Applied Energy Symposium 2019: Low carbon cities and urban energy systems
October 16-18, 2019, Xiamen, China
Paper ID: 0015

MULTI-TASK PREDICTION OF FUEL PROPERTIES OF HYDROCHAR DERIVED FROM WET MUNICIPAL WASTES WITH RANDOM FOREST

Jie Li, Xinzhe Zhu, Yinan Li, Yen Wah Tong, Xiaonan Wang*
Department of Chemical and Biomolecular Engineering, National University of Singapore, Singapore 117585, Singapore

ABSTRACT

Waste to energy is a promising way to ease the urban burden of waste treatment and hydrothermal carbonization (HC) can dewater the municipal wastes with high moisture efficiently with hydrochar left. The hydrochar with outstanding fuel characteristics can be used as fuel for incineration to generate power. To predict the fuel characteristics of hydrochar including the yield, higher heating value (HHV) and carbon content (C_char) based on the information of the wet municipal waste, machine learning methods have been explored in this work. Results show that the optimized Random Forest (80 trees with 10 maximum depths) has good multi-task prediction capability of fuel characteristics. The R^2 for the predictions of the yield, HHV and C_char are 0.80, 0.91 and 0.95, respectively. Moreover, according to the feature importance analysis, the yield of hydrochar is mainly determined by the temperature and water content of HC, while the HHV and C_char are dominated by the carbon and ash content of the feedstock, respectively.

Keywords: machine learning, multi-task prediction, waste to energy, hydrothermal carbonization, hydrochar

Nonmenclature

Abbreviations	
C_char	carbon content of hydrochar
Fc	fix carbon
HC	hydrothermal carbonization
HHV	higher heating value
ML	machine learning
MSE	mean squared error
MSW	municipal solid waste
PCC	Pearson correlation coefficient
RF	Random Forest
V	volatile matter
t	reaction time
T	reaction temperature
WC	water content
Symbols	
x_i	the value of *i* th input feature *x*
x_i^*	the normalized value of initial x_i
u	the mean of x_i
s	the standard deviation of x_i
ρ_{xy}	the value of PCC for feature to target or target to target
$\bar{x}$	the mean of input feature *x*
$\bar{y}$	the mean of output target *y*
R_j^2	the coefficient of determination of target *j*
$Y_{pred,i}^j$	the predicted value of target *j*
$Y_{exp,i}^j$	the experimentally value of target *j*
$Y_{exp,ave}^j$	the average of the experimental value of target *j*
MSE_j	the mean squared error of target *j*
MSE	the average mean squared error of multi-task targets

1. INTRODUCTION

With the development of economy and urbanization, more and more municipal solid waste (MSW) is generated by human activities due to the rising population and their living standards. It is reported that one person would produce about 477 kg per year averagely in Europe [1]. In order to dispose the MSW, lots of approaches have been investigated and the waste to energy technology is regarded as one of the most promising way. Incineration for power generation is a prevalent technology for converting MSW into energy because the relative high heating value (about 10 MJ/kg) of MSW [2]. However, some of MSW contain high moisture, such as food waste from catering industry and household, sewage sludge from municipal sewage

treatment plant, and animal manure from livestock and poultry industry. Those wet organic wastes are unable to be utilized for incineration before dewatering and drying. However, the dewatering and drying of those wet MSW (especially sludge) are energy-intensive processes due to their special water-trapped structure [3]. Hence, it is necessary to employ efficient dehydration technologies with lower energy consumption to pretreat those wet MSW.

Hydrothermal carbonization (HC) is a thermochemically decomposed process surrounded by hot compressed water, which is able to destroy the structure of wet MSW and let the bound water come out easily with hydrochar left [4, 5]. The hydrochar is a versatile product which can be used as fuel, fertilizer, and even adsorption material due to its high content of carbon, functional groups, and developed porous structure [6-8]. In addition, it was found that substituting coal-based energy sources with hydrochar to produce electricity can make contribution to the negative carbon emission [9]. Hence, it is extremely essential to understand the fuel characteristics of hydrochar. The yield, higher heating value (HHV) and carbon content (C_char) of hydrochar are critical indexes to judge whether it is qualified as a fossil fuel alternative for power generation [2]. In order to understand those important characteristics, the traditional way is to conduct the HC experiments and then detect its characteristics one by one. This process is quite expensive, labour intensive and time consuming [10]. Therefore, it is advantageous to employ machine learning (ML) methods trained on empirical data for those characteristics predictions of hydrochar. In the ML models, the relevance between input features and output targets can be achieved from a training dataset, and then the relevance can be further generalized to make predictions from other new input features [11].

Recently, the applications of ML have been widely investigated in waste treatment and other environmental issues, such as the prediction of bioavailability and toxicity of chemical mixtures in soil, the forecasting of MSW generation and the prediction of PM2.5 concentration in atmosphere [10, 12, 13]. Furthermore, the role of feedstock properties and process conditions on products from hydrothermal carbonization has been investigated by Li et al., and the models are applied by Ro, et al., while its prediction accuracy were not high [14, 15]. It can also be found that most of the above mentioned ML work is only one output target. Therefore, in this study, to predict the yield, HHV and C_char of hydrochar at the same with high accuracy, the multi-task learning of Random Forest (RF) was investigated based on the collected data of HC of food waste, sludge and manure. The 10-cross validation was used for hyper-parameter tuning of RF to get an optimized model with good prediction capacity. Moreover, the feature importance of feedstock and process conditions are analysed according to the optimized RF model.

2. MATERIAL AND METHODS

2.1 Data collection

A literature survey about the HC of food waste, sewage sludge and animal manure was conducted. The keywords such as food waste, sludge manure, hydrothermal carbonization, hydrochar were used in the search process. The element analysis (C, H, N, O) and proximate analysis including fix carbon (Fc), ash (A), volatile matter (V) data of those feedstock and the HC conditions data including reaction time (t), temperature (T), and water content (WC) were collected. As for the characteristics of hydrochar, the data of yield, HHV and C_char were gathered. A total of 248 data entries with the complete information were accumulated in this work.

2.2 Data preprocessing

In order to make uniform the domain of the variables, all the data of input feature and output targets were normalized and the value is calculated as equation (1)[16]:

$$x_i^* = \frac{x_i - u}{s} \qquad (1)$$

where x_i is the value of input feature *i*, x_i^* was the normalized value of initial x_i. The u and s were the mean and the standard deviation of x_i, respectively.

To make rough understanding of the correlation rapidly between the input feature and output targets, the Pearson correlation coefficient (PCC) was employed, and the ρ_{xy} was calculated according to Equation (2) [17]:

$$\rho_{xy} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})\sum_{i=1}^{n}(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2}} \qquad (2)$$

where ρ_{xy} is the value of PCC for feature to target or target to target, $\bar{x}$ or $\bar{y}$ denoted the mean of input feature *x* or output target *y*. The Pearson correlation coefficient is a measure of vector similarity. The range of ρ_{xy} is -1 to +1, 0 means no linear correlation, negative

and positive value means negative and positive correlation, respectively.

In addition, in the model building process, the total data were divided into two parts. The nine-tenth of the total data selected randomly were training data, and the left one-tenth data were used as test data for the final evaluation of the developed model. In order to improve the prediction ability of developed models, 10-fold cross validation was used in the training process for hyper-parameter tuning [18]. The origin training data were randomly divided into 10 equal subsamples. The 9 subsamples were used to train model, while the remaining one was retained as validation data to validate the performance of the trained model. The procedure was repeated by 10 times and the results were averaged to choose a final model.

2.3 Modelling methods and Model evaluation

In this work, the RF regression method in the Scikit-Learn was employed through Python to investigate the prediction of characteristics of hydrochar, and the final performance of the RF is the average of the trees. In the RF model, the number of trees and the maximum depth of the trees were important hyper-parameters. Hence they are adjusted from 1 to 100 and 1 to 20 to achieve the optimized model, respectively [19]. Furthermore, the importance of input features can be analysed through the optimized RF[20]. The deterioration of model quality is measured as descriptor importance when each feature is replaced by random noise in the out-of-bag validation [21]. Therefore, the relative importance of each input variable could be obtained according to the degree of deterioration.

The fit goodness of the RF model was evaluated using a coefficient of determination (R^2) [10]. The error associated with each model was assessed by using the mean squared error [22, 23], and they were calculated according to the equation (3-5):

$$R_j^2 = 1 - \frac{\sum_{i=1}^{n}\left(Y_{pred,i}^{j} - Y_{exp,i}^{j}\right)^2}{\sum_{i=1}^{n}\left(Y_{exp,i}^{j} - Y_{exp,ave}^{j}\right)^2} \quad (3)$$

$$MSE_j = \frac{\sum_{i}^{n}(Y_{pred,i}^{j} - Y_{exp,i}^{j})^2}{n} \quad (4)$$

$$MSE = \frac{\sum_{j}^{m} MSE_j}{m} \quad (5)$$

where, $Y_{pred,i}^{j}$ represents the predicted value of target *j*, $Y_{exp,i}^{j}$ represents the experimentally value of target *j*, and *n* represents the number of data, and $Y_{exp,ave}^{j}$ is the average of the experimental value of target *j*, MSE_j is the mean squared error of target *j* and *MSE* is the average mean squared error of multi-task targets.

3. RESULTS AND DISCUSSION

3.1 Statistical analysis of feedstock and hydrochar characteristics

In order to quickly understand the correlation for features to targets and targets to targets, the PCC of feature to targets and targets to targets was calculated and shown in Fig. 1. As can be seen, both HHV and C_char of hydrochar were the most positively correlated with the content of C of feedstock, and followed by the H and V contents, while a strongly negative response to the ash content. However, the yield was positively related to the water content of HC process, and negatively correlated with the T of HC. In terms of the targets to targets correlation, the three targets had an outstanding correlation to each other. The HHV and C_char had positively relevance to each other, and the yield was negatively related to the HHV and C_char. The results show that these three targets had a close correlation to each other, indicating the possibility for the multi-task prediction of machine learning.

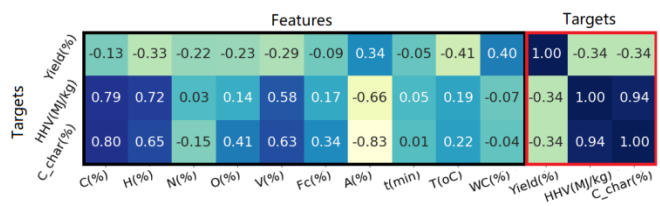

Fig 1 Pearson correlation coefficient of features to targets and targets to targets.

3.2 Hyper-parameter tuning and evaluation of RF

Fig. 2 shows the 10-cross validated MSE values with different values of hyper-parameters. It is found that the MSE decreased sharply first with the increasing of the maximum depth of trees from 1 to 5 in the RF, and then kept a stationary trend with the maximum depth of trees equal and more than 10. Another notable result is that the elevation of numbers of trees had slightly impact on the 10-times MSE under the condition of the maximum depth of trees was lower than 5. However, when the maximum depth of trees was more than 5, the 10-times MSE reduced significantly in the beginning with the increasing of numbers of trees, while the downward trend moderated with the numbers of trees from 10 to 80. From the above-mentioned results, it is found that the optimized hyper-parameters of RF are 80 trees with the maximum depth of 10.

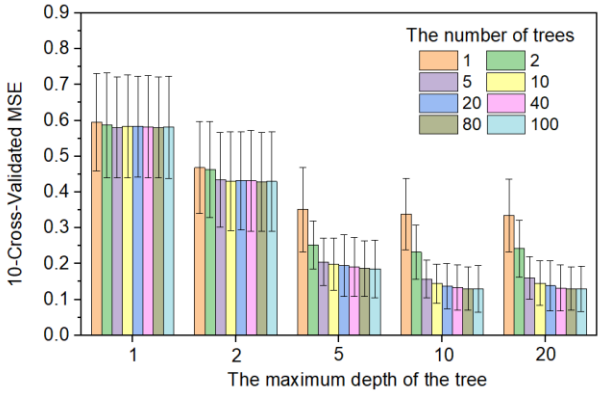

Fig. 2. Results of hyper-parameter tuning for RF.

In order to verify whether the optimized RF is appropriate for the non-training data, the left one-tenth collected data was used to evaluate its multi-task prediction capability. Both R^2 and MSE were used to evaluate the accuracy of the optimized RF. The value of R^2 is more closed to 1.0 and the MSE value is smaller, indicating a better accuracy of the RF. The value of the optimized model multi-task predicted data divided by the corresponding original data values from the experiment was plotted for evaluation. The prediction results of yield, HHV and C_char of hydrochar are shown in Fig. 3. A tighter cloud of points about the y=1 line indicates a better and more reliable predicting performance for the targets. Furthermore, the R^2 and MSE values of HHV and C_char were all more than 0.90 and lower than 0.10, implying that the optimized RF fit the test data of HHV and C_char quite well. It means that the optimized RF had a high capability in predicting HHV and C_char as long as given the basic information (values of input features) of feedstock. However, it is worthwhile to note that the R^2 of yield was 0.80 which suggests a lower accuracy of prediction.

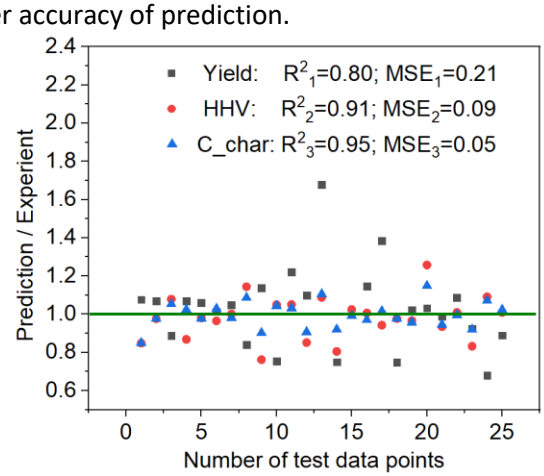

Fig 3 Evaluated results of multi-task prediction of the optimized RF.

3.3 Analysis of relative feature importance

In order to understand the feature importance of individual target, we re-run the RF with each single target three times to get the feature importance of the three targets. As shown in Fig. 4, different targets possessed different important influential factors. The yield of hydrochar was mainly influenced by three features which are the A, T and WC of the reactor. Furthermore, it is obvious that two of the three important factors were process parameters and they accounted for 43%, indicating that the process conditions, especially the T and WC had a significant influence on the yield of hydrochar. The similar result was also obtained by Li et al. [24]. As for HHV, the most important feature is C content which was followed by A and H contents, and all of them belonged to the characteristics of feedstock. It recommends that the HHV was determined by the performance of feedstock, especially by the C content. The explanation for the result can be found in the previous research work [4]. Hence, we can make a decision that whether the hydrochar is qualified as fuel for substituting of fossil fuel according to the C, A and H contents of feedstock with the RF model. Another notable result is that the A content accounted for as high as 69% in the analysis of feature importance for C_char. It is beyond our expectation that the dominated factor for determination of C_char of hydrochar is the A content rather than the C contents. In addition, as for the three targets together, it is interesting that the A is the most outstanding feature. The possible reason is that the A content does not only influence the content of valuable compounds in the feedstock, but also participates in the chemical reaction of compounds decomposition in the HC process [25]. In general, the above results imply that the yield of hydrochar is mainly related to the process conditions, while the HHV and C_char are dominated by the characteristics of the feedstock.

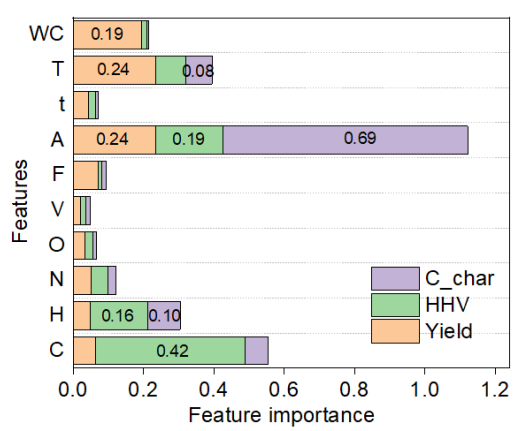

Fig 4 Evaluated results of multi-task prediction of the optimized RF.

3.4 Limitations of the optimized RF and future work

The prediction of those targets (yield, HHV and C_char) of hydrochar can make us understand the basic fuel characteristics of hydrochar derived for wet wastes by HC. This is beneficial to us on choosing the application approaches of hydrochar. However, some other attractive information such as the carbon recovery, energy recovery and the atoms ratios (H/C, O/C and N/C) of hydrochar are not predicted yet in this research. The carbon and energy recovery can provide the energy and resource recovery of waste treatment using HC. Besides, the atoms ratios of hydrochar can exhibit its similarity compared with coal, and they can indicate the stability of carbon in the hydrochar which is a significant index for assessing its ability of carbon capture and storage when used as soil remediation. Another limitation of this RF is that the accuracy of yield prediction is not excellent, and the discussion of target to target is not sufficient for the multi-task learning work. Therefore, the future work will mainly focus on much more targets predictions (carbon recovery, energy recovery and the atoms ratios). Moreover, the other machine learning models, such as deep neural network (DNN) and support vector regression (SVR), will be established to obtain higher accuracy of multi-task predictions for all the targets.

4. CONCLUSIONS

The yield, HHV, and C_char of hydrochar were predicted using RF based on the experimental research data of previous work. The optimized RF with 80 trees and 10 maximum depths has good multi-task prediction capability for fuel characteristics of hydrochar. The R^2 for the perditions of the yield, HHV, and C_char are 0.80, 0.91 and 0.95, respectively. In addition, the yield of hydrochar is mainly related to the HC condition, especially the temperature and water content, while the most important influencing factors for the HHV and C_char are the carbon and ash content of the feedstock. In future work, the accuracy of yield, the correlation among outputs, the prediction of other useful information about hydrochar and HC will be further studied to establish a comprehensive model and predict more targets with high accuracy.

ACKNOWLEDGEMENT

This work was supported by the National Research Foundation, Prime Minister's Office, Singapore under its Campus for Research Excellence and Technological Enterprise (CREATE) programme and the Singapore RIE2020 Advanced Manufacturing and Engineering (AME) Programmatic grant "Accelerated Materials Development for Manufacturing".

REFERENCE

[1] Malinauskaite J, Jouhara H, Czajczyńska D, Stanchev P, Katsou E, Rostkowski P, Thorne RJ, Colón J, Ponsá S, Al-Mansour F, Anguilano L, Krzyżyńska R, López IC, A.Vlasopoulos, Spencer N. Municipal solid waste management and waste-to-energy in the context of a circular economy and energy recycling in Europe, Energy 2017; 141:2013-2044.

[2] Porteous A. Why energy from waste incineration is an essential component of environmentally responsible waste management, Waste Management 2005; 25:451-459.

[3] Li C, Wang X, Zhang G, Li J, Li Z, Yu G, Wang Y. A process combining hydrothermal pretreatment, anaerobic digestion and pyrolysis for sewage sludge dewatering and co-production of biogas and biochar: Pilot-scale verification, Bioresource Technology 2018; 254:187-193.

[4] Berge ND, Ro KS, Mao J, Flora JRV, Chappell MA, Bae S. Hydrothermal carbonization of municipal waste streams, Environmental Science and Technology 2011; 45:5696-5703.

[5] Kambo HS, Dutta A. A comparative review of biochar and hydrochar in terms of production, physico-chemical properties and applications, Renewable and Sustainable Energy Reviews 2015; 45:359-378.

[6] Abel S, Peters A, Trinks S, Schonsky H, Facklam M, Wessolek G. Impact of biochar and hydrochar addition on water retention and water repellency of sandy soil, Geoderma 2013; 202-203:183-191.

[7] Cao X, Ro KS, Chappell M, Li Y, Mao J. Chemical structures of swine-manure chars produced under different carbonization conditions investigated by advanced solid-state 13C nuclear magnetic resonance (NMR) spectroscopy, Energy and Fuels 2011; 25:388-397.

[8] Cao X, Ro KS, Libra JA, Kammann CI, Lima I, Berge N, Li A, Li Y, Chen N, Yang J, Deng B, Mao J. Effects of biomass types and carbonization conditions on the chemical characteristics of hydrochars, Journal of Agricultural and Food Chemistry 2013; 61:9401-9411.

[9] Berge ND, Li L, Flora JRV, Ro KS. Assessing the environmental impact of energy production from hydrochar generated via hydrothermal carbonization of food wastes, Waste Management 2015; 43:203-217.

[10] Cao H, Xin Y, Yuan Q. Prediction of biochar yield from cattle manure pyrolysis via least squares support vector machine intelligent approach, Bioresource Technology 2016; 202:158-164.

[11] Wu G, Kechavarzi C, Li X, Wu S, Pollard SJT, Sui H, Coulon F. Machine learning models for predicting

PAHs bioavailability in compost amended soils, Chemical Engineering Journal 2013; 223:747-754.
[12] Abbasi M, El Hanandeh A. Forecasting municipal solid waste generation using artificial intelligence modelling approaches, Waste Management 2016; 56:13-22.
[13] Xiao Q, Chang HH, Geng G, Liu Y. An Ensemble Machine-Learning Model To Predict Historical PM2.5 Concentrations in China from Satellite Data, Environ Sci Technol 2018; 52:13260-13269.
[14] Li L, Flora JRV, Caicedo JM, Berge ND. Investigating the role of feedstock properties and process conditions on products formed during the hydrothermal carbonization of organics using regression techniques, Bioresource Technology 2015; 187:263-274.
[15] Ro KS, Flora JRV, Bae S, Libra JA, Berge ND, Álvarez-Murillo A, Li L. Properties of Animal-Manure-Based Hydrochars and Predictions Using Published Models, ACS Sustainable Chemistry & Engineering 2017; 5:7317-7324.
[16] Zhu X, Li Y, Wang X. Machine learning prediction of biochar yield and carbon contents in biochar based on biomass characteristics and pyrolysis conditions, Bioresource Technology 2019; 288:121527.
[17] Mu Y, Liu X, Wang L. A Pearson's correlation coefficient based decision tree and its parallel implementation, Information Sciences 2018; 435:40-58.
[18] Refaeilzadeh P, Tang L, Liu H, Cross-validation, in: Encyclopedia of database systems, Springer, 2009, pp. 532-538.
[19] Were K, Bui DT, Dick ØB, Singh BR. A comparative assessment of support vector regression, artificial neural networks, and random forests for predicting and mapping soil organic carbon stocks across an Afromontane landscape, Ecological Indicators 2015; 52:394-403.
[20] Raja K, Patrick M, Elder JT, Tsoi LC. Machine learning workflow to enhance predictions of adverse drug reactions (ADRs) through drug-gene interactions: Application to drugs for cutaneous diseases, Scientific reports 2017; 7:3690.
[21] Cao DS, Liang YZ, Xu QS, Zhang LX, Hu QN, Li HD. Feature importance sampling-based adaptive random forest as a useful tool to screen underlying lead compounds, Journal of Chemometrics 2011; 25:201-207.
[22] Parshetti GK, Chowdhury S, Balasubramanian R. Hydrothermal conversion of urban food waste to chars for removal of textile dyes from contaminated waters, Bioresource Technology 2014; 161:310-319.
[23] Levatić J, Ceci M, Kocev D, Džeroski S, Semi-supervised learning for multi-target regression, in: Lecture Notes in Artificial Intelligence (Subseries of Lecture Notes in Computer Science), 2015, pp. 3-18.
[24] Li L, Wang Y, Xu J, Flora JRV, Hoque S, Berge ND. Quantifying the sensitivity of feedstock properties and process conditions on hydrochar yield, carbon content, and energy content, Bioresource Technology 2018; 262:284-293.
[25] Buss W, Jansson S, Wurzer C, Mašek O. Synergies between BECCS and Biochar - Maximizing Carbon Sequestration Potential by Recycling Wood Ash, ACS Sustainable Chemistry and Engineering 2019; 7:4204-4209.