causal variables. Also, we will try to develop a prediction model that can help the company to estimate future loses. This work aims to help the company in improving the quality and productivity of the paper manufacturing process. Furthermore, process improvement through waste management and valuable material recovery can make the manufacturing process environmentally friendly with very less ecological damage. The formulation of the above-stated problem along with data description is presented in Section 4.

The problem can be viewed as a typical nonparametric regression problems where one can establish a relationship between the response variable (recovery percentage of FFRE) and the major causal variables (process parameters of DAFSP) without having any prior information about the data. Regression trees (RT), support vector regression (SVR) and artificial neural networks (ANN) have been applied for various prediction tasks in water quality improvement, water planning and many other related problems [3,4,5,6]. Even various hybrid regression models have been developed for performing regression task in many real-life problems like water demand forecasting and others [7,8,9,10]. To develop a prediction model for the waste recovery improvement that can also find out critical parameters among the set of possible parameters, we take recourse to nonparametric regression methodology. Tree-structured model, RT, is popular for its easy interpretability and built-in mechanism of important variable selection [11] but they are sometimes unstable. On the other hand, neural networks (NN) were developed to mathematically model human intellectual abilities by biologically plausible engineering designs [12] but it may fail when limited data are available. Among the family of ANNs, radial basis function networks (RBFN) have the advantages of having only one hidden layer, less time complexity, and easy interpretability. To harness their advantages as well as to overcome their drawbacks, several works of literature concentrated on the hybridization of RT and ANN models, viz. entropy nets [13] and neural tree [14].

Motivated by the above discussion, we have proposed a radial basis neural tree (RBNT) model which utilizes the power of both the tree-based models and ANN to solve the production process efficiency problem. We also fill the gap between theory and practice by showing the consistency of the proposed regression model. In our model, we combine RBFN with RT to enhance the predicting accuracy for this typical regression problem. In the case of FFRE dataset, parameter optimization within the proposed RBNT framework yields the proposed model that is intermediate between RT and RBFN and outperforms RT, RBFN and other commonly used nonparametric regression models. The proposed model has the advantages of higher accuracy, converges much faster than many other complex hybrid models and easy interpretability as compared to many advanced “black-box-like" models. The proposed model has no assumptions on the distribution of the input and output variables, unlike parametric regresion models.

When applied to solve the process efficiency problem, we found a set of most important process parameters from the available data that controls process recovery percentage. Our proposed model is found useful for forecasting future process efficiency in terms of recovery percentage till the company is able to make suggested changes to the process for its improvement. To summarize, the contributions are:

- We propose a pipeline model, namely, RBNT model for improving waste recovery process based on decision tree and neural networks.
- Optimization of model parameters along with the regularity conditions for the universal consistency of the proposed model is discussed.
- The robustness of the proposed algorithm is shown through experimental evaluation over five publicly available regression datasets from various applied fields.

The paper is organized as follows. In Section 2, we describe the proposed RBNT model and its statistical properties are discussed in Section 3. Waste recovery improvement problem, model evaluation and comparisons are presented in Section 4. To show the robustness and general applicability of the proposed model, we evaluated its performance on the publicly available regression datasets in Section 5. Finally, the concluding remarks are given in Section 6.

2 | RADIAL BASIS NEURAL TREE (RBNT) MODEL

RT uses a hierarchical segmentation of the input feature spaces whereas RBFN is a particular type of ANN model that uses a radial basis function as an activation function. In the proposed RBNT model, we first split the input feature space into areas by RT algorithm (for detailed discussion see Section 3.1). Based on feature rankings provided by RT, a set of important features are chosen and extracted from the training dataset. We then build the RBFN model using the important variables obtained through RT algorithm along with the prediction results obtained from RT as another input information in the input layer of the network. The effectiveness of the proposed classifier lies in the selection of important features and use of prediction results of RT followed by the application of the RBFN model. The inclusion of RT output as an additional input feature not only improves the model accuracy but also increases class separability. The proposed RBNT regression model can handle high-dimensional datasets through the implementation of RT in selecting features as well as the incorporation of its predicted outputs tied up with one hidden layered RBFN model. This hybridization improves the performances of RT, RBFN as well as reduces the biases and variances of these individual models. The informal workflow of our proposed model is as follows:

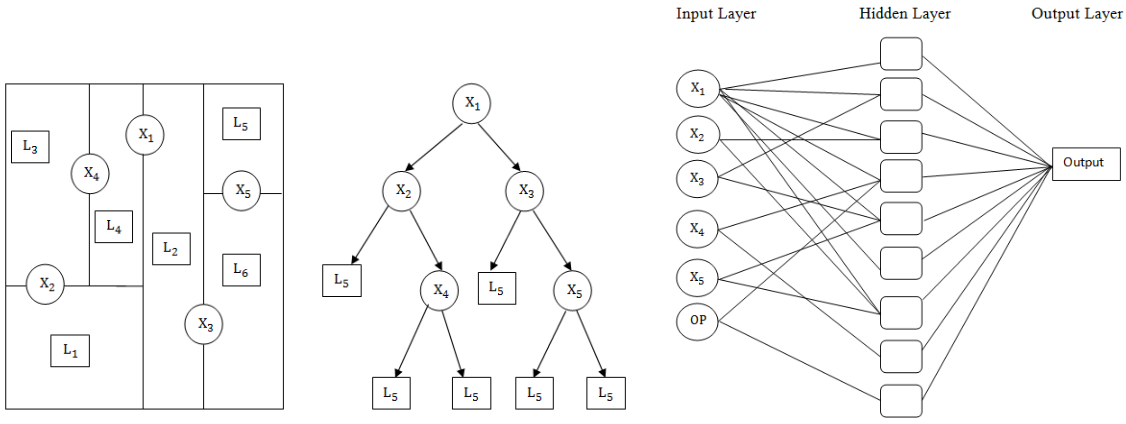

FIGURE 1 An example of RBNT with $X_i; i = 1, 2, 3, 4, 5$ as Important features obtained by RT, L_i as leaf nodes and OP as RT output. Creation of RT (Left) and one hidden layered RBFN model (Right).

- Apply RT algorithm to rank the feature, extracts important features and find the splits between different adjacent values of the features.
- Choose the features that have minimum mean squared error. We further grow unpruned RT and record its outputs. RT has an in-built important feature selection mechanism.
- Export important input variables along with additional feature (prediction values of RT algorithm) to the RBFN model and a neural network is generated.
- RBFN model uses a gaussian kernel as an activation function, and parameter optimization is done using a gradient descent algorithm (for detail discussion see Section 3.2). Finally, we record the final outputs.

RBNT model is a two-step pipeline approach such as choosing most important features among the set of available features and finally getting an improved prediction values. Since RT is robust in handling the curse of high-dimensional datasets, thus incorporation of its predicted class levels along with important features obtained by RT as input features in RBFN will necessarily improve the performance of the single classifiers. A flowchart of the RBNT model is presented in Figure 1.

3 | THEORETICAL PROPERTIES OF THE RBNT MODEL

The proposed RBNT model has the following architecture: (1) it extracts important features from the feature space using RT algorithm; (2) it builds one hidden layered RBFN model with the important features extracted using RT along with RT outputs as an additional feature. Now let us investigate the theoretical properties of the proposed model by introducing a set of regularity conditions for consistency of RT followed by the consistency results of RBFN algorithm. Further, an idea about parameter optimization in the RBNT model is proposed in Section 3.2.

3.1 | Regularity Conditions for Universal Consistency

Let $\underline{X}$ be the space of all possible values of d features, i.e. $\underline{X} = (X_1, X_2, ..., X_d)$ and $\underline{Y}$ be set of all possible values that the response variable can take and $Y \in [-K, K]$. We wish to estimate regression function $r(x) = E(Y|X = x) \in [-K, K]$ based on a training sample $L_n = \{(X_1, Y_1), (X_2, Y_2), ..., (X_n, Y_n)\}$, where $X_i = (X_{i1}, X_{i2}, ..., X_{id}) \in \underline{X}$ with n observations and let $\Omega = \{\omega_1, \omega_2, ..., \omega_k\}$ be a partition of the feature space $\underline{X}$. We denote $\widetilde{\Omega}$ as one such partition of Ω. Define $(L_n)_{\omega_i} = \{(X_i, Y_i) \in L_n : X_i \in \omega_i, Y_i \in [-K, K]\}$ to be the subset of L_n induced by ω_i and let $(L_n)_{\widetilde{\Omega}}$ denote the partition of L_n induced by $\widetilde{\Omega}$.

The criterion used to identify the best features involves mean squared error for regression problems. We don't make any assumption on the distribution of the pair $(\underline{X}, \underline{Y}) \in \mathbb{R}^d \times [-K, K]$. Mean squared error is used to partition feature space into a set $\widetilde{\Omega}$ of nodes. There exists a partitioning regression function $p : \widetilde{\Omega} \to Y$ such that p is constant on every node of $\widetilde{\Omega}$. Now let us define $\widehat{L}_n$ be the space of all learning samples and $\mathbb{D}$ be the space of all partitioning regression function then, binary partitioning and regression tree based rule $\Phi : \widehat{L}_n \to \mathbb{D}$ such that $\Phi(L_n) = (\psi \circ \phi)(L_n)$, where ϕ maps L_n to some induced partition $(L_n)_{\widetilde{\Omega}}$ and ψ is an assigning rule which maps $(L_n)_{\widetilde{\Omega}}$ to a partitioning regression function p on the partition $\widetilde{\Omega}$. A reasonable assigning rule ψ is the plurality rule $\psi_{pl}((L_n)_{\widetilde{\Omega}}) = p$ such that if $x \in \omega_i$, then

$$p(\underline{x}) = \arg \min_{Y_i \in [-K,K]} |\frac{1}{n} \sum_{i=1}^{n} (\Phi(L_n) - Y_i)^2|$$

The stopping rule in RT is decided based on the minimum number of split in the posterior sample called minsplit. If minsplit $\geq \alpha$ then ω_i will split into two child nodes and if minsplit $< \alpha$ then ω_i is a leaf node and no more split is required. Here α can be determined by the user, usually it is taken as 10% of the training sample size. Now we are going to evolve a sufficient regularity condition for the binary partitioning and regression tree based rule to be universally consistent.

Theorem 1. Suppose $(\underline{X}, \underline{Y})$ be a random vector in $\mathbb{R}^d \times [-K, K]$ and L_n be the training set of n outcomes of $(\underline{X}, \underline{Y})$. Finally if for every n and $w_i \in \tilde{\Omega}_n$, the induced subset $(L_n)_{w_i}$ contains at least k_n of $\underline{X}$, then empirically optimal regression trees strategy employing axis parallel splits are consistent when the size k_n of the tree grows as $o(\frac{n}{log(n)})$.

Proof. The proof of the theorem can be easily derived from the proof of consistency of the histogram regression estimation using data-dependent partitions [15]. □

Remark 1. Regression trees are consistent when the size of the tree grows as $o(\frac{n}{log(n)})$, where n is the number of training samples. It should be noted that the choice of important features based on RT is a greedy algorithm and the optimality of local choices of the best feature for a node doesn't guarantee that the constructed tree will be globally optimal [12].

Further, we build the RBFN model with RT extracted features and OP as another input feature in our model. The dimension of the input layer in the RBFN model, denoted by $d_m(< d)$ equals the number of important features obtained by RT + 1. Since

RBFN model consists of strictly one hidden layer, thus the proposed model is easily interpretable and fast in implementation. Our next objective is to discuss the sufficient condition for universal consistency of the RBFN model. After incorporating RT output in the feature space, we have n training sequence $\xi_n = \{(Z_1, Y_1), ..., (Z_n, Y_n)\}$ of n i.i.d copies of $(\underline{Z}, \underline{Y})$ taking values from $\mathbb{R}^{d_m} \times [-K, K]$, a regression estimate realized by a one-hidden layered neural network is chosen to minimize the empirical L_2 risk $\frac{1}{n} \sum_{j=1}^{n} \mid f(z_j) - Y_j \mid^2$.

RBFN is a family of ANNs, consists of only a single hidden layer and uses a nonlinear function called radial basis function as an activation function, unlike feed forward neural network. Figure 1 gives an example of RBFN with one hidden layer. RBFN is presented as a three-layered feed-forward structure where input layer distributes inputs to the hidden layer which contains neurons with nonlinear activation function[16]. Gaussian functions are most frequently used in this layer can be defined as follows:

$$\phi_i(z_i) = \phi\Big(\parallel z_i - c_i \parallel ; \sigma_i \Big) = exp\Bigg(- \frac{\parallel z_i - c_i \parallel^2}{2\sigma_i^2} \Bigg)$$

where z_i is an input vector, ϕ_i is the output of i^{th} hidden neuron in the hidden layer with centers c_i and σ_i^2 as the variance. Here we can select the center vector as cluster centers of the input data. For practical use, the number of clusters is generally chosen to be much smaller than the number of data points resulting in RBFN of less complexity than other types of ANNs. Let us now consider a RBF network with one hidden layer (an example is given in Figure 2) having k nodes for a fixed gaussian function is given by the equation:

$$f(z_i) = \sum_{j=1}^{k} w_j \ \phi\Big(\parallel z_i - c_j \parallel \Big) + w_0,$$

where $w_0, w_1, ..., w_k \in [-b, b]$ $(b > 0)$ and $c_1, c_2, ..., c_k \in \mathbb{R}^{d_m}$. The weights w_j and c_j are parameters of the RBFN and ϕ is the gaussian radial basis function with $\phi(z) \rightarrow 0 \ as \ x \rightarrow \infty$. The next Theorem due to krzyzak et al.[17] gives regularity conditions for the universal consistency of RBFN model.

Theorem 2. Consider a RBF network with one hidden layer having k nodes. If $k \rightarrow \infty, \ b \rightarrow \infty \ and \ \frac{kb^4 log(kb^2)}{n} \ \rightarrow \ 0$ as $n \rightarrow \infty$, then RBFN model is said to be universally consistent for all distribution of $(\underline{Z}, \underline{Y})$.

Remark 2. We can conclude that if the RBNT model satisfies the regularity conditions as stated in Theorem 1 and Theorem 2, then the proposed algorithm will be universally consistent. This is a fundamental property of any model for its robustness and general use. But computationally choosing parameters of RBFN by minimizing empirical L_2 risk will be very costly. In practice, parameters of the RBFN model are learned by a gradient descent algorithm as mentioned below.

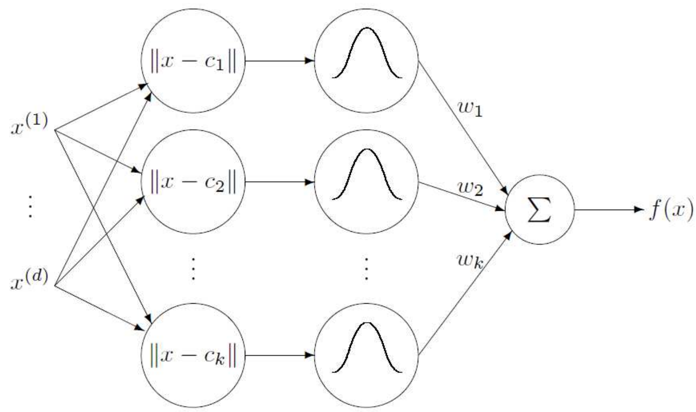

FIGURE 2 An example of Radial basis network with one hidden layer.

3.2 | Optimization of Model Parameters

Here we are going to discuss about the tuning parameters of the RBNT model. In the first stage of the pipeline model, 'minsplit' function (see Section 3.1) to be chosen as 10% of the training dataset is recommended as the stopping rule in RT algorithm. Further, we grow unpruned RT and use RT suggested features along with RT output as an additional feature in the input space of RBFN. RBFN is trained using a linear combination of gaussian basis functions. Therefore we need to use an optimization algorithm for empirical error (to be denoted by E in the rest of the paper) minimization on ξ_n [18]. Three important parameters to be optimized while training RBF network are: centers (c_i), standard deviation (σ_i) and weights (w_j) of each neuron. We use a gradient descent algorithm over E to perform the optimization task in the RBFN model [19], as follows:

$$\Delta c_i = -\rho_c \nabla_{c_i} E,$$

$$\Delta \sigma_i = -\rho_\sigma \frac{\partial E}{\partial \sigma_i},$$

$$\Delta w_j = -\rho_w \frac{\partial E}{\partial w_j},$$

where ρ_c, ρ_σ and ρ_w are small positive constants. Using this the parameters of the gaussian basis function will be optimized. Generalization error is usually estimated by cross-validation method, and the optimum value of k would be found by trial and error.

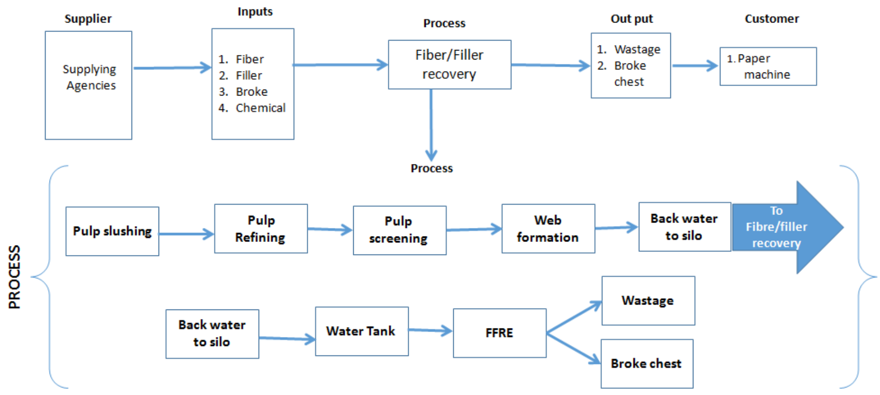

FIGURE 3 SIPOC diagram of fiber and filler recovery process

4 | APPLICATION TO WASTE RECOVERY PROBLEM

A modern paper manufacturing company is facing a problem of fiber and filler losses which account for a substantial monetary loss for the company. The objective of this work is to improve the waste recovery process by improving the percentage of recovery of the FFRE which will reduce the fiber and filler losses. The inlet to FFRE is the waste (lean backwater combined with some valuable materials) coming out of the paper machine which undergoes froth flotation when treated with chemical and air[20,21]. The chemical helps to form the flocks from the fines present in the lean backwater and air helps to form the cake over the FFRE for recovery. The cake is then scooped out, collected and sent back to the paper machine process stream. Remaining water is sent to water level tank for further to clarification. To give a better idea about the process, we presented the SIPOC diagram of fiber filler recovery process in Figure 3.

FFRE removes solids using air floatation and sedimentation. Turbulence caused by water movement is an important factor in floatation and greatly reduces the efficiency of the other types of floatation units. Conventionally there must always be water movement for the water to flow from inlet to outlet. In FFRE, the inlet and outlet of the process are not stationary, and it rotates about the center. The rotation is synchronized so that the water in the water level tank achieves zero velocity during floatation[22]. In practical terms, this allows better clarification in smaller surface areas and a much shallower tank. The processing time of water from the inlet to outlet takes around 2-3 minutes. Air is dissolved into the water using air dissolving tube (ADT) of FFRE[1], and unclarified water is released through a valve. The water flows in at the exact center, through a rotary joint. Coarse air is discharged through a vent pipe in the duct. The flow is directed to eliminate turbulence. Since the inlet distribution is moving forward at the same speed than the water is flowing out, the water stays in one spot in the tank without any movement during floatation. The floated material is recovered from the top surface through the FFRE spiral scoop. The scoop is designed to remove the floated material at the highest possible consistency, with a minimum surface disturbance. The level of water determines the consistency of the floated material removed. The floatation system in FFRE removes the solid content in the water by floating them to the

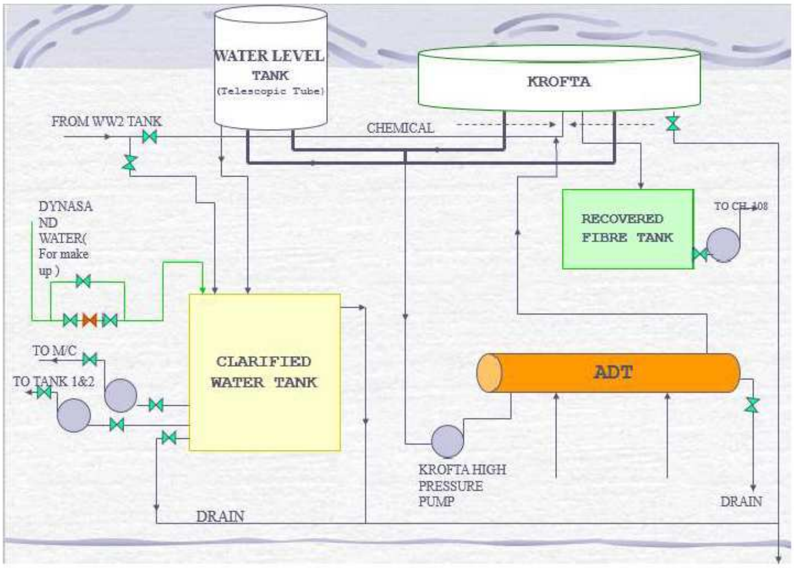

FIGURE 4 Schematic diagram of fiber and filler recovery process

TABLE 1 Sample data set

Sl. No.	Inlet Flow	Water Pressure	Air Pressure	Air-Left	Air-Right	ADT-D Left	ADT-D Right	Amount of chemical	Recovery Percentage
1	1448	6.4	5.8	1.0	2.1	3.2	4.0	2.0	96.80
2	1794	5.2	5.6	2.4	1.6	3.6	4.0	3.0	97.47
3	2995	6.0	6.0	1.5	4.5	4.0	4.8	4.0	28.87
4	1139	6.5	6.0	1.2	1.7	3.0	4.6	2.0	33.05
5	2899	6.2	5.7	2.0	1.2	3.1	4.0	2.0	97.91
6	1472	6.6	6.8	3.7	3.1	5.2	4.8	4.0	57.77
7	1703	6.2	6.0	2.9	1.0	3.0	4.2	2.0	26.94
8	1514	5.5	5.0	2.0	2.1	3.8	4.7	2.0	67.01
.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.

surface for removal. Chemicals are used to increase the clarifier efficiency by flocking out small and colloidal particles, that otherwise would not float or settle in the clarifier. The schematic diagram of the FFRE process is shown in Figure 4.

4.1 | Data

To identify the causal parameters affecting the FFRE efficiency, a cause and effect analysis is done for revealing the key relationship among causal variables with FFRE efficiency. A list of controllable parameters is prepared and accordingly we collected the data for the paper tissue from the process. The dataset collected for a year from the process on the following causal variables: Inlet Flow, Water Pressure (water inlet pressure to ADT), Air Pressure, Pressure of Air-Left, Pressure of Air-Right, Pressure of ADT-D Left, Pressure of ADT-D Right and Amount of chemical lubricants. The response variable (FFRE recovery percentage) lies between 20 to 100. Sample data set for the paper tissue is presented in Tables 1. This dataset will be used for finding crucial process parameters and also finding a prediction model that can help the company for forecasting future recovery percentage of FFRE.

4.2 | Performance metrics

We are going to use root mean square error (RMSE), mean absolute percentage error (MAPE), co-efficient of multiple determination (R^2) and adjusted R^2 ($AdjR^2$) to evaluate the performance of different regression models:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2};\ MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right|;$$
$$R^2 = 1 - \left[\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2}\right];\ AdjR^2 = 1 - \left[\frac{(1-R^2)(n-1)}{n-k-1}\right];$$

where, $y_i, \overline{y}, \hat{y}_i$ denote the actual value, average value and predicted value of the dependent variable, respectively for the i^{th} instant. Here n and k denote the number of data points and independent variables used for performance evaluation, respectively. The lower the value of RMSE and MAPE and the higher the value of R^2 and Adj R^2, the better is the model is.

4.3 | Analysis of Results

Here we elaborate about the experimental evaluation of the proposed RBNT model while applied to the waste recovery process (WRP) dataset. We first shuffled the observations of the WRP dataset randomly and split them into training and testing data in a ratio of 60 : 40. Each experiments are repeated 10 times with different randomly assigned training and test sets. We have considered here six well known machine learning models for comparison purpose. All the competing methods are implemented using standards libraries in R on a PC with 2.1 GHz processor and 8 GB memory. These algorithms have been executed with their default parameters. The performance of each model was evaluated with five statistical measures (viz. MAE, RMSE, MAPE, R^2 and Adj(R^2)) based on the prediction values relative to the observed values.

The FFRE efficiency corresponding to a tissue depends on several process features such as Water Pressure, Air Pressure, Inlet Flow, AIR-Left, AIR-Right, amount of chemical, ADT-D Left and ADT-D Right. Our task is to select the important features out the above and apply the model to improve the recovery percentage. We first apply the RT method to select the important FFRE features and then compute the RBFN model with the selected important features along with RT output as shown in Figure 1. RT is trained using *rpart* package implementation in R that uses gini index, and C_p for selection of variables to enter and leave the tree structure. To implement neural nets, we first standardize the datasets and run the ANN model. A transfer function ("logsig") is applied to bring back the original form at the end of modeling. We implemented the ANN model with different combinations of hidden layers without employing any other feature selection algorithm using *neuralnet* package. Due to the fact that the dataset is medium-sized, thus going beyond 2 hidden layered (2HL) neural net will overfit the dataset. The number of neurons in 1st hidden layer are chosen as 2/3 the size of the input layer and the number of neurons in 2nd hidden layer are chosen as 1/3 the size of the input layer. RBFN with gaussian radial basis kernel was trained using *RSNNS* package and the maximum

TABLE 2 Quantitative measures of performance for different regression models.

Regression Models	MAE	RMSE	MAPE	R^2	Adj(R^2)
RT	11.691	16.927	29.010	59.028	55.304
ANN	12.334	17.073	27.564	58.310	54.529
SVR	12.460	20.362	40.010	40.174	35.325
BART	12.892	16.010	30.038	59.380	56.458
RBFN	13.926	18.757	32.48	49.689	46.335
Neural tree	10.895	16.012	24.021	65.120	62.946
RBNT	**9.226**	**14.331**	**20.187**	**70.632**	**68.675**

number of iterations in all these NN implementations are chosen as 100. Execution time for RBFN model is much less than ANN but higher than RT.

The essential features for FFRE recovery dataset are: Water Pressure, Air Pressure, Inlet Flow, ADT-D Left and ADT-D Right. Table 2 exhibits the quantitative measures corresponding to MAE, RMSE, MAPE, R^2 and Adj(R^2) respectively for the data set and for all the competing methods. A method in predicting the data is said to perform well if it is minimize the MAE, RMSE, MAPE values and maximize the R^2 and Adj(R^2) values between the predictions relative to the observations. It can be observed from Table 2 that the MAE, RMSE and MAPE values obtained through the proposed model is always smaller than the values obtained from all other competing methods over the WRP dataset. Further, we compare our proposed model with different regressions models, viz. RT, ANN, RBFN, SVR, bayesian adaptive regression tree (BART) and neural tree. Neural tree provides the second best performance than the remaining models in terms of having low RMSE, MAE and MAPE values. Based on the recommendation of the model, we conducted design of experiments to find optimal level of essential parameters to improve FFRE efficiency. Also, the paper company can use our model for future prediction of recovery percentage.

5 | OTHER EXPERIMENTS

We now compare the RBNT models with standard RT, standard ANN, BART, etc. on five regression data sets from UCI Machine Learning Repository [23]. Nonnumerical features and samples with missing entries were systematically removed from the datasets. A brief summary of these datasets are available in Table 3.

To showcase the abilities of the RBNT, we picked diverse, but mostly small and medium-sized regression datasets. These are the sets where RT often outperforms NN, since the latter typically require plentiful training data to perform reasonably well. It is on these particular types of datasets where the benefits of neural optimization can usually not be exploited, but with the specific inductive bias of the proposed RBNT model it becomes possible. These are such types of datasets in which prediction is not the only task but also feature selection plays a vital role. To start, we first shuffled the observations in each of the five different

TABLE 3 Data set description: number of samples and number of features, after removing observations with missing information or nonnumerical input features.

Data set	Number of samples	Number of features
Auto MPG	398	7
Concrete	1030	8
Forest Fires	517	10
Housing	506	13
Wisconsin	194	32

TABLE 4 Average RMSE results for each of the models across the different datasets

Data	RT	ANN	SVR	BART	RBFN	Neural Tree	RBNT
Auto MPG	3.950	4.260	5.720	3.220	4.595	3.300	**3.215**
Concrete	8.700	10.180	11.588	**5.540**	10.210	7.420	7.063
Forest Fires	75.138	90.702	91.985	65.890	82.804	**62.478**	64.411
Housing	4.980	9.054	12.520	3.978	7.871	4.590	**3.077**
Wisconsin	41.059	34.710	41.220	32.054	38.495	40.700	**23.659**

datasets randomly and split them into training and test datasets in a ratio of 60 : 40. We have repeated each of the experiments 10 times with different randomly assigned training and test sets.

Table 4 summarizes the results in terms of RMSE for the different models. The first important comment is that the proposed model improves over the original RT, ANN, RBFN, SVR models consistently across all data sets. Proposed RBNT model outperforms all the similar types of machine learning models for Auto MPG, Housing, and Wisconsin dataset and found competitive on Concrete and Forest Fires dataset as well.

6 | CONCLUSIONS

In this paper, we proposed a novel distribution-free hybrid regression model which is a hybridization of RT and RBFN model for improving waste recovery process in a paper company. The proposed model demonstrated the best performance and offered a practical solution to the WRP problem of finding crucial process parameters to improve recovery percentage of fiber filler recovery equipment. The model can also be useful for future recovery prediction of the process that may give the company an indication about the stability of the process. Using the recommendation of the model, we further designed a set of experiments to find the optimal level of essential process parameters (as found in Section 4.3) causing the FFRE efficiency. This resulted in improved waste recovery from an existing average recovery percentage of 60% to 80% which added financial benefit to the paper company. Our proposed model takes into account the curse of high-dimensional data and performs well on small or medium-sized datasets. Through computational experiments, we have shown our proposed methodology performed quite well as compared to the other state-of-the-arts in the tissue paper dataset as well as five publicly available datasets. The proposed

RBNT model has the most desired statistical property, viz. universal consistency. The usefulness and effectiveness of the model lie in its robustness and easy interpretability as compared to complex “black-box-like" models. An immediate future work of this paper is to investigate the applicability of the proposed model on spatio-temporal datasets involving regression tasks.

References

1. Krofta M. Three zone dissolved air flotation clarifier with improved efficiency. 1998. US Patent 5,846,413.

2. Krofta M. Apparatus for clarification of water. 1986. US Patent 4,626,345.

3. Shrimali M, Singh K. New methods of nitrate removal from water. *Environmental pollution* 2001; 112(3): 351–359.

4. Mahuli S, Rhinehart R, Riggs J. pH control using a statistical technique for continuous on-line model adaptation. *Computers & chemical engineering* 1993; 17(4): 309–317.

5. Gmar S, Helali N, Boubakri A, Sayadi IBS, Tlili M, Amor MB. Electrodialytic desalination of brackish water: determination of optimal experimental parameters using full factorial design. *Applied Water Science* 2017; 7(8): 4563–4572.

6. Bhattacharya B, Solomatine DP. Neural networks and M5 model trees in modelling water level–discharge relationship. *Neurocomputing* 2005; 63: 381–396.

7. Brentan BM, Luvizotto Jr E, Herrera M, Izquierdo J, Pérez-García R. Hybrid regression model for near real-time urban water demand forecasting. *Journal of Computational and Applied Mathematics* 2017; 309: 532–541.

8. Sebri M. Forecasting urban water demand: A meta-regression analysis. *Journal of environmental management* 2016; 183: 777–785.

9. Cancho VG, Suzuki AK, Barriga GD, Louzada F. A non-default fraction bivariate regression model for credit scoring: An application to brazilian customer data. *Communications in Statistics: Case Studies, Data Analysis and Applications* 2016; 2(1-2): 1–12.

10. Lee TS, Chen IF. A two-stage hybrid credit scoring model using artificial neural networks and multivariate adaptive regression splines. *Expert Systems with Applications* 2005; 28(4): 743–752.

11. Breiman L. *Classification and regression trees*. Routledge . 2017.

12. Kuncheva LI. *Combining pattern classifiers: methods and algorithms*. John Wiley & Sons . 2004.

13. Sethi IK. Entropy nets: from decision trees to neural networks. *Proceedings of the IEEE* 1990; 78(10): 1605–1613.

14. Sirat J, Nadal J. Neural trees: a new tool for classification. *Network: Computation in Neural Systems* 1990; 1(4): 423–438.

15. Nobel A, others . Histogram regression estimation using data-dependent partitions. *The Annals of Statistics* 1996; 24(3): 1084–1105.

16. Györfi L, Kohler M, Krzyzak A, Walk H. *A distribution-free theory of nonparametric regression*. Springer Science & Business Media . 2006.

17. Krzyzak A, Linder T, Lugosi C. Nonparametric estimation and classification using radial basis function nets and empirical risk minimization. *IEEE Transactions on Neural Networks* 1996; 7(2): 475–487.

18. Poggio T, Girosi F. Regularization algorithms for learning that are equivalent to multilayer networks. *Science* 1990; 247(4945): 978–982.

19. Karayiannis NB. Reformulated radial basis neural networks trained by gradient descent. *IEEE transactions on neural networks* 1999; 10(3): 657–671.

20. Marques GA, Tenório JAS. Use of froth flotation to separate PVC/PET mixtures. *Waste Management* 2000; 20(4): 265–269.

21. Aldrich C, Marais C, Shean B, Cilliers J. Online monitoring and control of froth flotation systems with machine vision: A review. *International Journal of Mineral Processing* 2010; 96(1-4): 1–13.

22. Zhou Z, Xu Z, Finch J. On the role of cavitation in particle collection during flotation-a critical review. *Minerals Engineering* 1994; 7(9): 1073–1084.

23. Asuncion A, Newman D. UCI machine learning repository. 2007.

□