Machine Learning Based Prediction of Pyrolytic Conversion for Red Sea Seaweed

Muhammad Saleem[1], Imtiaz Ali[2]

Abstract—The complex and non-linear relationship between kinetic parameters and main biomass components is one of the major problem for model development. The application of non-linear modelling would be interesting to predict the behavior of such a complicated model. The machine learning approaches are widely used in many applications for modelling the non-linear relationship between input and output data. In this study, two machine learning tools, namely, Artificial Neural Network (ANN) and Support Vector Machine (SVM) are investigated in order to develop a prediction model for the pyrolytic conversion of Red Sea seaweed (*Sargassum* sp.). Pyrolysis of *Sargassum* sp. was carried out in a thermogravimetric analyzer (TGA) where the samples were non-isothermally heated at 10, 15, 20 and 25°C·min^{-1} from 40 to 900°C in an inert atmosphere. The obtained mass loss-time and temperature series data sets are used for the training, validation and testing the prediction model with ANN and SVM. The results show that ANN provide better accuracy than SVM. More importantly, higher R^2 and lower error rates renders machine learning based tools as reliable means for forecasting pyrolytic conversion.

Keywords—Machine learning, Prediction model, Pyrolytic conversion, *Sargassum* sp.

I. INTRODUCTION

Biomass is a complex material constituted from many organic and inorganic constituents. It is a rich source of carbon which is the most valuable element of widely produced chemicals and energy. This inherent potential of biomass can be exploited by converting it to valuable bio-chemicals and bioenergy through environmentally benign processes. Non-food biomass are abundantly available in nature and are also cultivated on land and in water. Direct firing of the biomass in the furnace to produced energy leads to large environmental impact. Biomass conversion is typically carried out either through biochemical or thermochemical processes. Thermochemical processes have some benefits over biochemical processes as the conversion rate are higher also lignin which is the main constituent of lignocellulosic biomass is difficult to digest through biochemical processes. Pyrolysis is simple, flexible and an efficient process compared to other thermochemical processes. The pyrolytic process parameters can be varied to obtain different yields of gaseous, liquid or solid products. Depending on the required output, the operating process parameters are optimized. Since it is an energy intensive process, to develop efficient pyrolytic systems, an understanding of the process kinetics is utmost important. Therefore, kinetic models can be developed and subsequently tested on the pyrolysis. One of the challenge in the model development is the interplay of biomass constituents during pyrolysis. So comprehensive kinetic models are needed which are in reality both time consuming and computationally expensive.

Sargassum sp. is composed of cellulose, hemicelluloses, lignin, lipids and protein [1], [2]. The most common equipment to conduct pyrolysis in the lab is a thermogravimetric analyzer (TGA), where mass loss profile is obtained as the samples are heated in an inert atmosphere under controlled conditions. In this work, Red sea seaweed (*Sargassum* sp.) was pyrolyzed in TGA. The obtained mass loss data with respect to time and temperature are used to predict the conversion rate with machine learning algorithms.

The machine learning approaches are extensively used in many applications for modelling the non-linear relationship between input and output data. Machine learning is described as the study of those computer algorithms which are capable of learning and automatically improving themselves overtime [3]. It can also be defined as the capability of computer algorithm to produce new information from existing or non-existing data-set in order to optimize performance. [4]. Machine learning techniques have proved to work well with massive amounts of data and provide outstanding results within seconds in different fields like computer vision, medical diagnosis, bio-surveillance, pattern recognition, and credit card fraud detection in financial institutions [5].

The supervised and unsupervised are two main groups of machine learning categorized on the basis of learning methods. Supervised learning contains algorithms that learn from history data instances and build general hypothesis to make predictions about future instances. In supervised learning, generally, response variable is known and steer the learning process. Supervised machine learning algorithms like linear and logistic regression, Artificial Neural Network (ANN), Decision Trees, Support Vector Machine (SVM), K-Nearest Neighbour (KNN) and Random Forests have been applied in different application domains. ANN and SVM have shown high reliability for the training of non-linear regression models

Muhammad Saleem[1], was with the Faculty of Engineering, University of Duisburg-Essen, Germany. He is now with the department of Industrial Engineering, King Abdulaziz University, Rabigh Campus, KSA (phone: +49 179 780 5633; e-mail: msaleim1@kau.edu.sa).

Imtiaz Ali[2] is with the department of Chemical and Materials Engineering, Rabigh Campus, KSA (e-mail: imtiaz_che@hotmail.com).

and are employed in this work to predict the pyrolytic conversion rate.

Artificial neural network (ANN) is machine learning model for learning multivariate and complex non-linear relationships inspired by biological neurons in the human brain [6], [7]. ANN algorithm is widely used in many applications for the tasks like prediction, classification, and pattern recognition. ANN have been applied to a number of chemical engineering problems [8]–[12] and even to kinetic model reduction [13], [14]. ANN have also been used in some other aspects of biomass pyrolysis modelling [15]–[18].

Formally, ANN consists of interconnected neurons. Multilayer perceptron (MLP) is a type of artificial neural network model in which perceptrons (simple neurons) are structured in layers. These layers are classified into three types; input, hidden and output layers. The neurons are interconnected with weighted connection lines between layers. The input layer is scaled and introduced into the hidden layer through weight and neurons in the hidden layer sum up the weighted inputs. Each layer receives its input from the previous layer and the output of each neural is forwarded to the next layer. The weights are adjusted iteratively during a training process using backpropagation algorithm [19] resulting in networks that can be highly accurate and extremely computationally efficient.

Support Vector Machine (SVM) is fairly new advanced machine learning method for supervised classification problems introduced by Vladimir Vapnik and colleagues [20]. SVM is a popular supervised learning algorithm which can handle high-dimensional data for desired level of accuracy. It has been widely used in different disciplines like fault diagnosis [21], image classification [22], geographical analysis [23], hand-written character recognition [24], and bioinformatics [25].

SVM is actually a kernel-based algorithm that convert data into high dimensional space and build a hyperplane. The main idea is to identify separating hyperplanes during the learning procedure that maximize the distance to the nearest data point of any of the input classes and minimize the classification errors [26]. SVM is originally developed to train binary classifiers but can be extended to multiple classes by dividing the multiclass problem into several binary classification problems [27].

The SVM were primarily developed for classification of linearly separable problems, however, it can be generalized to non-linear ones by using kernel methods [28]. SVM algorithm can perform well for both linear and nonlinear separable problems in classification and regression tasks.

II. MACHINE LEARNING BASED PREDICTION

A. Biomass Material

Sargassum sp. of Red sea origin was collected shores of Rabigh, Kingdom of Saudi Arabia (22.793151N, 38.951435E). The biomass material was washed with distilled water and air-dried. The moisture content was further reduced by subsequent oven-drying at 50 °C for approximately 4 h. The dried samples were pulverized to a particle size of less than 0.42 mm in a knife mill and the fine powder was stored in an air-tight jar for further characterizations.

B. Characterization

The proximate analysis of *Sargassum* sp. was carried out using ASTM standards and the ultimate analysis was carried out in a CHN/O elemental analyzer (2400 series II PerkinElmer). The details of which are available elsewhere [29]. Thermal characterization of the powder was carried out in TGA 4000 (Perkin Elmer Inc., USA). Pyrolytic conditions were maintained by sparing pure nitrogen gas at 20 mL·min^{-1}. Almost about 35 ± 5 mg of the sample was used for each experimental run. The samples were heated non-isothermal from 40 to 900°C at 10, 15, 20, and 25°C·min^{-1}, separately. Each experimental run was performed two times in order to minimize the error in analysis. The experiment was performed third time only when the conversion difference in the final mass loss (%) was more than 5%.

C. Methodology

Training of any supervised learning algorithm involves exposing it to a sufficient amount of input-output learning data. In this work, the set of experimental data used to develop the prediction model through the machine learning algorithms were obtained from the pyrolysis of Red sea seaweed (*Sargassum* sp.) in TGA. The input data was the heating rate and temperature whilst the output data were the conversion rate.

In supervised learning, available learning data is usually divided into training, validation and testing data sets. In ANN, a training set is used to adjust the weights between neurons in the network. A validation set is used to decide when to stop the training of network. This has been done by evaluating performance of the trained network on data that has not been exposed for the purpose of model selection. The test data assess the performance of the final selected model on an unknown data. To set up these splits, we first performed some pre-processing of our available data by randomly shuffling all learning pairs. The data was randomly separated into training (70%), validation (10%) and test (10%) subsets. In addition to this split, 10% of the data kept for the prediction purpose. This additional prediction data set is completely new for the model and will further confirms the accuracy with the machine learning tools. The number of training, validation, testing, and prediction data instances for different heating rates are given in Table 1.

TABLE I
DATA INSTANCES FOR MACHINE LEARNING MODELS

Heating Rate β [°C·min^{-1}]	Data Instances for Training, Validation & Test dα/dt [%·min^{-1}]	Data Instances for Prediction dα/dt [%·min^{-1}]
10	4643	516
15	3095	344
20	2322	258
25	1857	206

Feedforward ANN models containing multilayer percepton (MLP) has been developed using Neural Network Toolbox

available in MATLAB software. The network is trained using the Levenberg-Marquardt backpropagation algorithm. The structure of ANN has to be determined and then optimization is performed by minimizing the error between the target and calculated parameters. This trained ANN model is then saved and can be used to estimate the output function whenever required. The performance of the models in all data sets can be statistical measured by the linear regression coefficient R^2 between the experimental and predicted values.

Support Vector Machine (SVM) is another important supervised learning method considered to be very competitive and seems to be the newer alternative to ANN. We have used the same learning data, as for ANN, to train the SVM using SVM Toolbox in MATLAB software. Since our data represent the non-linear behavior so it is possible to build the prediction model using kernel methods in SVM. Different kernel functions, available in the toolbox, such as quadratic, cubic, fine Gaussian, medium Gaussian, and coarse Gaussian have been employed to the experimental data in order to find a best fit model.

III. RESULTS

Both machine learning classification algorithms, ANN and SVM, have been trained for the given data sets with different heating rates (β = 10, 15, 20, and 25°C·min^{-1}). The validation data have determined the optimum prediction model based on the performance. After the training procedure, each machine learning model has been evaluated based on its performance to estimate the output parameter. For this purpose, each model got input from the test data set which was not previously exposed to it. The models predictions are then compared with the expected output of the test data sets. The parameters of the algorithm are optimized by minimizing the difference between the calculated and target values through learning process. The optimum models, with lower error rate and higher coefficient of determination R^2, are then saved for future predictions.

Fig.1 (a-c) are the representative plots for training, validation and test data sets with ANN showing output measures with high R^2 for the heating rate β =10°C·min^{-1} while best validation performance plot for ANN is shown in Fig.1 (d).

The prediction data set (10% of the original data) is now used to predict with the saved models. This data is completely unknown for the model. Fig.2 shows the graph (temperature versus conversion rate) of the original and predicted data for this prediction data set along with the error plot using saved model of ANN. This makes it clear that the saved ANN model is able to predict all the output measures with excellent accuracy. The R^2 (0.999) measure for predicted and target output of the prediction data is shown in Fig.3.

Same prediction data set is also given to the saved SVM model for the prediction. The best prediction results are obtained from fine Gaussian kernel of SVM. The graphs similar to ANN model are shown in Fig.3 & 4. These plots show that SVM model has slightly less R^2 measure (0.982) compared to ANN but still competitive.
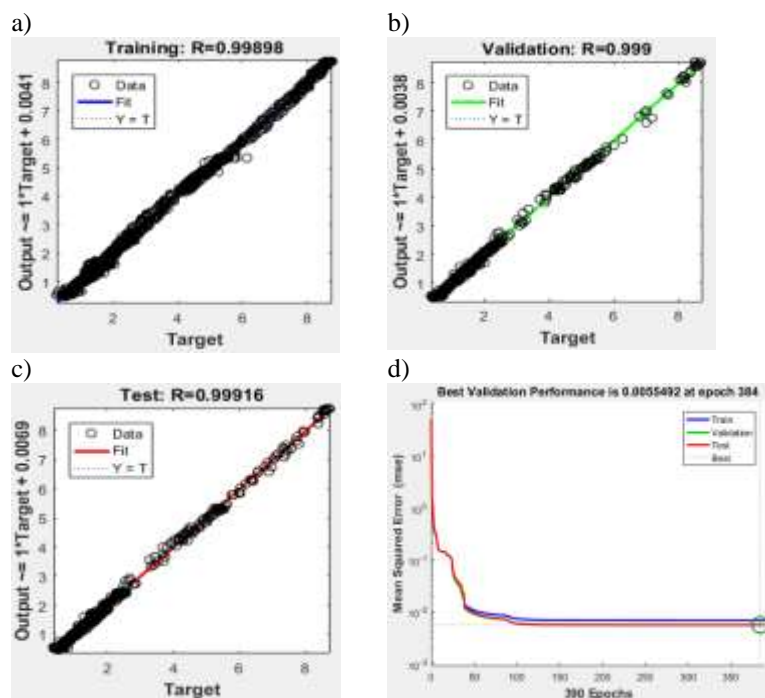

Fig. 1 ANN model fit plot for β =10 °C·min^{-1} a) Training, b) Validation, c) Test, and d) Best validation Performance

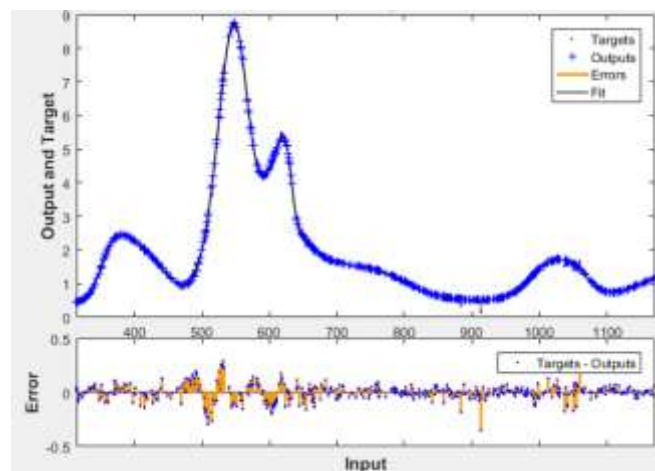Fig. 2 ANN Error graph for β =10 °C·min^{-1}

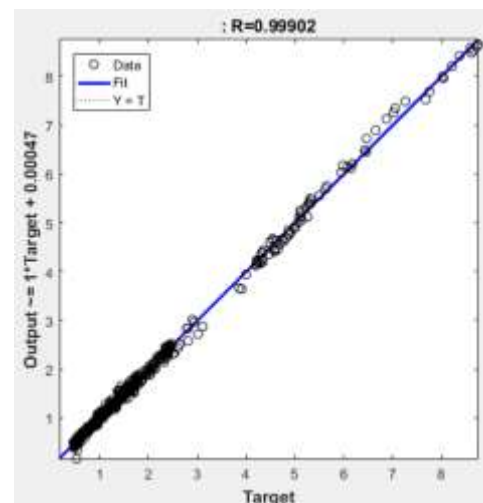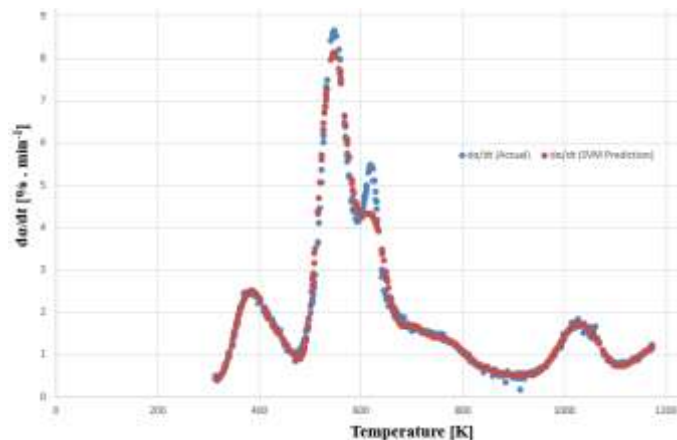Fig. 3 Model fit of ANN Prediction for β =10 °C·min^{-1}

Fig. 4 Actual data vs. SVM prediction for β =10 °C·min^{-1}

Fig. 5 Model fit of SVM Prediction for β =10 °C·min^{-1}

The R^2 measures for the prediction data sets, for all heating rates, with both machine learning models (ANN & SVM) are summarized in Table II. It is evident from Table II that both ANN and SVM algorithms performed really well as indicated by higher R^2 values

TABLE II
R^2 MEASURE FOR PREDICTION DATA SET

Heating Rate β [°C·min^{-1}]	R^2 (ANN)	R^2 (SVM)
10	0.9990	0.9823
15	0.9982	0.9752
20	0.9975	0.9743
25	0.9971	0.9733

IV. CONCLUSION

In this work, two different machine learning classification algorithms, Artificial Neural Network (ANN) and Support Vector Machine (SVM) have been tested and compared to develop a prediction model for the pyrolytic conversion of Red Sea seaweed (*Sargassum* sp.). The obtained data sets from the pyrolysis of *Sargassum* sp. were used to train, validate and test the prediction model with ANN and SVM. The input data to these models were heating rate and temperature while conversion rate was the output which needs to be predicted

from the developed models. The results showed that both algorithms performed better, however, ANN model provide better accuracy than SVM. The best prediction has achieved an average R^2 value of 0.998 using ANN model whereas SVM yielded R^2 up to an average of 0.976 proving competitive alternatives to ANN.

REFERENCES

[1] M. Radhakumari, D.J. Prakash, B. Satyavathi, "Pyrolysis characteristics and kinetics of algal biomass using tga analysis based on ICTAC recommendations," *Biomass Conv.* Bioref. 6, 2015, pp. 189–195. doi:10.1007/s13399-015-0173-7.

[2] H. Rabemanolontsoa, S. Saka, "Comparative study on chemical composition of various biomass species," *RSC* Adv. 3, 2013 pp. 3946–3956. doi:10.1039/C3RA22958K.

[3] T. Mitchell, *Machine Learning*, MIT Press, 1997.

[4] Alpaydin Ethem, *Introduction to Machine Learning*, Cambridge, Massachusetts, London, England: MIT Press. 2004.

[5] T. Mitchell, *The Discipline of Machine Learning*, Carnegie Mellon University, Pittsburgh, PA, USA, 2006.

[6] Bishop, C. M., *Pattern recognition and machine learning*, New York, Springer, 2006.

[7] Haykin, S. S., *Neural Networks and Learning Machines*, Prentice Hall, 2009.

[8] Asadi-Eydivand, M., Solati-Hashjin, M., Farzadi, A., & Abu Osman, N. A., "Artificial neural network approach to estimate the composition of chemically synthesized biphasic calcium phosphate powders," *Ceramics International*, 40, 2014, pp. 12439-12448.

[9] Beck, D. A. C., Carothers, J. M., Subramanian, V. R., & Pfaendtner, J., "Data Science: Accelerating Innovation and Discovery in Chemical Engineering," *Aiche Journal*, 62, 2016, pp. 1402-1416.

[10] Kayala, M. A., & Baldi, P., "Reaction Predictor: Prediction of Complex Chemical Reactions at the Mechanistic Level Using Machine Learning," *Journal of Chemical Information and Modeling*, 52, 2012, pp. 2526-2540.

[11] Molga, E. J., "Neural network approach to support modelling of chemical reactors: problems, resolutions, criteria of application," *Chemical Engineering and Processing*, 42, 2003, pp. 675-695.

[12] Rocabruno-Valdes, C. I., Ramirez-Verduzco, L. F., & Hernandez, J. A., "Artificial neural network models to predict density, dynamic viscosity, and cetane number of biodiesel," *Fuel*, 147, 2015, pp. 9-17.

[13] Conesa, J. A., Caballero, J. A., & Reyes-Labarta, J. A., "Artificial neural network for modelling thermal decompositions," *Journal of Analytical and Applied Pyrolysis*, 71, 2004, pp. 343-352.

[14] Molga, E. J., van Woezik, B. A. A., & Westerterp, K. R., "Neural networks for modelling of chemical reaction systems with complex kinetics: oxidation of 2- octanol with nitric acid," *Chemical Engineering and Processing*, 39, 2000, pp. 323-334.

[15] Mikulandrić, R., Lončar, D., Böhning, D., Böhme, R., & Beckmann, M., "Artificial neural network modelling approach for a biomass gasification process in fixed bed gasifiers," *Energy Conversion and Management*, 87, 2014, pp. 1210-1223.

[16] Puig-Arnavat, M., Hernández, J. A., Bruno, J. C., & Coronas, A., "Artificial neural network models for biomass gasification in fluidized bed gasifiers," *Biomass and Bioenergy*, 49, 2013, pp. 279-289.

[17] Sun, Y., Liu, L., Wang, Q., Yang, X., & Tu, X., "Pyrolysis products from industrial waste biomass based on a neural network model," *Journal of Analytical and Applied Pyrolysis*, 120, 2016, pp. 94-102.

[18] Sunphorka, S., Chalermsinsuwan, B., & Piumsomboon, P., "Application of artificial neural network for kinetic parameters prediction of biomass oxidation from biomass properties," Journal of the Energy Institute, 90, 2017, pp. 51-61.

[19] D. Michie, D. J. Spiegelhalter, C. C. Taylor, and J. Campbell, Eds., *Machine learning, neural and statistical classification*, Upper Saddle River, NJ, USA: Ellis Horwood, 1994.

[20] B.E. Boser, I.M. Guyon, V.N. Vapnik, "A training algorithm for optimal margin classifiers," *In 5th Annual ACM Workshop on COLT*, Pittsburgh, PA. ACM Press, 1992, pp. 144–152.

[21] N. Pooyan, M. Shahbazian, K. salahshoor, M. Hadian, "Simultaneous fault diagnosis using multi class support vector machine in a Dew point process," *J. Natural Gas Sci. Eng*. 23, 2015, pp. 373–379.

[22] S. Li, H. Wu, D. Wan, J. Zhu, "An effective feature selection method for hyperspectral image classification based on genetic algorithm and support vector machine," *Knowledge-Based Systems*, 24, 2011, pp. 40–48.

[23] B. Pradhan, "A comparative study on the predictive ability of the decision tree, support vector machine and neuro-fuzzy models in landslide susceptibility mapping using GIS," *Comput. Geosci*. 51, 2013, pp. 350–365.

[24] J. John, K.V. Pramod, K. Balakrishnan, "Unconstrained handwritten Malayalam character recognition using wavelet transform and support vector machine classifier," *Procedia Eng*. 30, 2012, pp. 598–605.

[25] G. Orr`u, W. Pettersson-Yeo, A.F. Marquand, G. Sartori, A. Mechelli, "Using support vector machine to identify imaging biomarkers of neurological and psychiatric disease: A critical review," *Neurosci. Biobehav*, Rev. 36, 2012, pp. 1140–1152.

[26] C. Campbell, *An Introduction to Kernel Methods*, 2000.

[27] J.A. Nasiri, N.M. Charkari, S. Jalili, "Least squares twin multi-class classification support vector machine," *Pattern Recognition*, 48, 2015, pp. 984–992.

[28] N. Cristianini, J. Shawe-Taylor, *An introduction to support vector machines: And other kernel-based learning methods*, Cambridge University Press, 2000.

[29] I. Ali, A. Bahadar, "Red Sea seaweed (Sargassum spp.) pyrolysis and its devolatilization kinetics," *Algal Research*. 21, 2017, pp. 89–97. doi:10.1016/j.algal.2016.11.011.