Selection of Optimum Features using PSO-SPG2 for Predicting COD Effluent Level in Wastewater

Vijayabhanu Rajagopal [#1], V.Radha [#2]

Department of Computer Science ,
Avinashilingam Institute for Home Science and Higher Education for Women,
Coimbatore, TamilNadu, India
vijayabhanuresearch@gmail.com
radharesearch@yahoo.co.in

***Abstract*—**

Wastewaters are waterborne solids and liquids which get discharged into sewers representing the wastes of public life. Wastewater Treatment is the process of removing contaminants from both runoff and domestic wastewater. Chemical Oxygen Demand (COD) is used to measure the degradable organic matter confined in wastewater effluents. An anaerobic process becomes popular method for the treatment of a number of waste streams. Up-flow Anaerobic Filter (UAF) is an anaerobic reactor which is used for the removal and digestion of organic matter present in wastewater. A systematic prediction of agro-food wastewater using Adaptive Neuro-Fuzzy Inference System (ANFIS) was proposed to predict COD effluent of a full-scale anaerobic wastewater treatment plant in accordance with UAF. Wastewater Treatment process initiates with Data collection and pre-processing preceded by normalization technique using Z-Score, which is followed by Feature selection algorithm and the prediction of COD is concluded using ANFIS by the model when applied to the description of UAF treating agro-food industrial wastewaters such as fruit juice wastewater. Prediction of agro-food wastewater can be improved using various feature selection algorithms such as Genetic Algorithm (GA), Particle Swarm Optimization (PSO) and Particle Swarm Optimization (PSO) With Spectral Projected Gradient (SPG2) algorithm. The experimental results proved that Combination of ANFIS with PSO-SPG2 is considered as a best fit to predict the COD effluent level with greater accuracy and with low Mean Square Error values.

***Keywords*--- Wastewater Treatment, COD, Z-Score Normalization, Genetic algorithm, PSO, SPG2, ANFIS.**

I. INTRODUCTION

Water is one of the useful natural resources and it is facing several problems with the fast development of large number of industries. Almost all the industries discharge the wastewater into any pure water bodies without treating it. This is turned out to be a biggest threat to human lives and many other aquatic organisms. Hence it is necessary to treat or purify the water before discharging it into other water bodies. COD is an essential test for evaluating the quality of effluents and wastewaters prior to discharge.

Anaerobic wastewater treatment differs from traditional aerobic treatment in that no aeration is used. The proposed method uses anaerobic treatment process to utilize anaerobic bacteria (biomass) to convert organic pollutants or COD into biogas in an oxygen free environment. Anaerobic digestion is a process in which organic matter is decomposed and bio-converted into biogas and microbial biomass in the absence of oxygen. Anaerobic digestion is basically an unstable process and variations of the input variables such as hydraulic flow rate; influent organic load etc. makes the process to be complex. Though, anaerobic digestion can be considered as a convenient method under necessary conditions if adequate monitoring is performed and suspiciously designed control strategies are implemented [8].

UAF is used as an anaerobic reactor in the proposed method. The effect on the environment of agro-food production (particularly due to water emissions) is the major issue which has received attention in the recent years. Wastewater from agro-food production contains a relevant concentration of hardly biodegradable compounds phyto-toxic chemicals [9]. The proposed method uses fruit juice wastewater.

Wastewater Treatment training on real dataset could be made more efficient by performing certain pre-processing steps on influents and target COD level. The normalization process for the raw dataset has significant effect on preparing the data to be suitable for the training. Feature selection allows reducing the future space and helps in improving the accuracy by reducing training time. The main goal of FSA is to determine the relevant features of the dataset based on relevance. In Feature selection process, subsets features available from the data are selected to apply the learning algorithm. The best subset selection of data has the least number of dimensions

which contributes to accuracy [5]. Prediction is a form of data analysis that can be used to extract models to predict future data trends. This type of data analysis helps to provide better understanding of the large data set. Proposed method uses prediction algorithms to find the COD effluent level in the fruit juice wastewater.

Fruit juice is one of the principal fruits which is canned. Most fruit juices which are canned are produced in Thailand, Philippines and Indonesia. These countries export 775x103 tons of canned Fruit juice annually . The cans of fruit juice produced in the world in a single year contain approximately 303 x103 tons of syrup, which contains glucose, fructose and sucrose, and can be converted into useful materials such as lactic acid by microbial systems.

Paper can be organized as follows: Section II describes about the algorithms used in the proposed methodology. Section III deals with the results obtained during experiment. The concluding work is presented in section IV.

II. METHODOLOGY

Automated wastewater treatment process is employed in the proposed method inorder to reduce the time consumption in detecting the COD effluent level in wastewater and to predict the COD effluent level in wastewater with greater accuracy. COD effluent level in wastewater can be formulated using following steps.

2.1 DATA PRE-PROCESSING

Missing values can arise from any stages of the experiment such as the data collection & preprocessing till prediction (Scheel et al. 2005). Hence it is crucial to obtain accurate and complete dataset. The feasible solution to recover the dataset with missing value is to repeat experiments whose dataset values are inefficient and infeasible. Therefore, the missing values in the original data should be predicted mathematically and the data must be estimated before further processing.

Accurate prediction of missing values guarantees the reliability of consequent steps. Simple allegations were used in the traditional approaches to fill the missing values with zeros using the row mean for computation. These methods produce inaccurate estimating values and influence the results of feature selection and prediction. Several complicated approaches have been proposed to predict the missing values.

Each sample in the raw dataset comprises certain parameters which are non linear. Usually real-world databases contain unknown values, outliers or missing values and erroneous data. There will be some missing values and outliers being accommodated in the multidimensional historical database of an UAF. In the anaerobic data, the total samples based on the effluent COD when using COD, Volatile Suspended Solid, Total Suspended Solid as influent into the wastewater treatment plant and it is treated in the UAF reactor. First, the raw data is subjected to an analysis using K-Nearest Neighbor algorithm (K-NN), a Data Mining Technique, in order to eliminate redundant and missing values in the input dataset.

$$Z = \frac{X - \mu}{\sigma} \qquad (1)$$

Where μ represents the arithmetic mean and σ represents the standard deviation of the given data.

The quantity z indicates the distance between the raw score and the population mean in units of the standard deviation. The value of Z is negative when the raw score is below the mean and Z value is positive when the raw score is above the mean. If μ and σ are not known they can be estimated from the dataset. Each feature will have the mean value 0 after z-score normalization. The unit of each value will be the number of standard deviations which will be far away from the mean. Z-score normalization will give exact values to the small values of σ.

2.2 FEATURE SELECTION

Feature selection provides a better understanding of data by offering their significant characteristic features. Feature selection abolishes unrelated and unnecessary words of text and thus minimizes the dimensionality of documents. A good feature selection approach can computationally enhance the learning algorithms.

2.2.1 Genetic Algorithm

Genetic Algorithms (GA) are search algorithms which are based on the concepts of natural selection and natural genetics. Genetic algorithm was enlarged to replicate some of the processes examined in natural evolution. The genetic algorithm searches along with a population of points and works with a coding of parameter set, instead of the parameter values. It utilizes objective function information without any gradient information [10]. Hence the transition scheme of the genetic algorithm is probabilistic. GA also provides means to search irregular space and thus it can be used for a variety of parameter estimation, function optimization, and machine learning applications. Because of these features genetic algorithm is employed in the proposed work for feature selection to improve the quality and speed in determining COD effluent level present in the wastewater.

The major steps involved in the genetic algorithm are the generation of a population of solutions, finding the objective function and fitness function. The procedure for genetic algorithm is described as follows:

Genetic Algorithm
Formulate initial population
Randomly initialize population
Repeat
Evaluate objective function
Find fitness function
Apply genetic operators
Reproduction
Crossover
Mutation
Until stopping criteria

Procedure for Genetic Algorithm

Consider the maximization problem

$$\text{Maximize} \qquad (2)$$
$$f(x) = x_i^l \leq x_i \leq x_i^u, i = 1,2, \ldots . n$$

Where x_i^l and x_i^u are the lower and upper bound of the variable. According to the following linear mapping rule, the string can be found to represent a point in the search space.

$$x_i = x_i^l + \frac{x_i^u - x_i^l}{2^\beta - 1} \qquad (3)$$

The variable x_i is coded with sub-string s_i of length. The decoded value of a binary sub-string s_i is computed as $\sum_{j=0}^{\beta} \gamma_j 2^j$ where $s_i \in (0,1)$ and the string s is represented as $s_{\beta-1}, s_{\beta-2} \ldots \ldots s_2, s_1, s_0$. The length of a sub-string representing a variable depends on the desired accuracy in that variable. It varies with the desired precision of the results, such that longer the string length, the more will be the accuracy. The relationship between string length β and precision α is given by

$$(x_i^u - x_i^l)\, 10^\alpha \leq (2^\beta - 1) \qquad (4)$$

The corresponding point $x = (x_1, x_2, \ldots . x_N)^T$ can be found if the coding of the variables is complete. The function value at the point x can also be calculated by substituting x in the given objective function f(x).

A fitness function F (i) is obtained from the objective function and used in consecutive genetic operations. Fitness is a measure of the reproductive efficiency of chromosomes. Fitness function is used to allocate reproductive attributes to the individuals in the population and an individual which has higher fitness value will have higher probability of being selected as candidates for further assessment [6]. Fitness function can be considered to be the same as the objective function for maximization problems i.e. $F(i) = O(i)$. In case of minimization problems, it is necessary to map the essential natural objective function to fitness function form inorder to create non-negative values in all the cases and to reproduce the comparative fitness of individual string.

$$F(x) = \frac{1}{1 + f(x)} \qquad (5)$$

This transformation does not change the position of the minimum, but it converts a minimization problem to an equivalent maximization problem. It can be also done using the following equation,

$$F(i) = V - \frac{O(i)P}{\sum_{i=1}^{p} O(i)} \qquad (6)$$

Where O(i)the objective function value of i^{th} individual, P is is the population size and V is a large value to ensure non-negative fitness values. The population is then operated by three main operators such as reproduction, crossover and mutation to create a new population of points. The function of these operators is to generate new solution vectors by selection, alteration or combination of the current solution vectors which has shown to be good temporary solutions. The new population is estimated and tested till termination.

Reproduction (or selection) operator creates several reproductions of improved strings in a new population. A crossover operator is used to recombine two strings to get a better string. Recombination process produces dissimilar individuals in the successive generations by joining substance from two individuals of the previous generation. Mutation inserts new information in a random way to the genetic search process and eventually assists to evade getting trapped at local optima.

Application of these operators on the current population creates a new population which is used to generate subsequent populations and so on, producing solutions that are closer to the optimum solution. The values of the objective function of the individuals of the new population are again determined by decoding the strings which express the fitness of the solutions of the new generations. It completes one cycle of genetic algorithm called a generation. In every generation, the improved solution is stored as the best solution which is repeated till convergence.

2.2.2 PARTICLE SWARM OPTIMIZATION METHOD

A population of random solutions is initialized to PSO called 'particles'. Each particle can be treated as a point in an S-dimensional space [3]. The ith particle is represented as Xi = (xi1,xi2,…..,xis).The best previous position (pbest, the position giving the best fitness value) of any particle is recorded and represented as Pi = (pi1,pi2,…..,pis). The index of the best particle among all the particles in the population is represented by the symbol 'gbest'. The rate of the position change (velocity) for particle i is represented as Vi = (vi1,vi2,…..,vis) . The particles are manipulated according to the following equation:

$$v_{id} = w * v_{id} + c_1 * \text{rand} * (p_{id} - x_{id}) + c_2 * \text{rand} * (p_{gd} - x_{id}) \quad (7)$$

$$x_{id} = x_{id} + v_{id} \quad (8)$$

where d = 1,2,..., S , w is the inertia weight which is a positive linear function of time changing according to the generation iteration. Appropriate selection of the inertia weight provides a balance between global and local exploration. This results in less iteration on average to measure the most favourable solution. The acceleration constants c1 and c2 in equation (7) indicates the weighting of the stochastic acceleration terms which pull each particle towards the pbest and gbest regions [11]. High values result in abrupt movement towards target regions while low values allow particles to roam away from the target regions before being pulled back. rand() and Rand() are two random functions in the range [0,1].

Particle Swarm Optimization

Inputs: m - the swarm size;
c1,c2 - positive acceleration constants;
w - inertia weight;
MaxV - maximum velocity of particles;
MaxGen - maximum generation;
MaxFit - maximum fitness value;
Output: Pgbest - Global best position
Algorithm:
Begin
Swarms $\{x_{id}, v_{id}\}$ =Generate(m);
/* Initialize a population of particles with random positions and velocities on S dimensions*/
Pbest(i)=0; i = 1,...,m, d = 1,...,S
Gbest=0; Iter=0;
While(Iter<MaxGen and Gbest<MaxFit)
{ For(every particle i)
{ Fitness(i)=Evaluate(i);
IF(Fitness(i)>Pbest(i))
{Pbest(i)=Fitness(i); pid = xid; d = 1,...,S }
IF(Fitness(i)>Gbest)
{Gbest=Fitness(i); gbest=i;}
}
For(every particle i)
{ For(every d)
{

$$v_{id} = w * v_{id} + c_1 * \text{rand}() * (p_{id} - x_{id}) + c_2 * \text{rand}(* (p_{gd} - x_{id})$$

IF(vid > MaxV) { vid = MaxV ;}
IF(vid < - MaxV) { vid = - MaxV;}

$$x_{id} = x_{id} + v_{id}$$

}
}
Iter=Iter+1;
}/*rand() and Rand() are two random functions in the range [0,1]*/
Return P_{gbest}
End

Procedure for PSO Algorithm

Equation (7) is used to calculate the particle's new velocity according to its previous velocity and the distances of its current position from its own best experience (position) and the group's best experience. According to equation (8), the particle flies toward a new position.

The performance of each particle is measured according to a fitness function. The fitness function is defined as the prediction quality of the extracted rules. Finally, the features or attributes within rules with highest indices are selected. Fitness function is given by,

$$\text{Fitness} = \alpha * \gamma_R(D) + \beta * \frac{|C|-|R|}{|c|} \qquad (9)$$

Where $\gamma_R(D)$ is the prediction quality of condition attribute set R relative to decision D, |R| is the '1' number of a position or the length of selected feature subset, |C| is the total number of features, α and β are two parameters corresponding to the importance of prediction quality and subset length, $\alpha \in [0,1]$ and $\beta = 1 - a$ .This formula indicates that the prediction quality and feature subset length have different significance for feature selection task.

PSO is easy to implement and there are few parameters to adjust. Thus by using PSO as a feature selection reduces noise in the dataset to improve classification accuracy.

2.2.3 SPG2 METHOD

Spectral gradient methods are non-monotone schemes that have recently received considerable attention in the numerical analysis and optimization literature. SPG method generally used to solve the problem

$$\text{Minimize } f(x) \text{ subject to } x \quad \Omega \qquad (10)$$

where f admits continuous first derivatives and $\Omega \subset IR^n$ closed and convex. A point $x \in \Omega$ is stationary if $\nabla f(x)^T d \geq 0$ for all $d \in IR^n$ such that $x + d \in \Omega$.

SPG method has been used as an "Inexact Variable Metric" method for solving (10). Let IB be the set of $n \times n$ positive definite matrices such that $\|B\| \leq L$ and $\|B^{-1}\| \leq L$. Thus, IB is a compact set of $IR^{n\times n}$. In the spectral gradient approach, the matrices will be of the form σI, with $\sigma \in [\sigma_{min}, \sigma_{max}]$.

Inexact Variable Metric Method

Input:Assume $\eta \in (0,1], \gamma \in (0,1), 0 < \tau 1 < \tau 2 < 1, M \geq 1$ an integer number.

Let $x_0 \in \Omega$ be an arbitrary initial point. Consider $g_k = \nabla f(x_k)$ for all $k = 0, 1, 2, \ldots$. Given $x_k \in \Omega$, $B_k \in IB$.

Procedure for kth iteration

Step 1. Compute the search direction

Consider the subproblem

Minimize $Q_k(d)$ subject to $x_k + d \in \Omega$ (a)

where

$$Q_k(d) = \frac{1}{2} d^T B_k d + g_k^T d.$$

Let $\overline{d}_k$ be the minimizer of (10) (This minimizer exists and is unique by the strict convexity of the subproblem, but does not need to be computed).

Let d_k be such that $x_k + d_k \in \Omega$ and

$$Q_k(d_k) \leq \eta\, Q_k(\overline{d}_k).$$

If $d_k = 0$, stop the execution of the algorithm declaring that x_k is a stationary point.

Step 2. Compute the steplength

Compute $f_{max} = \max\{f(x_{k-j}) \mid 0 \leq j \leq \min\{k, M-1\}\}$, $\delta \leftarrow (g_k, d_k)$ and set $\alpha \leftarrow 1$.

$$f(x_k, \alpha d_k \leq f_{max} + \gamma\alpha\delta \qquad (b)$$

Set $\alpha_k = \alpha$, $x_{k+1} = x_k + \alpha_k d_k$ and finish the iteration. Otherwise, choose $\alpha_{new} \quad |\tau_1\alpha, \tau_2\alpha|$, set $\alpha \leftarrow \alpha_{new}$ and repeat test (b).

Procedure for Inexact Variable Metric Method

The possibility $\eta = 1$ corresponds to the case in which the subproblem (a) is solved exactly. The Spectral Projected Gradient method is used to minimize a real smooth function 'f' on a closed and convex set Ω. The method, as well as its unconstrained counterpart [8] is defined as

$$x_{k+1} = x_k + \alpha_k d_k \qquad (11)$$

where the scalar α_k is the step length, x_k denotes the current iterate and d_k is the descent direction. The search direction d_k has been defined in [1] as

$$d_k = P\left(x_k - \frac{1}{\sigma_k}\nabla f(x_k)\right) - x_k \qquad (12)$$

where P denotes the Euclidean projection on Ω. If $d_k \neq 0$, then $f(x_k + \alpha d_k) \ll f(x_k)$ for α small enough. In principle, one could define convergent methods imposing sufficient decrease at each iteration. But it leads to disastrous practical results. Due to this reason, the spectral methods employ a non-monotone line search which does not impose functional decrease at each iteration.

The complete line search procedure is explained in the following algorithm.

Line Search Using SPG2

Compute $f_{max} = \max\{f(x_{k-j}) \mid 0 \le j \le \min\{k, M-1\}\}$, $x_+ \leftarrow x_k + d_k$, $\delta \leftarrow (\nabla f(x_k)), d_k)$ and set $\alpha \leftarrow 1$.

Step 1: Test nonmonotone Armijo-like criterion

If $f(x_+) \le f_{max} + \gamma\alpha\delta$ then set $\alpha \leftarrow \alpha$ and stop.

Step 2:Compute a safeguarded new trial steplength

Compute $\alpha_{tmp} \leftarrow -\frac{\frac{1}{2}\alpha^2\delta}{f(x_+)-f(x_k)-\alpha\delta}$

If $\alpha_{tmp} \ge \sigma_1$ and $\alpha_{tmp} \ge \sigma_2\alpha$ then set $\alpha \leftarrow \alpha_{tmp}$.

Otherwise, set $\alpha \leftarrow \frac{\alpha}{2}$.

Step 3: Compute $x_+ \leftarrow x_k + \alpha d_k$ and goto step 1.

Procedure for Line Search Algorithm Using Spectral Projected Gradient Method

If the first trial point is rejected, then the next ones are computed along the same direction. As a result, the projection operation is performed only once per iteration. The line search depends on a safeguarded quadratic interpolation and aims to satisfy an Armijo-type criterion with a sufficient decrease parameter γ (0, 1). When the minimum of the one-dimensional quadratic lies outside [τ1, τ2α], then the safeguarding procedure acts and not when it lies outside [τ1α, τ2α]. This means that when interpolation tends to reject 90% (for σ1 = 0.1) of the original search interval ([0, 1]), then that prediction is not reliable and the more conservative bisection should be preferred.

2.2.4 **HYBRID PSO-SPG METHOD***:*

Proposed method is obtained by hybridizing the PSO and the SPG methods.PSO contributes to hybrid approach in a way to ensure extrapolation of search space, while the SPG tests the effectiveness of the PSO in each iteration by using the best global particle of the current iteration of the PSO as the starting design for the SPG. Save the SPG improvements and PSO continue their iterations. The hybrid PSO-SPG2 algorithm is shown below:

Hybrid PSO-SPG2

Step 1: Intialize the parameters of PSO and SPG2

Step 2: Set N as the maximum number of iterations and Let E=inf and k=1.

Step 3: If it reaches the kth iteration of PSO, then set P_k^x be the position of the best.

Step 4: if(N % k = = 0)

{

Start local minimization of P_k^x using SPG2 algorithm.

Initialize X_{SPG} be the solution found using SPG2 algorithm.

If (f (X_{SPG}) <E)

{

E= f (X_{SPG});

}

Else if(K== N)

{

Algorithm terminates;

}

Else

{

K=K+1;

Goto step 3;

}

Else

Goto step 3;

}

Procedure for Hybrid PSO-SPG2 Algorithm

The quality and speed of the prediction level can be improved using better Feature selection approaches.

2.3 PREDICTION OF COD

COD test is a common water quality test used to indirectly measure the total amount of organic compounds in a water sample using a strong oxidizing agent such as potassium dichromate. A high COD value indicates a high concentration of organic matter in the water sample. Proposed Method used ANFIS with Modified LM as training algorithm.

2.3.1 ADAPTIVE NEURO FUZZY INFERENCE SYSTEM (ANFIS):

ANFIS is a framework of adaptive technique to assist learning and adaptation. This kind of framework formulates the ANFIS modeling highly organized and not as much of dependent on specialist involvement. ANFIS is a method based on the input–output data of the system under consideration. Success in obtaining a reliable and robust ANFIS is proposed as a core neuro-fuzzy model that can incorporate human expertise as well as adapt itself through repeated learning. It is composed of an appropriate combination of neural and fuzzy systems. The hybrid combination of ANFIS allows dealing with both the verbal and the numeric power of intelligent systems.

ANFIS network depends heavily on the choice of process variables involved as well as the available data set and the domain used for training purposes [2].

Assume a fuzzy inference system with two inputs x and y, and one output , a common rule set with two fuzzy if –then rules is defined for the first-order Sugeno fuzzy model is as follows:

Rule 1: If x is A1 and y is B1, then

$$f1 = p1x + q1y + r1 \quad \text{----(13)}$$

Rule 2: If x is A2 and y is B2, then

$$f2 = p2x + q2y + r2 \quad \text{----(14)}$$

Here type-3 fuzzy inference system proposed by Takagi and Sugeno [4] is used. In this inference system the output of each rule is a linear combination of input variables added by a constant term. Average weight of each rule's output is the final output. Based on the system, the parameters of the membership functions can be initialized so that the convergence speed is increased.

ANFIS output is clearly a linear function of the adjustable defuzzifier parameters. Gradient descent method is applied at the adjustment of [p q r] T vector. The fuzzifier possesses six inputs and the rule base contains 25 rules and the defuzzifier has four outputs.

ANFIS architecture comprises of five layers. Each node in layer 1 is an adaptive node with a node function which may be a Gaussian membership function or any membership functions. Each node in layer 2 is a fixed node labeled Π which indicates the firing strength of each rule. All nodes in layer 3 is a fixed node which is labeled as N represents the normalized firing strength of each rule. In Layer 4 every node is an adaptive node with a node function. In layer 5, single node is a fixed node labeled Σ, represents the overall output (Z). It is defined as the summation of all incoming signals. Gaussian membership function is employed for the input variable. The hybrid learning algorithm is employed to determine the parameters of Sugeno-type fuzzy inference systems. The combination of the least-squares method and the back-propagation gradient descent method is utilized for a given training dataset to update FIS membership function parameters.

III. EXPERIMENTAL RESULTS

Raw data is gathered from the day by day operation of an anaerobic filter treating fruit juice wastewater for a period of six months [7]. COD effluent level in the wastewater is predicted automatically using proposed methods. At first, missing values in the data can be eliminated. Then the data is subjected to z-score normalization in which data can be arranged within a particular range. Feature selection algorithms are applied on the data to select the best feasible features. Prediction algorithm helps to deduce the COD concentration in wastewater.

The main purpose of this paper is to find the best feature selection algorithm for the prediction of COD level and to improve the prediction accuracy and to increase the speed in predicting COD effluent level. PSO-SPG2 approach is found to be the best for the selection of best features.

The model was developed using MATLAB software and the performance of ANFIS with each feature selection techniques was evaluated by calculating the prediction accuracy and execution time.

Three statistical indices are used in order to evaluate the performance of the proposed model: Mean Square Error (MSE) and Regression (R) values that are derived in statistical calculation of observation in model output predictions can be defined as:

$$MSE = \sum_{i=1}^{N} \frac{(x-y)^2}{N} \quad (15)$$

$$R = \frac{\sum(x-\bar{x})(y-\overline{y})}{\sum(x-\overline{x})^2} \quad (16)$$

where x is the actual values of Xt+1 with {i = 1, 2, …n} observations, $\bar{x}$ is the average of Xt+1 , n is the total observation number and y is the predicted Yt+1 value.

Table I shows the comparison of parameters for the prediction of COD using various feature selection algorithms. It is clear from the table that the PSO-SPG2 with ANFIS gives best results for the prediction of COD effluent level in fruit juice wastewater.

TABLE 1. COMPARING THE PARAMETERS FOR THE VARIOUS FEATURE SELECTION ALGORITHMS

Method	**Z-Score With ANFIS**		
	GA	**PSO**	**PSO-SPG2**
Execution Time	0.4219	0.212	0.1598
Accuracy	89	92	93
No. Of Features Selected	9	8	7
Regression	0.906	0.926	0. 954
RMSE	1.246	1.052	1.041
MAE	1.472	1.33	1.30
MSE	3.528	2.587	2.387

It is observed from the above table that PSO-SPG2 takes very less execution time to predict the COD concentration in fruit juice wastewater whereas GA algorithm takes more time than PSO and PSO-SPG2

Prediction accuracy is calculated for all the feature selection techniques in the prediction of COD level. In order to find the best feature selection technique for the prediction of COD in the ANFIS model, prediction accuracy is calculated using various feature selection technique and the results are shown in Fig 1. It is clear that the accuracy constantly increases from GA to PSO-SPG2.

Thus PSO-SPG2 gives better results to improve the quality and speed of predicting the COD level in fruit juice wastewater.

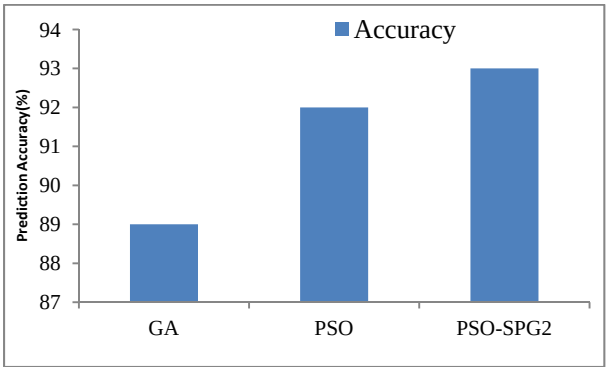

Fig. 1 Comparison of Prediction Accuracy Based On Various Feature Selection Algorithms

IV. CONCLUSION

Wastewater treatment incorporates the separating, removing and disposing the pollutants present in the wastewater. The strength of wastewater is usually measured using accurate analytical techniques. COD is an essential test for evaluating the quality of effluents and wastewaters prior to discharge. This paper investigates the effect of various feature techniques on the performance of a prediction level of COD effluent level using ANFIS. It is observed that the ANFIS with PSO-SPG2 predicted the effluent COD level with minimum execution time based on the selection of optimum features than GA and PSO.SPG2 tests the effectiveness of the PSO in each iteration by using the best global particle of the current iteration thus increasing the prediction accuracy. Thus, hybridizing PSO with SPG2 method provides better selection of features for the ANFIS to predict the effluent COD level with improved speed and accuracy than GA and PSO approaches.

REFERENCES

[1] Birgin, E. G. , Mart´ınez. J. M. and Raydan. M. (2000), Nonmonotone spectral projected gradient methods on convex sets, *SIAM Journal on Optimization*, Vol. 10, Pp. 1196–1211.

[2] Buragohain.M and Mahanta. C (2008), "*A novel approach for ANFIS modelling based on full factorial design*" , Applied Soft Computing , Pp. 609–625.

[3] Eberhart. R.C. and Shi. Y. (2001), "Particle swarm optimization: Developments, applications and resources", *Proceedings of the IEEE International Conference On Evolutionary Computation*. Seoul, Pp. 81-86.

[4] Jang. J.S.R. (1993), ANFIS: Adaptive-network-based fuzzy inference systems, *IEEE Trans. Syst. Man Cybern*, Vol. 23, Pp.665-685.

[5] Ladha. L and Deepa.T (2011), Feature Selection Methods and Algorithms, *International Journal on Computer Science and Engineering (IJCSE)*, Vol. 3, No. 5, May. ISSN: 0975-3397.

[6] Michalewicz.Z (1999), "Genetic Algorithms + Data Structures = Evolution Programs", *Springer-Verlag*, 3rd edition.

[7] Rajagopal Rajinikanth, Rengaraj Ganesh, Pradeep Kumar and Joseph V. Thanikald (2009), High rate anaerobic filter with floating supports for the treatment of effluents from small-scale agro-food industries, *Desalination and Water Treatment*, Vol. 4, Pp. 183–190.

[8] Raydan.M (1997), The Barzi lai and Borwein gradient method for the large scale unconstrained minimization problem, *SIAM Journal on Optimization* Vol.7, Pp. 26–33.

[9] Simona Caudo, Chiara Genovese and Siglinda Perathoner (2007), Copper-pillared clays (Cu-PILC) for agro-food wastewater purification with H2O, *Elsevier Inc*, doi:10.1016/j.micromeso.2007.05.011.

[10] Van Dijck,G, Van Hulle. M. M., and Wevers. M. (2005), Genetic Algorithm for Feature Subset Selection with Exploitation of Feature Correlations from Continuous Wavelet Transform: a real-case Application, *World Academy of Science, Engineering and Technology*, Vol. 1.

[11] Xiangyang Wang, Jie Yang, Xiaolong Teng, Weijun Xia, Feature Selection based on Rough Sets and Particle Sarm Optimization