**INTRODUCTION**

Data analysis is extensively used to design models for facilitating prediction that provides better understanding of a large dataset. Prediction is a form of data analysis that can be used by both human environmental scientists and computer-aided Environmental Decision Support Systems. Predictive tasks ensure safety operation in environmental process. The restoration of the normal operation condition can be sustained by accurate predictive tasks. The environmental managers are guided by accurate predictions thus insisting on the development of better prediction techniques.

1.1 **Applications of Prediction models**

Prediction models are implemented in various applications such as

- Home appliances
- Energy Consumption
- Faults
- Software reliability
- Network usage

The prediction models have a significant role greater in the field of environmental engineering, particularly in

- Water quality
- Wastewater treatment
- Air pollution
- Weather and climate conditions.
Prediction task identifies the current state of the environmental process. Afterwards, a right decision is made to lead the environmental process to a safe operational state, inorder to prevent from any possible environmental hazards. For example, in the field of air pollution, prediction of the air quality is required to find out the maximum level of pollution hazardous for the human population in a particular area so that bad consequences or outcomes can be avoided. The process of an accurate prediction of the current operational state of wastewater treatment plant is necessary for a safe normal operational state, as it avoids the high contamination degree of water outflow. Prediction tasks in environmental processes are key elements for their safe operation.

Among the several environmental issues, Wastewater Treatment System (WWTS) has become vital, as it plays a significant role in conserving and protecting the aquatic life and also in reducing the impact of wastewater on human health. WWTSs have to be enhanced on par with the growing population and the rapid industrialization.

1.2 Overview of Wastewater and Treatment Process

Wastewater may be defined as a combination of liquids or water-carried wastes discharged by residences, institutions, and commercial and industrial establishments, together with groundwater, surface water and storm water as may be present (Cheremisinff, 2002).

1.2.1 Types of Wastewater

There are different types of wastewater discharged from various sources. The categories include cotton textile mill wastewater, simulated textile wastewater, textile industry wastewater, food solid waste leachate, pulp and paper industry effluents, hospital wastewater, olive mill wastewater, slaughterhouse wastewater, synthetic wastewater, municipal wastewater,
domestic sewage, complex industrial wastewater and agro-food wastewater. Increasing urbanization, industrialization, and other anthropogenic activities result in degrading the quality of water day by day. These types of wastewater are hazardous and cannot be discharged directly on the ground or into the water bodies.

Agro and allied industries are considered to be the major contributing factors to worldwide industrial pollution. Small-scale agro-industries play a major role in the economy and at the same time they remain one of the most crucial industrial generators of water pollution and consumers of fossil fuel energy (El-Gohary et al., 2000). Hence, wastewater has to be treated to prevent environmental pollution.

1.2.2 Wastewater Treatment Process

The two basic types of biological wastewater treatment processes are aerobic and anaerobic process. Aerobic process can proceed with air or oxygen. When a culture of aerobic heterotrophic microorganisms is placed in an environment containing a source of organic material, the microorganisms will degrade and remove this material. A fraction of the organic material removed is utilized for the synthesis of new microorganisms, resulting in a biomass increase. The remaining material is oxidized to carbon dioxide, water and soluble inert material, providing energy for synthesis, metabolism and maintenance of the microorganisms’ vital functions (Zupancic and Ros, 2008). Anaerobic process proceeds without air or oxygen and anaerobic bacteria digest the waste via acid fermentation. Anaerobic treatment of waste is a complex biological process involving several groups of microorganisms. This technology has a positive net energy production and the biogas (mainly methane) produced can also replace fossil fuel sources and therefore, has a direct positive effect on greenhouse gas reduction (Batstone et al., 2002). Thus in recent years, treatment of agro-food wastewater has been shifted from aerobic to anaerobic process.
**Anaerobic Process**

Anaerobic wastewater treatment differs from conventional aerobic treatment methods. The non-existence of oxygen leads to controlled conversion of complex organic pollutions into Carbon dioxide and Methane (Mrowiec, 2005).

Anaerobic treatment of wastewater is extremely suitable for industries releasing exceedingly concentrated wastewater, with moderate Nitrogen concentrations (Dixon *et al*., 2005).

The key biological and chemical stages of anaerobic digestion are:

- Hydrolysis
- Acidogenesis
- Acetogenesis
- Methanogenesis

The anaerobic digestion initially starts with bacterial hydrolysis of the input materials in order to breakdown carbohydrates and these carbohydrates are available for other bacteria. Acetogen then renovates the sugars and amino acids into Ammonia, Carbon dioxide and Organic acid. These resulting organic acids are converted by acetogenic bacteria into acetic acid, along with additional Hydrogen, Ammonia and Carbon dioxide. Finally, these products are converted to Methane and Carbon dioxide by methanogen. The main advantages of anaerobic treatment process are:

- Removal of higher organic loading
- Removal of high pathogen
- Bio-gas production
- Low sludge production and
- Low energy consumption
High organic loading rates and low sludge production are among the many advantages of high-rate anaerobic process. In conventional anaerobic processes, slow rate and process instability are the major drawbacks. Space requirements and costs involved also increases as the process stability increases. In addition, the biogas technology for large scale applications has been limited due to these problems. High-rate anaerobic reactors are becoming increasingly popular for the treatment of various types of wastewaters because of their low initial and operational costs, smaller space requirements, high organic removal efficiency and low sludge production, combined with a net energy benefit through the production of biogas (Jawed and Tare, 2000). In order to achieve high system loading rates, minimum HRTs should be used, while at the same time maintaining positive net solids (biomass retention).

The various types of high-rate anaerobic reactors used to treat the wastewater are:

- Anaerobic Contact (AC)
- Up-flow Anaerobic Filter (UAF)
- Fluidized-Bed Reactor (FBR)
- Up-flow Anaerobic Sludge Blanket reactor (UASB)

These types of anaerobic reactors or digesters allow higher wastewater volume as well as higher organic loading rates and are suitable for wastewater treatment of high organic material content. The wastewater is treated using the laboratory scale anaerobic reactors or digesters.

The common parameters used to characterize wastewater are Biological Oxygen Demand (BOD), pH, Total Suspended Solids (TSS), Chemical Oxygen Demand (COD), Total Phosphorus and Total Nitrogen (Jules, 2008). A high
COD value indicates a high concentration of organic matter in the water sample and it is the most important in controlling the total content of pollutants. COD test is performed regularly in laboratories and industries. On treating the agro-food wastewater, the levels of effluent COD must be predicted before it is discharged.

1.3 Research Challenges

The main problems observed in wastewater treatment process are as follows:

i) Construction of Wastewater Treatment Plant (WWTP) and handling of dry sludge disposal are very costly.

ii) Mechanical systems which are more suitable to treat the wastewater are very expensive to build and operate.

iii) The complexity and cost of wastewater treatment technologies increase with the quality of the effluent produced.

iv) Imperfect and less significant technologies take longer duration to detect the amount of impurities present in wastewater.

v) In conventional method, the Chemical Oxygen Demand (COD) test is incorporated and an element of complexity exists in controlling the pollutants.

vi) In addition to the complexity in the control of pollutants, the accuracy in predicting the amount of COD level is unpredictable in conventional methods.

Hence, formulation of an effective approach in wastewater treatment technology is essential, so that rigid environmental regulations are fulfilled. For these reasons, many soft computing techniques have been undertaken to solve these predictive tasks in the past.
1.4 Soft Computing Techniques

Soft Computing is a combination of methods namely, Neural Network (NN), Fuzzy Logic (FL), Evolutionary and Genetic Computing and, which are tolerant to imprecision, uncertainty and partial truth. These methodologies can be used altogether to create powerful hybrid intelligent systems. Better results can be obtained by employing the ingredients of soft computing in combination rather than in a stand-alone mode. ANN, Genetic Algorithm (GA) and FL are Artificial Intelligence (AI) based modelling techniques.

The Artificial Intelligence based prediction models have greater potential to solve complex environmental applications that incorporate large amount of independent parameters and non-linear relationships. Several AI based modeling techniques, such as ANN, Fuzzy Logic, and Adaptive Neuro-Fuzzy Inference Systems (ANFIS) have recently been conducted in modeling of various real-life processes in environmental engineering field because of their predictive capabilities and non-linear characteristics (Turkdogan and Yetilmezsoy, 2010).

1.4.1 Artificial Intelligence Techniques

Reliable prediction of Wastewater Treatment System (WWTS) is becoming more important, to fulfill the need for safety environmental measures. The technique called AI can automate the prediction of WWTS of the influent characteristics. In this research, AI based prediction system is developed and the concepts related to the techniques are provided in the following sections.

Artificial Neural Network (ANN)

An ANN is a computational system consisting of a large number of neuron – like processing elements and having weighted connections between these elements. The ANN concepts are originally developed to simulate the human brain which is a close network of connected biological neurons by nerves.
The connections between the elements represent the distributed knowledge acquired by the network through a learning process. An ANN combines the human association with the analytical power of a computer. The structure of an ANN is shown in Figure 1.1, which defines the overall architecture of the network, including one input layer, one output layer and usually one or two hidden layers. The architecture of a neural network depends on the number of layers, the number of neurons in each layer, each layer’s transfer function and the connection among the layers.

![Figure 1.1 Typical ANN structure](image)

The connection pattern determines the categories of ANN architecture that is either feed-forward or feedback neural networks. In feed-forward neural network, neurons are organised into layers that have unidirectional connections between them. Feed-forward networks are static, that is, they produce only one set of output values from a given input and their response to an input is independent of the previous network state. Recurrent, or feedback, networks, on
the other hand, are dynamic systems having feedback paths that lead the network to enter a new state by modifying the inputs to each neuron. An appropriate learning algorithm is chosen based on the architecture of ANN.

The learning methods used for ANN are classified into two categories namely, supervised and unsupervised learning. In supervised learning, both the inputs and the outputs are provided. The network then processes the inputs and compares its resulting outputs against the desired outputs (http://www.ureason.com/mathematical-statistical/neural-networks). Errors are then calculated, causing the system to adjust the weights which control the network. In unsupervised learning, for the set of given patterns, the network has to learn by itself. In this research work, supervised learning is considered in which a neural network generates its own rules by learning from shown examples.

An ANN consists of many artificial neurons joined together. It combines the values of these input paths, usually by summation and a transfer function then modifies the combined input. The transfer function can be a simple mathematical function like sigmoid, hyperbolic tangent, sine or linear function. The output of the transfer function is passed to the input paths of other neurons.

Based on the calculated differences between the derived or measured output of a certain process and the results obtained by ANN, the so-called residuals or errors, parameters or weights are automatically adjusted in order to minimize the final and total error. ANN searches automatically for the best linear or non-linear relationships between cause (input) and effect (output). Neural computing differs from traditional computing in several important ways. Unlike traditional expert systems where knowledge is made explicit in the form of rules, neural networks generate their own rules by learning from examples shown to them. Learning is achieved through a learning rule which adapts or changes the
connection weights of the network in response to the example inputs and (optionally) the desired outputs of these inputs. There are two main phases in the operation of an ANN, learning and recall. Learning is the process of adapting or modifying the connection weights in response to examples. Recall refers to how the ANN performs when new but representative data are presented to the ANN (prediction). Before starting with the learning phase it is very important to verify and screen the data. Figure 1.2 shows a stepwise network operation.

![Figure 1.2 A Stepwise Network Operation](image)

**Fuzzy Logic**

Yet another popular AI-based modeling technique, the fuzzy logic methodology has also been analysed by many researchers as a promising method for modeling of various types of environmental problems in recent years (Garcia...
et al, 2007). These models exhibit the non-linear functions in an easy and understandable way by explaining the reasoning linguistically rather than with numerical quantities. The reasoning of non-linear functions linguistically involves fuzziness and the fuzzy logic systems with fuzzy theory make effective decision on the basis of imprecise linguistic information. The desired performance of the fuzzy logic system is achieved depending upon the tuning of the parameters in the membership function and building the fuzzy rules.

Fuzzy Inference System (FIS) is a process of mapping from given inputs to outputs by using the theory of fuzzy sets (Zadeh, 1965). FIS has been successfully applied in various fields of applications (Wong and Gedeon, 2000; Wong et al., 2003). FIS reflects the uncertainty as variables to be “partial true” and/or “partial false”. FIS derives an output by using an inference engine which is based on a form of IF-THEN rules. Basically, the two usual approaches to defuzzify the output fuzzy sets are Mamdani (Mamdani and Assilian, 1975) and Sugeno (Sugeno, 1985) approaches. The Mamdani approach defuzzifies output fuzzy sets by finding the centroid of a two-dimensional shape by integrating across a continuously variation function. In the Sugeno approach, output fuzzy sets are in the form of singleton, a fuzzy set with unity membership grade at a singleton point and zero elsewhere else on the universe of discourse. The output centroid is calculated by the weighted average method. In this study, the Sugeno approach is used because it is computationally efficient, works well with optimization and adaptive techniques and guaranteed continuity of the output surface.

Adaptive Neuro – Fuzzy Inference System (ANFIS)

ANFIS combines the two approaches of Artificial Neural Network (ANN) and Fuzzy Logic System. If both intelligent approaches are integrated, it will help to achieve good reasoning in quality and quantity. These tools apply fuzzy
inference techniques to data modeling. The acronym ANFIS derives its name from Adaptive Neuro-Fuzzy Inference System and it is a network structure consisting of a number of nodes connected through directional links (Jang, 1993). Each node has a node function with adjustable or fixed parameters. Learning or training phase of network is a process to decide parameter value to adequately fit the training data. The basic learning rule is the common back-propagation method which tries to minimize sum of squared differences between network’s outputs and desired outputs (Kaya et al., 2002).

The FIS is classified into two types; Mamdani’s system, Sugeno’s system depending on the types of inference operations using “if-then rules”. Mamdani’s system is the most commonly used; while Sugeno’s system is more compact and computationally efficient with concrete output. Mathematically intractable defuzzification operation with the use of adaptive techniques are the basis for sample-data based fuzzy modeling with minimum time consumption in Sugeno (Takagi and Sugeno, 1985).

ANFIS constructs a FIS whose membership function parameters are tuned (adjusted) using Back Propagation (BP) algorithm. The function parameters are also tuned by the combination of BP algorithm and least square type of method. The learning algorithms used to tune the function parameters allow the fuzzy systems to learn from the data to be modeled. In general, this type of modeling works well if the training data presented to ANFIS is fully representative of the features of the data. In some cases, data is collected using noisy measurements, and the training data cannot represent all the features of the data presented to the model. At this point of time, the model validation becomes predominant. ANFIS models provided better prediction capabilities because they generally offer the ability to model more complex non-linearity and interactions than linear and exponential regression models can offer. Good prediction capabilities are
required on machining for today’s industries. ANFIS can be adopted in predicting machining response from given input conditions with much non-linearity relation among the various process parameters.

The neuro-adaptive learning techniques have procedures to learn information about a dataset by involving fuzziness. The fuzziness computes the membership function parameters to track the given input/output data. The parameters associated with the membership functions will change through the learning process. The gradient vector computes these parameters and the computation provides a measure of how well the fuzzy inference system is modeling the input/output data for a given set of parameters. Several optimization routines could be applied in order to adjust the parameters so as to reduce the error measure between the actual and the observed outputs. ANFIS uses either BP or a combination of least squares estimation and back propagation for membership function parameter estimation.

In the last two decades, soft computing techniques have been extensively studied and applied for water engineering problems. This research focuses on the development of an enhanced prediction system for the Anaerobic Digestion (AD) of WWTS to predict the effluent COD level, thus providing with the necessary evidences for the environmental managers to take timely remedial actions. Despite the existence of various prediction systems, a system with high prediction accuracy and minimum error rate is yet to be developed. The proposed system focuses on improving the prediction system in its performance. Pre-processing, efficient feature selection method and prediction using data mining and soft computing techniques are the series of steps being adapted to improvise the prediction system.
1.5 Problem Justification

The characteristics of influent in any WWTS vary depending on the type of wastewater. In addition, the type of influent is also time dependent and it is difficult to have a homogeneous influent to a system. This may result in an operational risk on the system. Serious environmental and public health problems may result from improper operation and control of a wastewater treatment system, as discharging contaminated effluents to a receiving water body can cause or spread various diseases to human beings. Therefore, to overcome these issues, environmental regulations have restrictions emphasizing on the quality of effluent that must be met by any WWTS. These strict discharge standards and time-dependent non-uniform influent characteristics, have led to the development of proper management of treatment systems. This has made researchers and scientists to have major attention in developing wastewater treatment systems. WWTS involves several complex physical, biological and chemical processes. Necessary treatment is required before discharging this wastewater. However, the quantum of pollutants released by small scale industries may be equivalent to that of a large-scale industry.

Inspite of various techniques available in literature to treat wastewater and its further purification by means of prediction of effluents, yet it is essential to predict the level of effluent before discharging. But these conventional filtering systems have lot of limitations. Some of the problems of the traditional wastewater systems are unpredictable effluent levels, highly non-linear influent parameters and system changes. Such problems in conventional methods enable new prediction systems to be developed.

Research problem is formulated with the aim to design a prediction system for estimating the effluent COD level intelligently and intuitively for agro-food wastewater using soft computing techniques based on anaerobic digestion influent parameters.
1.6 Objectives

To build the prediction system for the above stated problem, the following objectives have been formulated

i. To produce quality data by applying techniques for replacement of missing values found in the raw data.

ii. To adapt an enhanced normalization technique to scale the input features for effective learning.

iii. To improve the prediction rate and minimize the execution time using better feature selection approach.

iv. To increase the prediction accuracy of the effluent COD and to reduce the errors using effective learning method in ANFIS.

1.7 Research Contributions

The following are the research contributions:

i. A modified normalization technique, Dynamic Score Normalization with Mahalanobis distance (DSN-M) has been proposed.

ii. Particle Swarm Optimization with Spectral projected Gradient (PSO-SPG2) is proposed for feature selection.

iii. An Enhanced prediction system is proposed using

- Adaptive Neuro Fuzzy Inference System (ANFIS) with Modified Levenberg Method (MLM).
1.8 Layout of the Thesis

This chapter introduced applications of prediction models, types of wastewater and outlined the specific objectives of the research. The rest of the thesis is organised as below.

A critical look at the various available literatures related to the present research work is given in Chapter 2, Review of Literature.

Chapter 3, Methodology, presents the research methodology and discusses the different phases of the proposed methodology. The various algorithms and techniques used are explained in this chapter.

Chapter 4, Results and Discussion, tabulates and discusses the various results obtained while testing. The results obtained are analysed and discussed in this chapter. The findings of the study are summarized along with future research directions in Chapter 5, Summary and Conclusion.

The work of several researchers are quoted and used as evidence to support the concepts explained in this thesis. All such evidences used are listed in the Bibliography.

1.9 Chapter Summary

This chapter discussed the overview of the prediction models and its applications. The various types of wastewater and wastewater treatment process using anaerobic process are elaborated. The objectives of the thesis and the research contributions are given in this chapter. A review of the previous research works was studied and the scrutinized works are summarized in the next chapter, Review of Literature.