Indian Journal of Chemical Technology
Vol. 31, January 2024, pp. 11-19
DOI: 10.56042/ijct.v31i1.6357

Application of machine learning in optimizing thermochemical conversion processes with pre-treatment to get higher bio-oil yield from biomass waste

Murugan Kamarajan[1], Kandasamy Sundaresan Srinivasan[1]* & Cingaram Ravichandran[2]

[1]Department of Electronics and Communication Engineering, Mohamed Sathak AJ College of Engineering, Chennai - 603103, Tamil Nadu, India

[2]Department of Chemistry, Easwari Engineering College, Chennai - 600089, Tamil Nadu, India

*E-mail: srinivasanks2022@gmail.com

Received 23 October 2023; accepted 13 December 2023

Improving the bio-oil yield is a challenging part in the thermochemical conversion processes of biomass. Implementing suitable pre-treatment technology to improve the biomass characteristics is an effective technique to increase the yield. In this study, a multi-variate random forest algorithm has been used to optimize the pre-treatment method in order to improve the biomass characteristics. The data collected from many previous studies are analysed to identify the importance of biomass characteristics in bio-oil yield. The correlation between biomass characteristics and bio-oil yield, is analysed using Pearson method and the important influencing parameters %C and %H have a very good positive correlation with a coefficient value range 0.455 to 0.818. Among the six pre-treatment methods analysed, thermochemical pre-treatment method was found effective with more than 95% improvement of many biomass characteristics. The range of voting given to the parameters identify %H be the important characteristic to be optimized first. The suggested method is validated by laboratory experiments and % accuracy between predicted and calculated biomass characteristic values showed more than 90% accuracy for all the biomass characteristic parameters tested in this study.

Keywords: Biomass waste, Bio-oil, Machine learning, Pearson matrix, Pre-treatment, Thermochemical liquefaction

Introduction

Interest in renewable energy sources has grown over the past few decades due to negative environmental effects generated by the traditional fossil fuels. Biomass is an alternative energy source that ranks as the fourth-most significant primary energy source in the world. Biomass is regarded as a desirable, pure, and eco-friendly raw material with little gas emission[1]. As a result of economic advancement and population increase, there has been an imbalance between the supply and demand of energy in practically every sector over time. There has been a substantial increase in the energy consumption by the transportation sector from 23% in 1971 to 29% in 2017, as compared to the overall final energy consumption by sector around the world[2]. Mainly, the emission of greenhouse gases has increased global warming concerns and researchers are tend to find out alternative sources as renewable options to minimize the dependence on fossil fuels. The biofuel production in worldwide has been scaled up. Also there have been instances of biorefineries that create uncertainty about their environmental effects[3]. The first generation of feedstock to biofuel has shown several issues such as water pollution, land use and degradation, high water consumption, and biodiversity loss. Second-generation feedstocks based on non-edible crops have been identified as a means of addressing the aforementioned issues, but they have some fundamental drawbacks, such as a high land demand and seasonal variations in the feedstocks[3]. In the coming years, it will be crucial to control the high demand for energy (particularly oil, which now dominates as a fuel) from the transportation sector. Additionally, according to BP's statistical analysis of the global energy sector, both energy consumption and carbon emissions are rising, at rates of 2.9% and 2%, respectively. Renewable and sustainable biomass sources can meet this escalating energy demand caused by the lack of non-renewable fossil fuel supplies[4].

Specifically grown plants, crop residues, timber, algae, fatty acids, edible plant oils, sewage waste, and food waste all makeup biomass feedstock. Pre-treatment techniques are crucial to deal with the resilience of lignocellulosic biomass, microalgae, and macroalgae biomass and make subsequent processes accessible and cost-effective[5]. Pre-treatment refers to procedures carried out on biomass feedstock before the primary transformation to overcome the natural barrier of resistance and disassemble the lignified cell wall, enhancing the effectiveness of subsequent processing into bioproducts[6]. Pretreatment techniques for lignocellulosic biomass include biological, physical, chemical, and physicochemical techniques. Physical techniques include milling, chipping, and grinding. Chemical techniques include using acid, alkaline, oxidants, organosolv, deep eutectic solvent (DES), and ionic liquid[7]. After the pretreatment of renewable sources, biochemical and thermochemical methods are typically used to convert biomass into biofuels. Through pyrolysis and gasification, the thermochemical processes can transform biomass, both food and non-food related into fuel products. A potential method is thermochemical gasification, which can extract the energy contained in different kinds of biomass and transform it into useful products appropriate for diverse industrial uses[8]. Agricultural crop leftovers, forest residues, energy crops, and organic municipal wastes are typical feedstocks for gasification. Pyrolysis, hydrodeoxygenation, and hydrothermal process are the most preferred techniques used to produce bio-oil from biomass[9,10]. There were many pretreatment techniques used to improve the yield of bio-oil. Especially, hydrothermal pretreatment is widely used in many commercial processes due to its lower cost, minimal erosive impacts on equipment, and low concentration of produced inhibitors[5].

According to the desired product, the hydrothermal process can be divided into three categories: hydrothermal liquefaction (HTL), hydrothermal gasification (HTG), and hydrothermal carbonization (HTC)[3]. To turn wet biomass into bio-oil at a moderate temperature (200-380°C), pressure (5-20 MPa), and time, the HTL process was chosen (15–60 min)[11]. A strategy to treat lignocellulosic biomass that uses liquid water or vapour water is known as hydrothermal pretreatment. The same method is also known by other names including autohydrolysis, hot-compressed water treatment, or liquid hot water pretreatment. Pyrolysis is a sophisticated thermal process that creates syngas from a material, similar to gasification, but at lower temperatures and without oxygen[12]. Additionally, it always comes first in the processes of gasification and combustion, where it is followed by either full or partial oxidation of the primary products. Despite having a higher calorific value than a gas created through gasification, pyrolysis often produces significantly less gas because there is no oxygen carrier.

Biomass pyrolysis and thermochemical processes are intricate processes and varies depending on several aspects like the lignocellulosic material composition, heating rate, and inorganic material content, etc.[13]. The cellulose, hemicellulose, and lignin content of biomass influence the gasification and pyrolysis, suggesting a wide variety of the effectiveness of various biomass and the processes employed.

Artificial intelligence and machine learning models help in predicting the required results based on the training given to it[15-17]. Supervised machine learning algorithms can be applied over the dataset for classification and regression[18-20]. Regression schemes can be applied over the dataset where input and output values are known[21], whereas classification schemes can be applied over the dataset which contains generic output variables[22,23]. Since thermochemical process involves known input and output variables, classification and regression schemes of supervised models provides high prediction precisions[24,25]. Supervised machine learning models having multiple input and output variables required a multi-variate regression[26,27]. Gopirajan *et al* have applied machine learning models to predict the bio-oil yield from the biomass through liquefaction and gasification process[28,29]. Cloud computing offers huge volume of storage capacity for dispersed dataset over wireless network[30]. This proposed study uses third party cloud storage offered by Amazon Web Services (AWS) to store and retrieve dataset[31,32]. Dataset stored in the cloud repository will be treated as the input to the machine learning model. This proposed study uses multiple biomass characteristics such as M (Moisture), VM (Volatile Matter), FC (Fixed carbon), A (Ash), C (Carbon), H (Hydrogen), O (Oxygen), N (Nitrogen) and S (Sulphur) which are considered as the dependent variables correlated with each other. *'Random Forest Classifier'* class package present in python was included in this proposed Multi-variate Dependent Random Forest (MDRF) model. An exclusively developed Multi-variate-based Decision Support System (MDSS) classification model suggests optimal pre-treatment method and suitable conditions for the optimization of the entire thermochemical conversion process.

Experimental Section

Biomass collection and characterization

The various biomasses were cultured and collected from Chennai, District of Tamil Nadu, India. The collected biomass was washed using the deionised water, air-dried and grinded into fine powder for further experiments. The characteristics of the biomass was then examined for its Carbon (%), Hydrogen (%), Nitrogen (%), Oxygen (%), and Sulphur (%) content of different biomass collected using an Elemental analyser (PerkinElmer 2400 series CHNS analyser). Further, the moisture content (%) and ash content (%) of different collected biomasses was also determined according to the standards of ASTM, 2006 and ASTM, 1995.

There are six types of pre-treatment technique considered in this study. They are P1: Ultrasound treatment, P2: Acid treatment, P3: Alkali treatment, P4: Thermochemical treatment, P5: Biological treatment and P6: Enzymatic treatment. For ultrasound treatment, about 100 g of biomass was suspended in 500 mL of water and ultrasonic waves of frequency 220 KHz were irradiated for about 30 min, then the biomass was filtered, dried and used for further studies. For acid treatment, 100 g of biomass was added to 500 mL of 10% HCl and mixed well using a mechanical shaker for 1 h at the speed of 100 rpm. After that, the biomass was filtered, washed with water three times and then dried to be used for further studies. For alkali treatment, about 10% NaOH solution of volume 500 mL was taken. 100 g of biomass was suspended in this solution and mixed well in a mechanical shaker for 1 h at a speed of 100 rpm. The biomass was then filtered, washed well with water for three times, dried and then used for further studies. For biological treatment, 100 g of biomass was suspended in 500 mL of water and 6 mL of *Cellulomonas flavigina* inoculum was added in to it. This mixture is kept in shaking incubator at a temperature of 35°C for 72 h. The shaking speed was maintained in 100 rpm. Then the biomass was filtered, washed with water, dried and then used for further studies. For enzymatic treatment, 100 g of biomass was suspended in 500 mL water and about 5 mL volume of 20% cellulase enzyme was added and incubated at a temperature of 35°C for 6 h. The mixture was shaken well at 100 rpm speed. After the reaction, the biomass was filtered, washed with water, dried and used for further studies.

Hydrothermal liquefaction

Each of the biomasses were liquefied individually using the stainless steel 250 mL capacity hydrothermal reactor (4598, PARR model reactor, Par Instrument Co, Moline, IL) equipped with an auto temperature controller. During the in-between process of each biomass in the hydrothermal reactor, the left-out residues in the reactor were washed with heated water at 400ºC for 4 h. About 15 g of biomass was processed by heating at varying temperatures from 200 – 340ºC for 1 h using 200 mL of solvents like ethanol and acetone. The hydrothermal liquefaction reaction was performed in the presence of a nitrogen atmosphere of about 5 MPa with the constant stirred reactor conditions at 720 rpm. Once after the processing of HTL experiments, the reactor was allowed to cool down and collected the gaseous products in airtight bags. Finally, dichloromethane was used to collect the bio-oil from the slurry obtained. Further, the excess dichloromethane present in the bio-oil was removed and the concentrated bio-oil was collected using a rotary vacuum evaporator. The obtained bio-oil yield was then estimated using the following Eq. (1).

$$Bio-oil\ yield\ (\%Y) = \left(\frac{Mass\ of\ bio-oil\ obtained}{Mass\ of\ biomass\ added}\right) \times 100 \quad \ldots (1)$$

Machine learning model

Fig. 1 shows the overall process flow of the proposed machine learning model used in this study. Input variables such as M, A, VM, FC, C, H, O, N and S were given to the central MDRF model. Decision tree with all the subsets were built based on the dependency between the input variables. This multivariate decision tree includes all the biomass characteristics, processing methods and conditions for predicting the best-fit pre-treatment process based in the voting suggested by the MDSS classification model. MDSS model predicts the yield based on the newly entered input variables as well as the existing input variables stored in the cloud repository. Finally, the biomass characteristics, predicted processing conditions (Temperature, Pressure and Time) and methods suggested by the decision support system were stored in the repository to form a dataset.

MDRF model

The process flow of the proposed MDRF model and complete process flowsheet for bio-oil extraction are shown in Figs 2 and 3, respectively. Data collection and pre- processing is the first step for dataset preparation. This study includes dataset collected from published[28,29,33–44] and unpublished work from the authors research team. The dataset

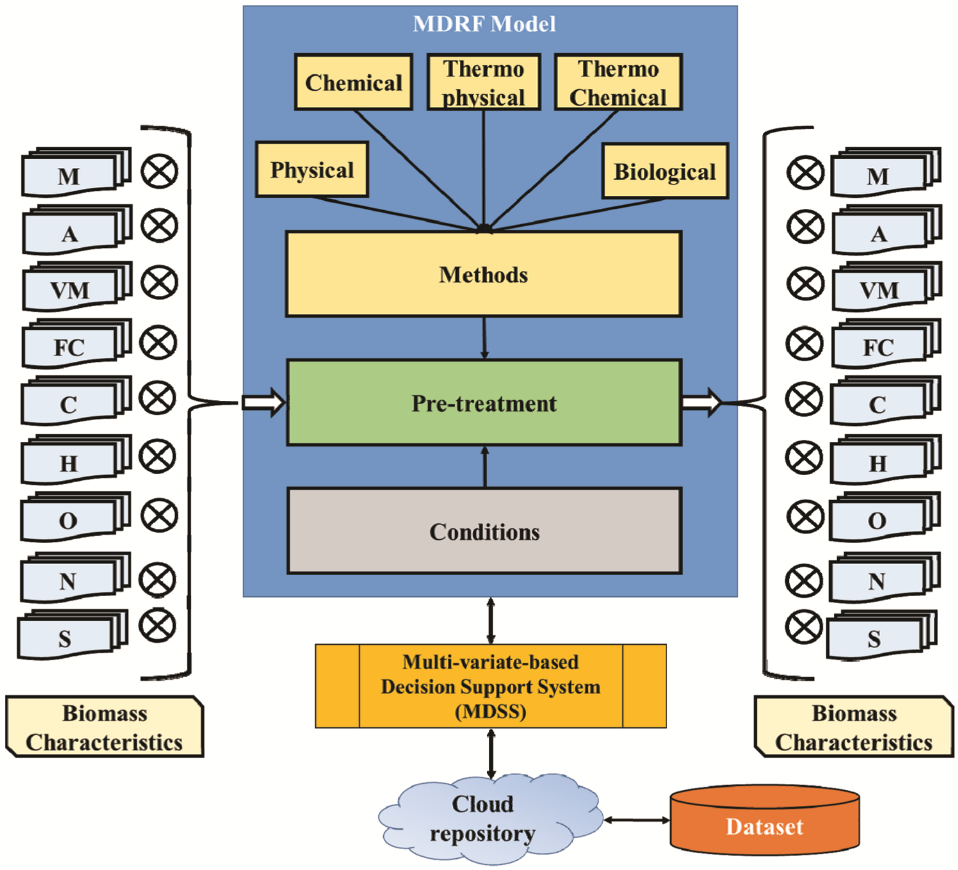


Fig. 1 — Machine learning model

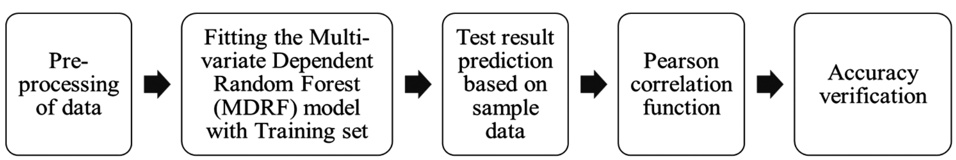


Fig. 2 — Machine learning model implementation process flow

includes, the biomass and its characteristics which are discussed in this paper, process conditions as mentioned in the paper and the corresponding bio-oil yield. Collected dataset was organized as '.csv' file and stored in the repository after the removal of redundant and noisy data. Error free dataset was fitted as the training set in the MDRF model using 'sklearn' library package available in Python. This proposed study uses 80:20 training and test data suggested by Roshan Joseph[45]. Around 700 biomass datasets were deposited in the cloud repository, out of which, 560 were considered as training set and 140 as test set. Pearson function suggests the correlation among the variables and helps to classify the variables as highly correlated and weakly correlated[46–50]. Pearson correlation function was calculated over the dataset to provide the correlation between the biomass variables. Accuracy of the proposed system was calculated with the results available obtained through experiments.

MDSS model

Multivariate based decision support system (MDSS) was constructed considering the following steps:

Step 1: In Multivariate Dependent Random Forest model, input all the records taken from the data set having 700 number of records.

Step 2: Construct Decision trees individually for each biomass including the subsets.

Step 3: Obtain output from all the subsets of all the decision trees.

Step 4: Calculate Subjective Majority Voting (SMV):

Step 4.1 Check If all the subsets of decision tree are complete

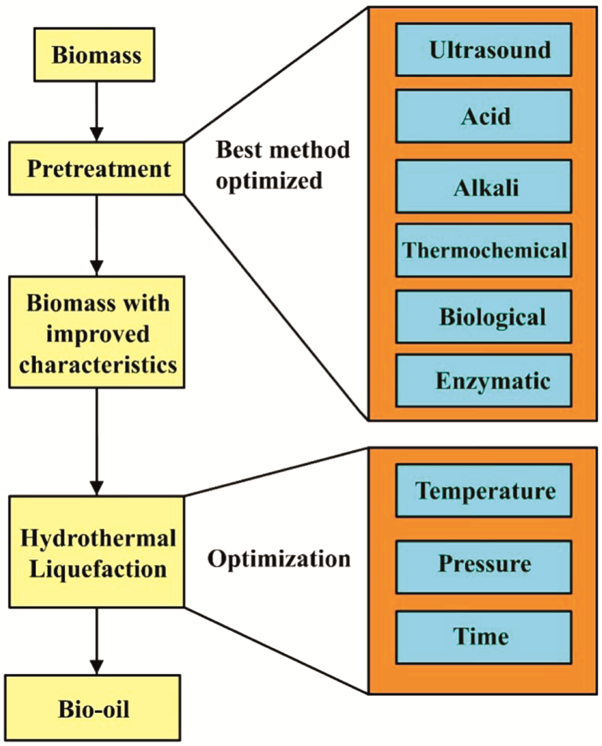


Fig. 3 — Complete process flowsheet

Step 4.2 Else perform Averaging for Classification and regression.

Step 5: Input

Majority voting helps in assigning individual strengths to the variables[51–54]. This proposed MDSS model uses majority voting (for complete decision tree) and averaging (for incomplete decision tree).

Results and Discussion

Effect of biomass characteristics on bio-oil yield

Bio-oil yield may be affected by various biomass characteristics. Fig. 4 shows the effect of biomass characteristics such as moisture content (M), ash content (A), volatile matter (VM), fixed carbon (FC), carbon (C), hydrogen (H), oxygen (O), nitrogen (N) and sulphur (S) on %yield. The characteristics such as M, VM, FC, C, H and N have positive effect on bio-oil yield, whereas others like ash content (A), O and sulfur (S) have negative effect. The presence of ash content reduces the available carbon which reduces the yield. Oxygen and sulphur inhibit the liquefaction reaction and their presence reduces the oil yield. Based on this observation it will be easy to decide the factors that should be reduced and the factors that should be improved by pre-treatment. This will help in choosing the suitable method to influence the change of favourable characteristics. The trend observation further shows that carbon and hydrogen

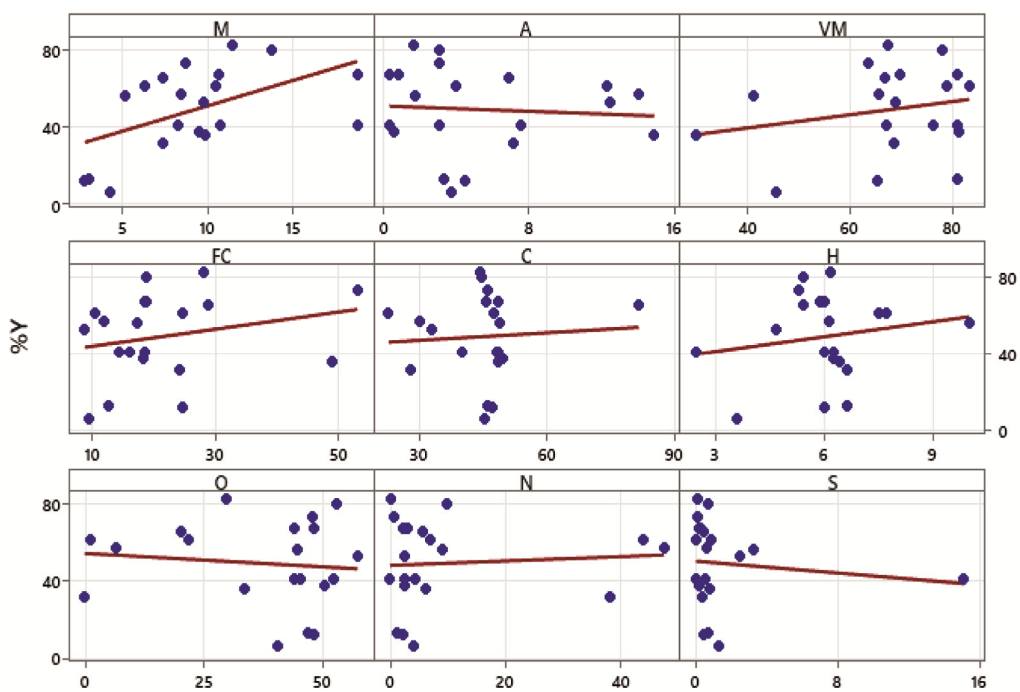


Fig. 4 — Trend analysis of influence of input biomass characteristics on bio-oil yield

content have synergistic effect on bio-oil yield. This can be clearly seen from Fig. 3 as the data points in the same level have a drastic change in the % Y.

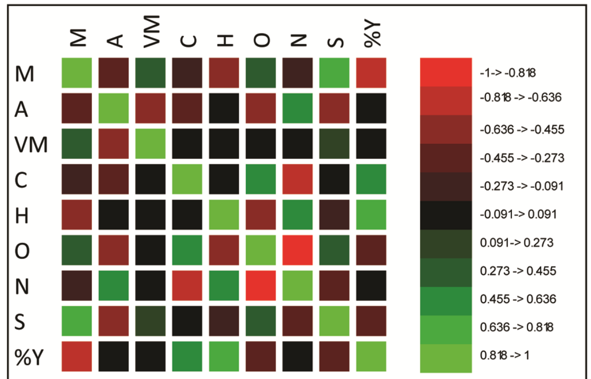
Correlation between biomass characteristics and bio-oil yield

Analysis of correlation between biomass characteristics and bio-oil yield was carried out using Pearson correlation method. The correlation matrix is shown in Fig. 5. This matrix helps in understanding the intensity of correlation between the parameters. Also, this correlation shows the way the machine learning model was trained to consider the parameters. The importance to various parameters will be given based on the correlation coefficient and its positivity or negativity to the response, in this case bio-oil yield (%Y). The important influencing parameters %C and %H have a very good positive correlation with a coefficient value range 0.455 to 0.818 positive in which %H has higher value whereas, the %ash content and %Oxygen has higher negative coefficient values -0.2 to -0.6. These values suggest the importance of negative characteristics which should be considered for modification. Interestingly, %moisture content has higher coefficient on the negative side (-0.818 to -0.636), which during the trend statistical analysis, showed a positive trend.

Technically, the higher moisture content of biomass results in the lower carbon content and other biodegradable contents. This would further reduce the oil yield and result in the less %Y values. Despite the addition of water in the liquefaction process, the presence of residual moisture in the biomass has a negative influence. Though the mixture of interactions among parameters studied and %Y, the interaction between two individual characteristics would highly depend on the way they affect each other. For example, the moisture content affects each parameter in a different way. Presence of moisture may interact with ash content in a negative way so the %Y would reduce. Similarly the interaction with oxygen would be positive for the parameters but the %Y will be affected in a negative way because oxygen is not favorable in the liquefaction reaction. The carbon content is affected by moisture in the negative manner and this would reduce the yield. The positive interactions between yield and other parameters are important. All the positive effectors would have positive effect when they interact with each other in a positive manner. Even the negative effectors interact each other in a positive manner, it will surely affect the yield in a negative way.

Effect of pre-treatment on biomass characteristics

There are six types of pre-treatment technique considered in this study. They are P1: Ultrasound treatment, P2: Acid treatment, P3: Alkali treatment, P4: Thermochemical treatment, P5: Biological treatment and P6: Enzymatic treatment. The improvement of each biomass characteristic was calculated based on before and after pre-treatment values. Fig. 6 shows the %Improvement of



M: Moisture; VM: Volatile Matter; FC: Fixed carbon, A: Ash, C: Carbon, H: Hydrogen, O: Oxygen; N: Nitrogen; S: Sulphur

Fig. 5 — Pearson correlation matrix for the biomass characteristics after pre-treatment methods and conditions versus bio-oil yield (%Y)

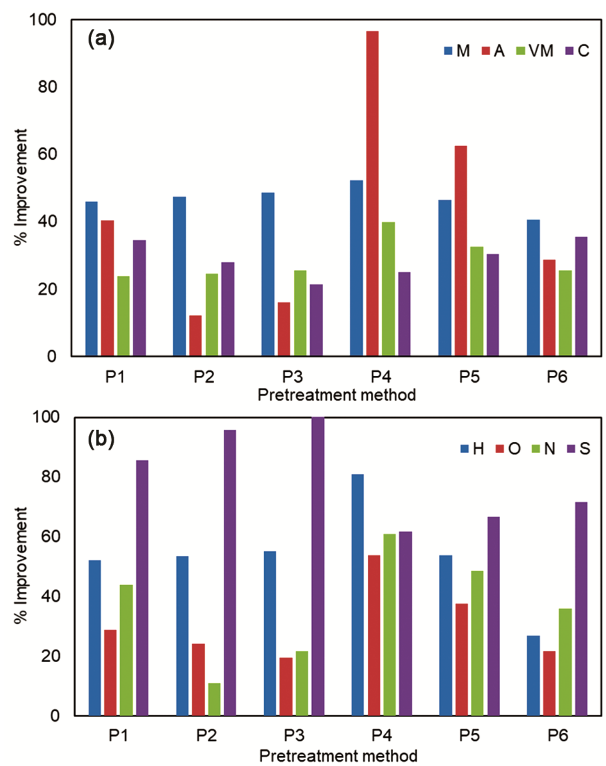


Fig. 6 — %Improvement of biomass characteristics by various pre-treatment methods (P1: Ultrasound, P2: Acid, P3: Alkali, P4: Thermochemical, P5: Biological, P6: Enzymatic)

each biomass characteristic according to the pre-treatment method. The results clearly show that each pre-treatment method has a significant improvement in the biomass characteristic. This is due to the removal of unwanted impurities and concentrating the effective biomass levels. Despite the similarity in %Improvement in all the methods, P4: Thermochemical pre-treatment method has shown high improvement of all the biomass characteristics.

Preference of selection in biomass characteristics for optimization

The optimization algorithm works based on the preferences for the selection of biomass characteristics to be optimized. The algorithm assigns a range of voting to optimize the parameters. Table 1 shows the range of voting parameters assigned by the algorithm for various biomass characteristics. V1 being the highest range, the votes descend based on the importance of parameters to a level of V8. The frequency of number of times the particular parameter is voted to a particular level is mentioned in the table. Taking V1 as an example, it is clearly seen that the biomass characteristic ‘H’ (%hydrogen) was frequently voted for this level. The voting score range is 90, whereas the ‘C’ (%carbon) was voted to the second level in this range. Similarly the voting range varies for every parameter. The algorithm works based on this voting range and starts optimizing the value highly voted in V1 and moving on to the next values. The high frequented value in V8 was optimized at the last. Also it can be noted that, the initial voting ranges were clearly organized and the final values were stumbling. But, the margin of range was very clear that the parameter for each voting line was selected properly.

Validation of predicted parameters

MDRF algorithm predicted the suitable pre-treatment method based on the input biomass characteristic values and also predicted the final biomass characteristic values that improve due to the pre-treatment process. These data were validated by laboratory experiments to analyse the suitability of the algorithm. Fig. 7 shows the %accuracy values of biomass characteristics predicted by the algorithm. A model biomass was taken for the analysis. It is noted that all the biomass characteristics were predicted with more than 90% accuracy by the algorithm. Many of these parameters were predicted even with more than 95% accuracy which shows that this algorithm is suitable for the optimization process pertaining to the thermochemical conversion techniques.

Table 1 — Range of voting for parameters

Parameter	V1	V2	V3	V4	V5	V6	V7	V8
M	0	1	82	2	3	2	3	3
A	0	1	2	2	3	69	9	11
VM	0	1	2	1	2	8	11	59
C	9	88	7	1	3	0	4	4
H	90	6	3	0	1	3	5	5
O	0	2	2	79	10	3	2	3
N	1	1	1	4	3	10	62	12
S	0	0	1	11	75	5	4	3

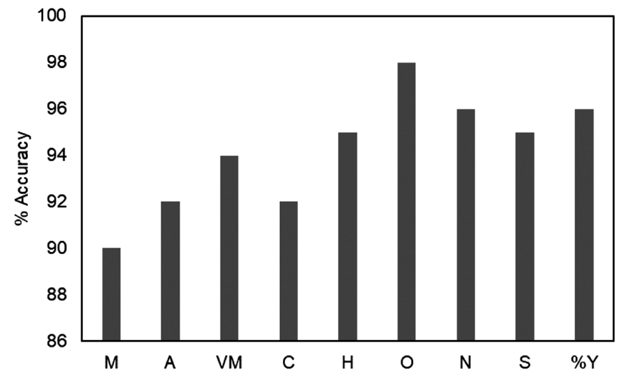


Fig. 7 — %Accuracy of biomass characteristics predicted based on the MDRF algorithm

Conclusion

This study involves the development of a machine learning algorithm based on random forest technique to optimize the pre-treatment method applied to the biomass to improve its characteristics to get optimum bio-oil yield using thermochemical conversion processes. The correlation of parameters showed that %H and %C were the important characteristics that influence the oil yield values. Furthermore, a range of voting was applied by the algorithm to initiate the optimization process from the foremost important parameter which in this case was %H. The validation of the predicted parameters was carried out based on the laboratory studies and the %accuracy values showed that all the biomass characteristics have been predicted with more than 90% accuracy. This MDRF algorithm is a promising technique to optimize the pre-treatment based thermochemical conversion systems, to get higher bio-oil yields from any suitable biomass.

References

1 Mateus M M, Bordado J M & dos Santos R G, Estimation of higher heating value (HHV) of bio-oils from thermochemical liquefaction by linear correlation, *Fuel*, 302 (2021) 121149.

2 Ahamed T S, Anto S, Mathimani T, Brindhadevi K & Pugazhendhi A, Upgrading of bio-oil from thermochemical conversion of various biomass–Mechanism, challenges and opportunities, *Fuel*, 287 (2021) 119329.
3 Brindhadevi K, Anto S, Rene E R, Sekar M, Mathimani T, Chi N T L & Pugazhendhi A, Effect of reaction temperature on the conversion of algal biomass to bio-oil and biochar through pyrolysis and hydrothermal liquefaction, *Fuel*, 285 (2021) 119106.
4 Chan Y H, Loh S K, Chin B, Yiin C L, Shen H B, Cheah K W, Wong M K, Loy A C M, Gwee Y L, Lo S L Y, Suzana Y & Lam S, Fractionation and extraction of bio-oil for production of greener fuel and value-added chemicals: Recent advances and future prospects, *Chem Eng J*, 397 (2020) 125406.
5 Sun Q, Wei-Jing C, Pang B, Sun Z, Lam S S, Sonne C & Tong-Qi Y, Ultrastructural change in lignocellulosic biomass during hydrothermal pretreatment, *Bioresour Technol*, 341 (2021) 125807.
6 Badiei M, Asim N, Jahim J M & Sopian K, Comparison of chemical pretreatment methods for cellulosic biomass, *APCBEE Procedia*, 9 (2014) 170.
7 Li X, Shi Y, Kong W, Wei J, Song W & Wang S, Improving enzymatic hydrolysis of lignocellulosic biomass by bio-coordinated physicochemical pretreatment—A review, *Energy Rep*, 8 (2022) 696.
8 Zabed H M, Akter S, Yun J, Zhang G, Awad F N, Qi X & Sahu J N, Recent advances in biological pretreatment of microalgae and lignocellulosic biomass for biofuel production, *Renew Sust Energy Rev*, 105 (2019) 105.
9 Sankaran R, Cruz R A P, Pakalapati H, Show P L, Ling T C, Chen W H & Yang T, Recent advancements in the pretreatment of microalgal and lignocellulosic biomass: A comprehensive review, *Biores Tech*, 298 (2020) 122476.
10 Chandraratne M R & Daful A G, Recent Advances in Thermochemical Conversion of Biomass, In *Recent Perspect Pyrol Res*, Intech open (2021).
11 Gopirajan P V, Gopinath K P, Sivaranjani G & Arun J, Optimization of hydrothermal liquefaction process through machine learning approach: Process conditions and oil yield, *Biomass Convers Biorefin*, 13 (2021) 1.
12 Djandja O S, Yin L, Wang Z, Guo Y, Zhang X & Duan P, Progress in thermochemical conversion of duckweed and upgrading of the bio-oil: A critical review, *Sci Total Environ*, 769 (2021) 144660.
13 Pandey A, Bhaskar T, Stöcker M & Sukumaran R, *Recent advances in thermochemical conversion of biomass*, 1st Edition, Elsevier (2015).
14 Canabarro N, Soares J F, Anchieta C G, Kelling C S & Mazutti M A, Thermochemical processes for biofuels production from biomass, *Sust Chem Process*, 1 (2013) 1.
15 McArthur J J, Shahbazi N, Fok R, Raghubar C, Bortoluzzi B & An A, Machine learning and BIM visualization for maintenance issue classification and enhanced data collection, *Adv Eng Inform*, 38 (2018) 101.
16 Antonopoulos I, Robu V, Couraud B, Kirli D, Norbu S, Kiprakis A, Flynn D, Elizondo-Gonzalez S & Wattam S, Artificial intelligence and machine learning approaches to energy demand-side response: A systematic review, *Renew Sust Energy Rev*, 130 (2020) 109899.
17 Sun H, Burton H V & Huang H, Machine learning applications for building structural design and performance assessment: State-of-the-art review, *J Build Eng*, 33 (2021) 101816.
18 Corrales D C, Ledezma A & Corrales J C, A case-based reasoning system for recommendation of data cleaning algorithms in classification and regression tasks, *Appl Soft Comput J*, 90 (2020) 106180.
19 Kamilaris A & Prenafeta-Boldú F X, Deep learning in agriculture: A survey, *Comput Electron Agric*, 147 (2018) 70.
20 Dogan A & Birant D, Machine learning and data mining in manufacturing, *Expert Syst Appl*, 166 (2021) 114060.
21 Wen S, Buyukada M, Evrendilek F & Liu J, Uncertainty and sensitivity analyses of co-combustion/pyrolysis of textile dyeing sludge and incense sticks: Regression and machine-learning models, *Renew Energy*, 151 (2020) 463.
22 Bozek J, Mustra M, Delac K & Grgic M, A Survey of Image Processing Algorithms in Digital Mammography BT - Recent Advances in Multimedia Signal Processing and Communications, Grgic M, Delac K & Ghanbari M, Eds., Berlin, Heidelberg: Springer Berlin Heidelberg, (2009) 631.
23 Ma X, Sha J, Wang D, Yu Y, Yang Q & Niu X, Study on a prediction of P2P network loan default based on the machine learning LightGBM and XGboost algorithms according to different high dimensional data cleaning, *Electron Commer Res Appl*, 31 (2018) 24.
24 Habyarimana E, Piccard I, Catellani M, De F P & Agata M D, Towards predictive modeling of sorghum biomass yields using fraction of absorbed photosynthetically active radiation derived from sentinel-2 satellite imagery and supervised machine learning techniques, *Agronomy*, 9 (2019) 203.
25 Dekker A, Dehing-Oberije C, Ruysscher D D, Lambin P, Komati K, Fung G, Yu S, Hope A, Neve W D & Lievens Y, Survival prediction in lung cancer treated with radiotherapy: Bayesian networks vs. support vector machines in handling missing data, *8th International Conference on Machine Learning Applications*, *ICMLA* (2009) 494.
26 Zhao Y, Xia X H, Yang Z F & Wang F, Assessment of water quality in Baiyangdian Lake using multivariate statistical techniques, *Procedia Environ Sci*, 13 (2012) 1213.
27 Buvé C, Saeys W, Rasmussen Mo A, Neckebroeck B, Hendrickx M, Grauwet T & Loey A V, Application of multivariate data analysis for food quality investigations: An example-based review, *Food Res Int*, 151 (2022) 110878.
28 Gopirajan P V, Gopinath K P, Sivaranjani G & Arun J, Optimization of hydrothermal liquefaction process through machine learning approach: Process conditions and oil yield, *Biomass Convers Biorefin*, 13 (2021) 1213.
29 Gopirajan P V, Gopinath K P, Sivaranjani G & Arun J, Optimization of hydrothermal gasification process through machine learning approach: Experimental conditions, product yield and pollution, *J Clean Prod*, 306 (2021) 127302.
30 Deepavathi P & Mala C, Privacy-secure link utilization routing algorithm (PSLU) for improving performance in IoT

and cloud computing BT - IoT and cloud computing for societal good, Verma J K, Saxena D & González-Prida V, Eds., Cham: *Springer International Publishing*, (2022) 297.

31 Hofmann W, Lang S, Reichardt P & Reggelin T, A brief introduction to deploy Amazon Web Services for online discrete-event simulation, *Procedia Comput Sci*, 200 (2022) 386.

32 Radford C J, Challenges and solutions protecting data within Amazon Web Services, *Netw Secur*, 2014 (2014) 5.

33 Vinitha N, Vasudevan J & Gopinath K P, Bioethanol production optimization through machine learning algorithm approach: Biomass characteristics, saccharification, and fermentation conditions for enzymatic hydrolysis, *Biomass Convers Biorefin*, 13 (2022) 7287.

34 Wang J, Zhang X, Li Z, Ma Y & Ma L, Recent progress of biomass-derived carbon materials for supercapacitors, *J Power Sources*, 451 (2020) 227794.

35 Toor S S, Rosendahl L & Rudolf A, Hydrothermal liquefaction of biomass: A review of subcritical water technologies, *Energy*, 36 (2011) 2328.

36 Bhaskar T, Sera A, Muto A & Sakata Y, Hydrothermal upgrading of wood biomass: Influence of the addition of K_2CO_3 and cellulose/lignin ratio, *Fuel*, 87 (2008) 2236.

37 Gautam P, Neha, Upadhyay S N & Dubey S K, Bio-methanol as a renewable fuel from waste biomass: Current trends and future perspective, *Fuel*, 273 (2020) 117783.

38 Karagöz S, Bhaskar T, Muto A & Sakata Y, Catalytic hydrothermal treatment of pine wood biomass: Effect of RbOH and CsOH on product distribution, *J Chem Technol Biotechnol*, 80 (2005) 1097.

39 Kruse A, Hydrothermal biomass gasification, *J Supercrit Fluids*, 47 (2009) 391.

40 Ali S, Shafique O, Mahmood S, Mahmood T, Khan B A & Ahmad I, Biofuels production from weed biomass using nanocatalyst technology, *Biomass Bioenergy*, 139 (2020)105595.

41 Yang J, He Q & Yang L, A review on hydrothermal co-liquefaction of biomass, *Appl Energy*, 250 (2019) 926.

42 Biswas B, Kumar A, Fernandes A C, Saini K, Negi S, Muraleedharan U D & Baskar T, Solid base catalytic hydrothermal liquefaction of macroalgae: Effects of process parameter on product yield and characterization, *Biores Tech*, 307 (2020) 123232.

43 Yan L, Wang Y, Li J, Zhang Y, Ma L, Fu F, Chen B & Liu H, Hydrothermal liquefaction of Ulva prolifera macroalgae and the influence of base catalysts on products, *Biores Tech*, 292 (2019) 121286.

44 Muppaneni T, Reddy H K, Selvaratnam T, Dandamudi K P R, Dungan B, Nirmalakhandan N, Schaub T, Holguin F O, Voorhies W, Lammers P & Deng S, Hydrothermal liquefaction of Cyanidioschyzon merolae and the influence of catalysts on products, *BioresourTechnol*, 223 (2017) 91.

45 Joseph V R, Optimal ratio for data splitting, *Stat Anal Data Min*, 15 (2022) 531.

46 Sedgwick P, Pearson's correlation coefficient, *BMJ (Online)*, 345 (2012) 1.

47 Zhou H , Deng Z, Xia Y & Fu M, A new sampling method in particle filter based on Pearson correlation coefficient, *Neurocomputing*, 216 (2016) 208.

48 Baak M, Koopman R, Snoek H & Klous S, A new correlation coefficient between categorical, ordinal and interval variables with Pearson characteristics, *Comput Stat Data Anal*, 152 (2020) 107043.

49 Mu Y, Liu X & Wang L, A Pearson's correlation coefficient based decision tree and its parallel implementation, *Inf Sci*, 435 (2017) 40.

50 Fu T, Tang X, Cai Z, Zuo Y, Tang Y & Zhao X, Correlation research of phase angle variation and coating performance by means of Pearson's correlation coefficient, *Prog Org Coat*, 139 (2019) 105459.

51 Chen Z, Jiang L & Li C, Label augmented and weighted majority voting for crowdsourcing, *Inf Sci*, 606 (2022) 397.

52 Cardak B A, Glomm G & Ravikumar B, Majority voting in a model of means testing, *Eur Econ Rev*, 122 (2020) 103351.

53 Xu Z, Mei X, Wang X, Yue M, Jin J, Yang Y & Li C, Fault diagnosis of wind turbine bearing using a multi-scale convolutional neural network with bidirectional long short term memory and weighted majority voting for multi-sensors, *Renew Energy*, 182 (2022) 615.

54 Witt C, How majority-vote crossover and estimation-of-distribution algorithms cope with fitness valleys, *Theor Comput Sci*, 940 (2022) 18.