



Review

Machine learning for hydrothermal treatment of biomass: A review



Weijin Zhang^{a,1}, Qingyue Chen^{b,1}, Jiefeng Chen^a, Donghai Xu^c, Hao Zhan^a, Haoyi Peng^a, Jian Pan^b, Mikhail Vlaskin^d, Lijian Leng^{a*}, Hailong Li^a

^a School of Energy Science and Engineering, Central South University, Changsha, Hunan 410083, China

^b School of Minerals Processing and Bioengineering, Central South University, Changsha, Hunan 410083, China

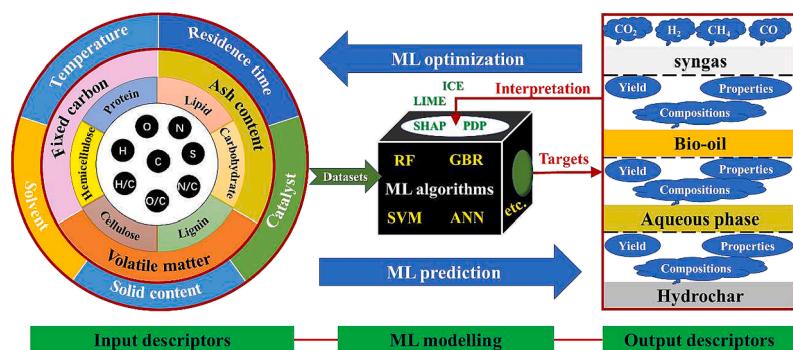
^c Key Laboratory of Thermo-Fluid Science & Engineering, Ministry of Education, School of Energy and Power Engineering, Xi'an Jiaotong University, Xi'an, Shaanxi Province 710049, China

^d Joint Institute for High Temperatures of the Russian Academy of Sciences, Moscow 125412, Russia

HIGHLIGHTS

- Biomass compositions and hydrothermal treatment (HTT) parameters predict products.
- Machine learning (ML) models and procedure were introduced and compared.
- ML can predict yield, compositions, and properties of products from HTT accurately.
- ML-aided prediction, optimization, and application are promising in HTT field.

GRAPHICAL ABSTRACT



ARTICLE INFO

Keywords:

Machine learning
Hydrothermal carbonization
Hydrothermal liquefaction
Hydrothermal gasification
Biomass

ABSTRACT

Hydrothermal treatment (HTT) (i.e., hydrothermal carbonization, liquefaction, and gasification) is a promising technology for biomass valorization. However, diverse variables, including biomass compositions and hydrothermal processes parameters, have impeded in-depth mechanistic understanding on the reaction and engineering in HTT. Recently, machine learning (ML) has been widely employed to predict and optimize the production of biofuels, chemicals, and materials from HTT by feeding experimental data. This review comprehensively analyzed the application of ML for HTT of biomass and systematically illustrated basic ML procedure and descriptors for inputs and outputs of ML models (e.g., biomass compositions, operation conditions, yield and physicochemical properties of derived products) that could be applied in HTT. Moreover, this review summarized ML-aided HTT prediction of yield, compositions, and physicochemical properties of HTT hydrochar or biochar, bio-oil, syngas, and aqueous phase. Ultimately, future prospects were proposed to enhance predictive performance, mechanistic interpretation, process optimization, data sharing, and model application during ML-aided HTT.

* Corresponding author.

E-mail addresses: l.leng2019@csu.edu.cn, lljchs@126.com (L. Leng).

¹ These authors contributed equally to this work.

1. Introduction

Burning of fossil fuels has brought out many problems, such as global climate change, energy crisis, and environmental pollution (Seo et al., 2022). It is necessary to develop renewable energy utilization technologies. Biomass energy (bioenergy) is considered as the only carbon-based renewable energy, which undoubtedly plays a critical role in replacing petroleum (Xiu and Shahbazi, 2012). Hydrothermal treatment (HTT), including hydrothermal carbonization, liquefaction, and gasification, is a kind of emerging thermochemical process for disposing and valorizing wet biomass (without prior drying), such as lignocellulosic biomass (e.g., agricultural waste and wood) (Velvizhi et al., 2022), animal manure, food waste, sewage sludge, and algae (Becker et al., 2014; Leng et al., 2021a; Peterson et al., 2008). The numerical and category descriptors for biomass are shown in Table 1. During the HTT, biomass raw materials are distributed into four phases, namely hydrochar or biochar, bio-oil, syngas, and aqueous phase. Under 180–250 °C, the solid hydrochar is the main product, which can be used as solid fuels and carbon materials, is the primary product of HTT. With higher temperatures, valuable oil and gas products that can be considered biofuels and biochemicals to replace fossil resources are generated gradually, and supercritical water conditions favor syngas production (Ragauskas et al., 2014; Tuck et al., 2012). Still, limited solid residue would be formed within higher temperature. On the other hand, the aqueous phase is produced inevitably in large amounts irrespective of processing temperatures (Leng et al., 2020c; Leng and Zhou, 2018).

The HTT process is a complex reaction system covering various chemical reactions (Marzbali et al., 2021). Many factors influence HTT performance, and it is challenging to investigate the HTT reaction mechanisms and transformation behavior through traditional experimental studies. Applying various modeling methods to simulate the HTT of biomass is preferable. The reactor and kinetic modeling for waste biomass thermochemical conversion have been systematically discussed (Perera et al., 2021). Physical and mathematical models for biomass HTT process systems could be converted into numerical models. The results are often preferable to understand the process of biomass to products. However, these models are mainly based on kinetic equations and transport phenomena (mass, energy, and momentum), which can only be obtained from vast experiments. Thus, reactor and kinetic modeling are time-consuming and labor-intensive. Moreover, kinetic models are generally applied for biomass model compounds (e.g., carbohydrate, protein, lipid, cellulose, hemicellulose, and lignin) modeling but are hard to be generalized to real-world biomass (compositions too complex) (Liu et al., 2022). In summary, the quantitative assessment of the impact and process optimization with traditional experimental or simulation investigations are arduous (Li et al., 2022a), due to the research objects are often limited to one or a few raw materials types (specific biomass or model compound).

Artificial intelligence (AI) has quickly been developed with the advancement of computer science and big data technology (Jordan and Mitchell, 2015). Machine learning (ML), an essential part of AI, is popularly applied in a wide range of areas, such as data mining, computer vision, pattern recognition, and natural language processing (LeCun et al., 2015). Additionally, ML is adept at solving complex nonlinear classification or regression problems in biomass treatment domains because of its strong data fitting ability (Elmaz et al., 2020a; Li et al., 2022a; Wang et al., 2021). Recently, ML has also been applied to aid the HTT of biomass for prediction and optimization. Few studies reported using ML to study HTT of biomass before 2019. Till now, extensive studies have been performed pertaining to ML-aided prediction of the quantity (yield) and quality (compositions and properties) of different products from HTT of biomass. More recently, reviews regarding ML-aided thermochemical treatment of biomass have also been reported just recently. For example, the applications of ML in the thermochemical conversion of biomass were briefly mentioned (khan et al., 2023). Besides, there are some studies overviewed the use of ML

for gasification (Ascher et al., 2022a); gasification and pyrolysis (Ascher et al., 2022b); and anaerobic digestion, hydrothermal carbonization (HTC), and hydrothermal gasification (HTG) (Li et al., 2022b). These reviews were generally organized by 1) introducing the various ML algorithms used in this field; 2) summarizing the prediction performance and discuss the model results; and 3) putting forward prospects, while HHT was mentioned with only a few words. Products poly-generation from HTT was not considered by any of those, and hydrothermal

Table 1
The input and output descriptors for hydrothermal treatment of biomass.

Objects	Quantitative simulation	Experimental descriptors
Biomass	Biomass type	Elemental compositions Proximate compositions Biochemical compositions Ash compositions
Hydrothermal treatment process	Operating conditions	Temperature Pressure Residence time Solid content Heating rate Catalyst, etc.
	Solvent	Reaction solvent Product extraction solvent
	Operating steps	Product extraction procedure Reaction mode (batch/continuous)
	Atmosphere	Inert gases (N ₂ , He, etc.) Catalytic hydrogenation, etc.
Hydrochar (solid phase)	Quantity Compositions	Yield Elemental compositions Chemical compositions Proximate compositions Water and metal content, etc. Structural compositions (functional groups) Ash content and composition Calorific value Density and porosity Energy and carbon recovery
	Fuel properties	Element atom ratios (H/C, O/C, and N/C) Material properties
		Surface area Pore volume and pore size distribution Hydrophobicity and pH Mechanical stability and grindability Ion exchange property Thermal conductivity and heat capacity Yield Elemental compositions Chemical components Calorific value Density and viscosity Ignition point Boiling point Energy recovery Element atom ratios (C/H and C/O)
Bio-oil (oil phase)	Quantity Compositions	Yield Elemental compositions Chemical components Calorific value Density and viscosity Ignition point Boiling point Energy recovery Element atom ratios (C/H and C/O)
	Fuel properties	Yield Gas compositions (CO ₂ , H ₂ , CO, CH ₄ , C ₂ H _n , etc.) Calorific value Tar content
Syngas (gas phase)	Quantity Compositions	Yield Gas compositions (CO ₂ , H ₂ , CO, CH ₄ , C ₂ H _n , etc.) Calorific value Tar content
Wastewater (aqueous phase)	Quantity Compositions	Yield Total organic carbon Total phosphorus Total nitrogen Metal content Glucose Sugar, etc. pH Chemical oxygen demand, etc.
	Properties	

liquefaction (HTL) was not addressed. In addition, the descriptors of inputs and outputs for HTT process were not stated comprehensively, and aqueous phase products were ignored. In particular, the ML-aided HTT of biomass by model interpretation and optimization has not been appropriately elaborated.

This review aims to summarize up-to-date development in ML-aided HTT of biomass and propose future directions. The review covers brief introduction to basic ML procedure and classic algorithms (Section 2), analysis and discussion of recent advance in ML-aided predictions and applications for multiple products from HTC, HTL, and HTG (Sections 3, 4, and 5), research gap and future prospects (Section 6), and concluding remarks (Section 7).

2. ML procedure and algorithms

2.1. ML procedure

Basic ML procedure includes data collection, data preprocessing, model development, evaluation and interpretation, and model sharing (Fig 0.1). Data collection is the initial core step of ML, and data are generally collected from reliable literature with data extraction softwares, such as PlotDigitizer (Zhang et al., 2021) and Getdata (Tang et al., 2021). The descriptions of biomass, HTT process, and products are shown in Table 1. Data include not only input features but also output variables. Feature quantification is to describe fuzzy variables with numerical data. When the number of input features is much larger than the number of output variables, dimensionality reduction methods (i.e., principal component analysis, discriminant analysis, and independent component analysis) are possibly required for integrating features. The statistical analysis indicates data distribution for all features and can further test outliers. If the range of values for different variables varies greatly, data normalization is conducted to transfer these data into same magnitude within -1 and 1. After data collection and preprocessing, suitable ML algorithms should be chosen based on various datasets. To balance the accuracy and generalization of models, 70–90 % data among whole data are generally randomly selected to train ML models, and the

remaining data are used as testing data, which are vital for evaluating models. ML models have two kinds of parameters, called model parameters and hyperparameters. Model parameters are configuration variables inside the model (e.g., the weights coefficients of neurons in neural networks models), whose values can be estimated based on data. Hyperparameters are configurations outside the model (e.g., the depth of decision trees in tree-based models), whose values cannot be directly determined from the data training. Hyperparameters are generally applied to configure a ML model or to minimize the loss function of the specific algorithm, and must be adjusted before learning a ML model because they define the model frame. The optimal ML models should be developed to fit the data with best hyperparameters. Hyperparameters of ML training models must be tuned, and k-fold cross-validation is an efficient method for determining optimal hyperparameters (Leng et al., 2022d). The optimal average performance of the k-fold cross-validation corresponds to optimal hyperparameters. The final ML models are re-trained with the best hyperparameters using the training data and assessed with the testing data. During the hyper-parameters tuning process, traversing methods (Li et al., 2021c; Zhang et al., 2021) and various optimization algorithms (Leng et al., 2022d) can be used for fitting ML models.

In general, the coefficient of determination (R^2) and root mean squared error (RMSE) are used for evaluating regression models. And receiver operating characteristic (ROC) curve and area under curve (AUC) value are commonly applied for classification models. A ML model with high R^2 (near 1) and low RMSE values is generally accurate to fit data. The R^2 of ML models used in HTT are mostly higher than 0.85, and some can reach as high as 0.99 (Tables 2–5). However, relying on these statistics indicators alone cannot explain the effects and mechanisms of the input features on the outputs. Feature importance analysis and partial dependence plots (PDP) (Goldstein et al., 2015) are thus conducted to study the correlations between inputs and outputs. The reasonability of the model interpretation may not necessarily be favorable, although with acceptable predictive performance. ML-based optimizations for desired outputs and inputs are practical to obtain optimal solutions. Experimental schemes can be designed with these solutions to

Table 2

Machine learning aided prediction for yield, compositions, and properties of hydrochar from hydrothermal carbonization.

Biomass Types	Input descriptors ^a	Data points	Output ^b	Models and predictive performance			Reference
				Algorithms ^c	Test R^2	Test RMSE ^c	
Wet wastes	PC, EC, OC	248	Yield	SVR	0.88	7.83	(Li et al., 2020a)
	PC, EC, OC		ER		0.92	6.22	
	PC, EC, OC		ED		0.89	0.09	
Biomass	PC, EC, BC, OC	622	C	RF	0.95	2.40	(Li et al., 2018)
	PC, EC, AR, OC		Yield	DNN	0.90	7.05	
Biomass	EC, solvent, OC, catalyst	488	ER		0.88	7.60	(Castro Garcia et al., 2022) (Shafizadeh et al., 2022)
	EC, OC		C		0.95	2.91	
			H/C		0.89	0.08	
			O/C		0.91	0.06	
			N/C		0.89	0.01	
			Yield	RF	0.94	3.84	
Woody and herbaceous Biomass	PC, EC, OC	296	Yield	GPR	0.94	— ^d	(Mu et al., 2022)
			Yield	NNR	0.80	—	
				GAM	0.85	—	
				SVR	0.89	—	
				PSO-NN	0.86	5.46	
			HHV		0.90	1.15	
Sewage sludge	PC, EC, OC	138	DHD		0.85	0.04	(Djandja et al., 2021) (Djandja et al., 2022)
			DCD		0.85	0.08	
			Ash		0.98	1.10	
			N/C		0.96	0.37	
	PC, EC, OC		TP	ANN	0.97	—	
	PC, EC, OC			RF	0.85	5.71	

^a : EC: elemental compositions of biomass; AR: atomic ratios of biomass; BC: biological compositions of biomass; PC: proximate compositions of biomass; OC: operating conditions of hydrothermal process;

^b : N: nitrogen content of hydrochar; TP: total phosphorus of hydrochar.

^c : SVR: support vector machine; RF: random forest; DNN: deep neural network; GPR: gaussian process regression; NNR: GAM: generalized additive model; PSO: particle swarm optimization; ANN: artificial neural network.

Table 3

Machine learning aided prediction for yield, compositions, and properties of bio-oil from hydrothermal liquefaction.

Biomass Types	Input descriptors ^a	Data points	Output ^b	Models and predictive performance			Reference
				Algorithms ^c	Test R ²	Test RMSE	
Algae	EC, OC	310	Yield	GBR	0.86	5.52	(Zhang et al., 2021)
	AR, OC			RF	0.85	6.16	
	BC, OC	310	Yield	GBR	0.88	5.07	
	EC, AR, BC, OC			RF	0.85	5.08	
		310	Yield	GBR	0.90	4.75	
				RF	0.86	6.06	
				GBR	0.90	4.69	
				RF	0.87	5.72	
		N	Yield	GBR	0.90	1.68	
				RF	0.87	1.91	
		O	Yield	GBR	0.90	0.68	
				RF	0.88	0.74	
Biomass	EC, BC, AR, OC	325	Yield	XGB	0.90	4.77	(Katongtung et al., 2022)
			HHV		0.87	1.47	
Biomass	EC, solvent, OC, catalyst	488	Yield	RF	0.90	6.03	(Castro Garcia et al., 2022)
Organic wastes	BC, OC	525	Yield	RF	0.78	8.07	(Cheng et al., 2022)
Biomass	EC, AR, OC	108	Yield	RF	0.85	5.83	(Leng et al., 2022c)
Biomass	EC, AR, BC, OC	91		RF	0.92	3.31	
Biomass	EC, OC, solvent	448	Yield	DRT	0.82	8.07	(Li et al., 2021c)
				RF	0.89	6.17	
				GBR	0.90	6.14	
	BC, OC	229	N	RF	0.75	0.95	
			ER	RF	0.80	8.84	
Biomass	EC, OC	650	Yield	GPR	0.95	0.04 ^d	(Shafizadeh et al., 2022)
		268	C		0.91	0.03 ^d	
			H		0.98	0.03 ^d	
			O		0.92	0.04 ^d	
		263	N		0.99	0.02 ^d	
			S		0.99	0.02 ^d	
			HHV		0.97	0.01 ^d	
Pine	EC, OC	–	HHV	GBR	0.98	52.15	(Zhou et al., 2022)
				RR	0.52	225.45	

^a : EC: elemental compositions of biomass; AR: atomic ratios of biomass; BC: biological compositions of biomass; PC: proximate compositions of biomass; OC: operating conditions of hydrothermal process.

^b : HHV: higher heating value; C, H, O, N, and S represents carbon, hydrogen, oxygen, nitrogen, and sulfur content of bio-oil, respectively.

^c : GBR: gradient boosting regression; RF: random forest; XGB: extreme gradient boosting; RR: ridge regression DRT: decision regression tree; GPR: gaussian process regression.

^d : Calculated with normalized data.

verify models' reliability further.

Model sharing provides convenient communication platforms for researchers. Both online platforms based on websites (Leng et al., 2022d) and offline applications (Shafizadeh et al., 2022) are available to develop graphical user interfaces (GUIs) for the final optimal ML models. With these GUIs, output values can be obtained quickly and accurately by inputting values of independent variables. The above ML modeling, optimization, and sharing can be realized in some open-access coding platforms, such as C/C++ language (Serrano et al., 2017), Python (Castro Garcia et al., 2022), R language (Natekin and Knoll, 2013), Matlab (Qi et al., 2021), and Julia language (Gao et al., 2020), etc.

2.2. ML algorithms

ML algorithms can be classified into three types: supervised, semi-supervised, and unsupervised algorithms (Fig. 2b). The classic and common algorithms, such as random forest (RF), gradient boosting decision tree (GBDT), support vector machine (SVM), and neural network (NN), can be used in different fields. More details about the principle and implementation of ML algorithms can be referred to (Ouadah et al., 2022; Yan et al., 2021). Many ML libraries and frameworks are available for carrying out ML algorithms, e.g., Scikit-learn (the most popular one), TensorFlow, PyTorch, and Keras. Each ML algorithm has its hyper-parameters, which must be tuned to fit specific datasets into different problems. Hyper-parameters adjusting can be realized manually (i.e., grid search or random search) or by an automatic process using optimization algorithms. Due to the high computational cost of manually

optimizing hyper-parameters, it is preferable to use optimization algorithms for finding the best hyper-parameter configurations. The theory and practice of hyper-parameter optimization libraries and frameworks for developing ML models were discussed in ref. (Yang and Shami, 2020). Particle swarm optimization (PSO) (Juneja and Nagar, 2016), genetic algorithm (GA) (Oliveira et al., 2010), simulated annealing algorithm (Lin and Chen, 2014), and Bayesian optimization (BO) are recommended for hyper-parameter optimization. In general, NN algorithms are more complex than traditional ML algorithms (tree-based algorithms), but they can probably achieve high prediction accuracy.

First, it is necessary to determine whether it is a regression problem or a classification problem according to the different problems in HTT field. For example, the regression models are often used to conduct the hydrothermal products prediction (yield, composition, and properties) and optimization, the classification models are employed for distinguishing different biomass or hydrothermal products, and similar properties of biomass or products could be labeled by the clustering models (Fig. 2). Then, the suitable models could be selected based on different data types (i.e., data numbers, features numbers, character types, digital data, and image data). Traditional ML models includes logistic regression, naive Bayes, decision trees, SVM, RF, GBDT and so on, which are sensitive to digital data. The NN models are preferable to read and treat image data (usually contains amount of data and features) than traditional models (Zhu et al., 2022). The advantages and disadvantages of using common models in the prediction and optimization can be seen in supplementary material.

Table 4

Machine learning aided prediction for yield and compositions of syngas from hydrothermal gasification.

Biomass Types	Input descriptors ^a	Data points	Output	Models and predictive performance			Reference
				Algorithms ^b	Test R ²	Test RMSE	
Biomass Wet wastes	EC, OC	216	Yield	GPR	0.86	0.10 ^c	(Shafizadeh et al., 2022)
	EC, OC	295	CO	NN	0.95	0.31	
			CO ₂		0.97	1.06	
			CH ₄		0.94	0.45	
			H ₂		0.97	1.27	
	EC, OC, alkali catalyst	117	CO	NN	0.86	0.40	
			CO ₂		0.82	2.22	
			CH ₄		0.80	0.74	
			H ₂		0.91	2.08	
	EC, OC, transition-metal catalyst	73	CO	NN	0.87	0.51	
Biomass			CO ₂		0.98	0.96	
			CH ₄		0.92	1.58	
			H ₂		0.93	1.60	
	EC, OC, with or without catalyst	233	CO	NN	0.80	0.26	
			CO ₂		0.91	2.38	
			CH ₄		0.86	1.22	
			H ₂		0.95	2.77	
	EC, OC	95	H ₂	GPR	0.95	2.93	(Zhao et al., 2021)
			H ₂	ANN	0.92	1.85	
Food wastes	OC	40	H ₂	SVM	0.98	0.87	(Shenbagaraj et al., 2021)
			H ₂	RF	0.98	1.02	
			CO	NN	0.99	0.15	
			CO ₂		0.98	0.52	
Biomass	PA, EC, OC	125	CH ₄		0.99	0.32	(Haq et al., 2022)
			H ₂		0.99	0.46	
			H ₂	GPR	0.96	0.20	
			H ₂	DT	0.92	1.23	
Biomass			H ₂	SVM	0.69	2.97	
			H ₂	ANN	0.86	1.99	

^a : EC: elemental compositions of biomass; AR: atomic ratios of biomass; BC: biological compositions of biomass; PC: proximate compositions of biomass; OC: operating conditions of hydrothermal process.

^b : GPR: gaussian process regression; NN: neural network regression; ANN: artificial neural network; SVM: support vector machine; RF: random forest; DT: decision tree.

^c : Calculated with normalized data.

Table 5

Machine learning aided prediction for yield, properties and compositions of aqueous phase from hydrothermal treatment.

Biomass Types	Input descriptors ^a	Data points	Output ^b	Models and predictive performance			Reference
				Algorithms ^c	Test R ²	Test RMSE	
Biomass	EC, OC	197	Yield	GPR	0.87	0.08 ^d	(Shafizadeh et al., 2022)
				NN	0.80	— ^e	
				GAM	0.98	—	
				SVR	0.76	—	
Biomass	EC, BC, OC	257	pH	GBDT	0.93	0.36	(Leng et al., 2022d)
				RF	0.87	0.49	
	EC, BC, OC	224	TN	GBDT	0.96	1.25	
				RF	0.95	1.45	
	EC, BC, OC	174	TOC	GBDT	0.91	9.06	
				RF	0.86	14.87	
	EC, BC, OC	136	TP	GBDT	0.90	0.11	
				RF	0.86	0.13	
	EC, BC, OC	110	pH, TN, TP	GBDT	0.85	0.55	
				RF	0.82	0.64	
Microalgae and residue	PC, EC, OC, H ₂ SO ₄	49	Glucose	MARS	0.93	—	(Wei-Hsin Chen et al., 2022)
			Glucose	NN	0.99	—	
Mesocarp Fiber	OC, Catalyst	45	Sugar	SVR	0.79	8.57	(Lee et al., 2022)

^a : EC: elemental compositions of biomass; BC: biological compositions of biomass; PC: proximate compositions of biomass; OC: operating conditions of hydrothermal process.

^b : TN: total nitrogen; TOC: total organic carbon; TP: total phosphorus.

^c : GPR: gaussian process regression; NN: neural network regression; GAM: generalized additive model; SVR: support vector machine; GBDT: gradient boosting decision tree; RF: random forest; MARS: multivariate adaptive regression splines.

^d : Calculated with normalized data.

^e : Not available.

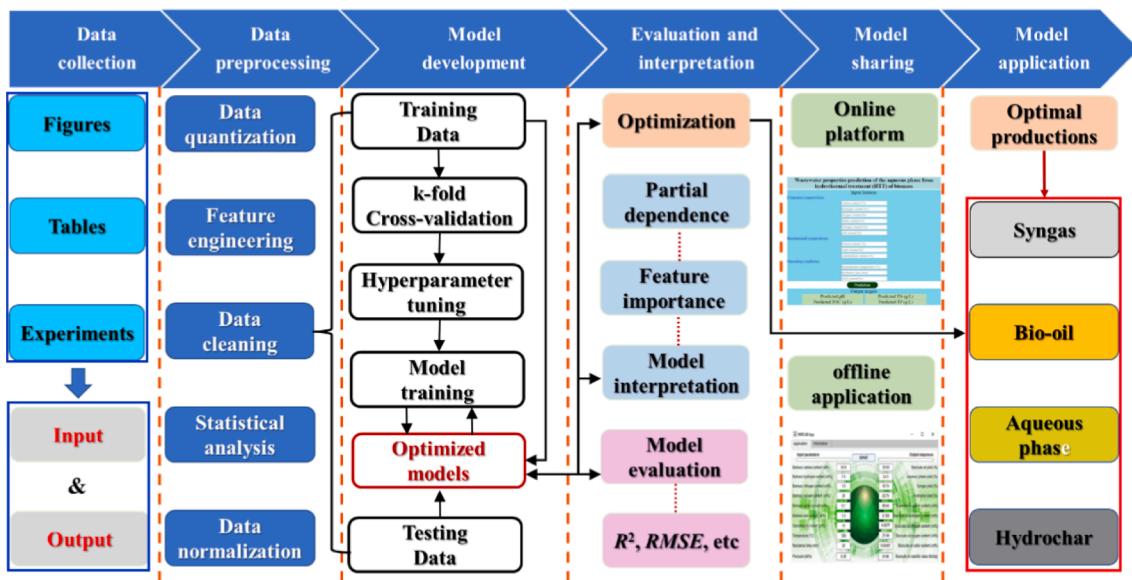


Fig. 1. The basic procedure of machine learning applications. Figures of the online platform and offline application cited from ref. (Leng et al., 2022d) and (Shafizadeh et al., 2022), respectively. Copyright Elsevier 2022.

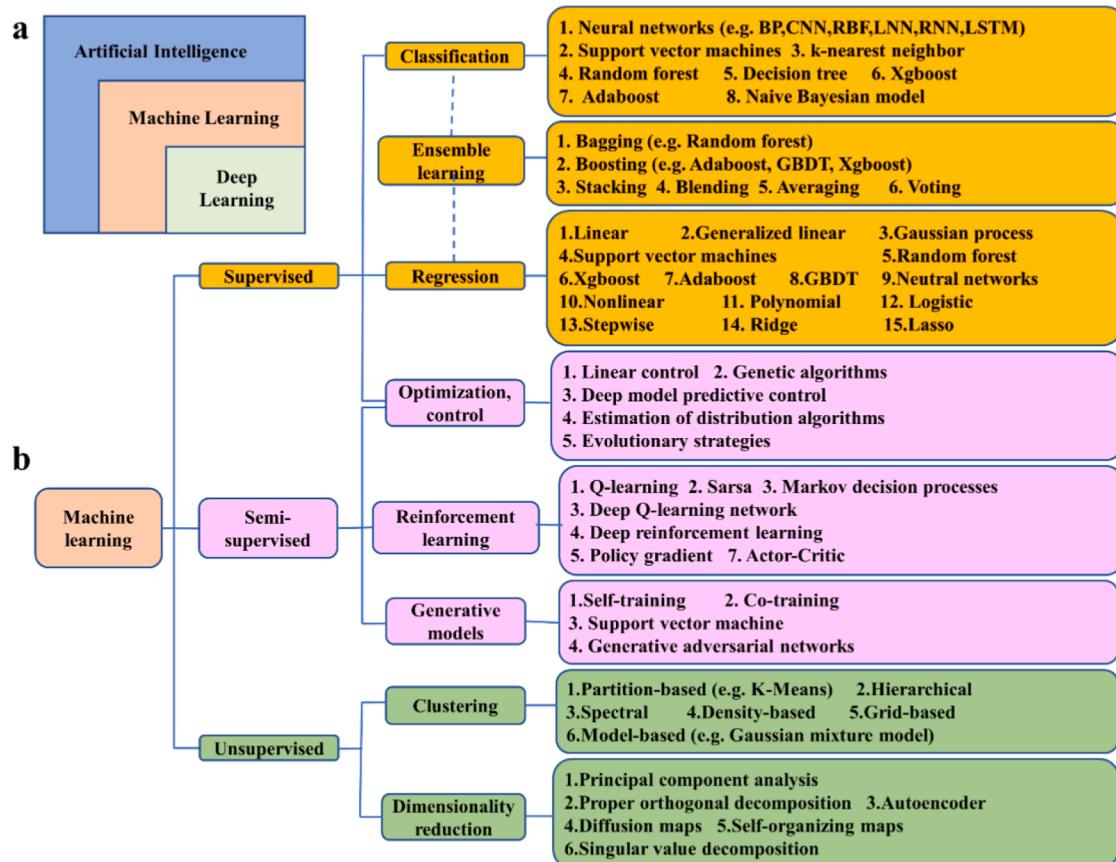


Fig. 2. (a) The relationships among artificial intelligence (AI), machine learning (ML), and deep learning (DL); (b) The possible categorization of ML algorithms. Figure adapted from ref. (Zhu et al., 2022).

3. Hydrothermal carbonization

3.1. Hydrothermal carbonization process

HTC, or so-called wet torrefaction, works at low temperatures

(180–260 °C) and low pressures (2–5 MPa), and the main product of the HTC process is a solid phase product called hydrochar (Peng et al., 2017; Leng et al., 2021e; Zhai et al., 2017). Limited gas (the main compound is CO₂) and a massive amount of aqueous phase will be generated. The principle reaction pathways in HTC for the formation of hydrochar have

been investigated (Kruse et al., 2013; Leng et al., 2021d). The formation of hydrochar can have mainly-two different pathways to form primary char via pyrolysis and secondary char via repolymerization. When biomass is heated in subcritical water, cellulose and hemicellulose are hydrolyzed to glucose or other saccharides such as fructose, protein to amino acids, lignin to guaiacol, and lipid is hydrolyzed to fatty acids (Li et al., 2019; Shahbeik et al., 2022). Higher temperatures correspond to higher pressures under an environment where water is the medium, and the most hydrochar with desired energy density was prepared at temperature near 255 °C and pressure near 5 MPa (Hoekman et al., 2011; Mu et al., 2022). High-pressure (20 MPa) water can promote cellulose HTC in low temperatures (117 °C) by cleaving the hydrogen bonding, activating the C—H bonds, and facilitating the dehydration (Yu et al., 2022). The residence time of HTC is generally from one hour to several hours, and even one day, if not more. As the reaction goes on, these hydrolyzed products are gradually decomposed and recombined to form hydrochar through chemical reactions, particularly interactive reactions (Funke and Ziegler, 2010; Li et al., 2019). Among the biochemical compositions of biomass, carbohydrates are the primary source of hydrochar during the HTC process of biomass. Generally, the increase of ash contents could reduce hydrochar yield. However, the release of alkali and alkaline earth metals (e.g., Ca) in ash compositions to the hydrochar surface could obtain the higher yield (He et al., 2022a), because the presence of inherent alkali and alkaline earth metals (especially Ca) in the biomass can catalyze the hydrolysis of biomass to hydrochar (Liu et al., 2021a). And the generation of hydrochar with high aromaticity from organic acids could be restrained in an alkaline environment (Liu et al., 2021c), indicating lower H/C (<1.4) and O/C (<0.4) ratios of the hydrochar (He et al., 2022a). The accumulated Ca in the hydrochar could also decreased combustion properties, due to formed stable bonds between Ca and volatile carbon (He et al., 2022c). As reported in previous literature, a series of reactions occurred (e.g., hydrolysis, dehydration, aromatization, decarboxylation, and polymerization) from carbohydrate to hydrochar (Shi et al., 2019; Zhai et al., 2017). The H/C and O/C of hydrochar are lower than raw biomass, which is close to some kinds of coal (Xu et al., 2021). Thus, HTC is considered a promising technology for fuel production to obtain coal-like char. The other fuel properties, such as calorific value, density, porosity, and energy/carbon recovery (Table 1), are also decisive to hydrochar characteristics for replacing coal (E. Leng et al., 2021). H/C, O/C, and the ratio of fixed carbon/volatiles can be treated as hydrochar carbon stability indexes representing the potential of climate change mitigation of char (Chen et al., 2021; Leng et al., 2019). Besides, hydrochar has been studied for use in many other applications by acting as adsorbents, catalysts, carbon-based materials (Zhu et al., 2021), fertilizers, soil additives, etc. (Mihajlović et al., 2018). These various applications depend on material properties, for example, surface area, pore volume, pore size distribution, pH, cation exchange capacity, thermal conductivity, and so on (Kan et al., 2016). For example, the surface area and pore volume of hydrochar determine the adsorption of pollutants in the wastewater (Sigmund et al., 2020) and the adsorption of CO₂ and H₂ in gases (Maulana Kusdhany and Lyth, 2021). Cation exchange capacity is the most important to remove heavy metals from wastewater (Zhu et al., 2019). The elemental compositions and surface characteristics of hydrochar can be used to predict soil heavy metal immobilization (Palansooriya et al., 2022). By tailoring the hydrothermal conditions, hydrochar could also be used for not only sludge management in waste water treatment plant (e.g., enhance anaerobic digestion performance) but also contaminants control in industrial wastewater treatment (He et al., 2022b). Hence, it is critical to explore the effects of biomass compositions and hydrothermal parameters on these application-performance-decisive properties of hydrochar.

3.2. ML-aided HTC

High yield and preferable compositions and properties of the

hydrochar are the major focuses of HTC's effectiveness (Xiang et al., 2020). The chemical compositions (C, H, O, N, S, H/C, O/C, N/C, etc.), proximate compositions (volatile matter, fixed carbon, ash, and moisture), and functional groups (especially O/N/S-containing functional groups) are usually employed to assess the potential of hydrochar in numerous applications (Leng et al., 2021d; Leng et al., 2022a). RF and NN are the most frequently used ML algorithms for predicting hydrochar with single-target or multi-targets models. Representative examples of ML-aided prediction and optimization for HTC are shown in Fig. 3 and Table 2.

Table 2 reviewed ML-aided prediction for yield, compositions, and properties of hydrochar from HTC. A gaussian process regression (GPR) model was used to predict yield with high performance ($R^2 = 0.94$), considering elemental compositions and HTC operating conditions as inputs (Shafizadeh et al., 2022). The temperature and C content of biomass were the two most important input variables to hydrochar yield (Table 6). A hybrid PSO-NN model was built for yield prediction with proximate analysis, elemental compositions, and operating conditions as inputs (Mu et al., 2022). Herein, the PSO algorithm was coupled with NN models for hyper-parameter adjusting, which helped to reduce the relative error of predictive models. However, the performance of the hybrid model ($R^2 = 0.86$) is inferior to that of the GPR model (Shafizadeh et al., 2022), which may result from the differences in input features and datasets. Among biomass compositions, the pure lignin is necessary to be used as raw material for valorization, which is the dominant aromatic polymer in nature (Ragauskas et al., 2014). However, the reaction behavior of lignin to hydrochar is challenge to understand because of the many possible interactions between operating conditions. Recently, a RF model has been developed to predict the hydrochar yield from HTC of lignin with the addition of solvents and catalysts (Castro Garcia et al., 2022). The lignin/solvent ratio, catalyst zeolite, and catalyst/solvent ratio were considered as descriptors for solvents and catalysts. The prediction accuracy can reach a high level of $R^2 = 0.94$ and RMSE = 3.84. The feature importance showed that temperature was also the most crucial factor, accounting for 26 % of the weight. However, the hydrochar yield was not dependent on the solvents or catalysts compared to other variables (Castro Garcia et al., 2022; Li et al., 2022c). The compositions and properties of hydrochar were predicted and interpreted with ML models in addition to hydrochar yield. ANN and RF models were applied to predict the N content and total phosphorus (TP) of hydrochar with favorable performance, respectively (Djandja et al., 2022, 2021). In addition, various ML models, including decision tree (DT), RF, SVM, radial basis function (RBF), adaptive network-based fuzzy inference system (ANFIS), and multilayer perceptron (MLP), were developed and compared for predicting specific surface area (SSA) (Leng et al., 2022b), higher heating value (HHV) (Chen et al., 2022b), and energy recovery of hydrochar (Rasam et al., 2021). The predicted values of these three targets from best SVM models ($R^2 \sim 0.97$) were consistent with the experimental values (relative error < 3 %), showing strong prediction and generalization ability of ML models in the HTC process. Furthermore, the particle size of biomass was the most influential parameter for hydrochar SSA compared to temperature and residence time. The reduction of particle size brought about more water permeation into pores of the biomass and thereby, a bigger hydrochar SSA would be obtained under the enhancement of the heat and mass transfer (Rasam et al., 2021). Long residence time and moderation temperature could improve HHV and energy recovery of hydrochar due to biomass oxygen removal or hydrolysis of hemicellulose (Li et al., 2023; Rasam et al., 2021). Though the accuracy of these models for predicting hydrochar yield, compositions, and properties were slight different, the R^2 can be higher than 0.90 in optimal ML models with various datasets. The critical reaction factors (especially temperature and C content) can be considered to improve the accuracy of prediction models for hydrochar properties, which are also important to hydrochar yield.

Different multi-task models (RF, SVR, and DNN) were used to

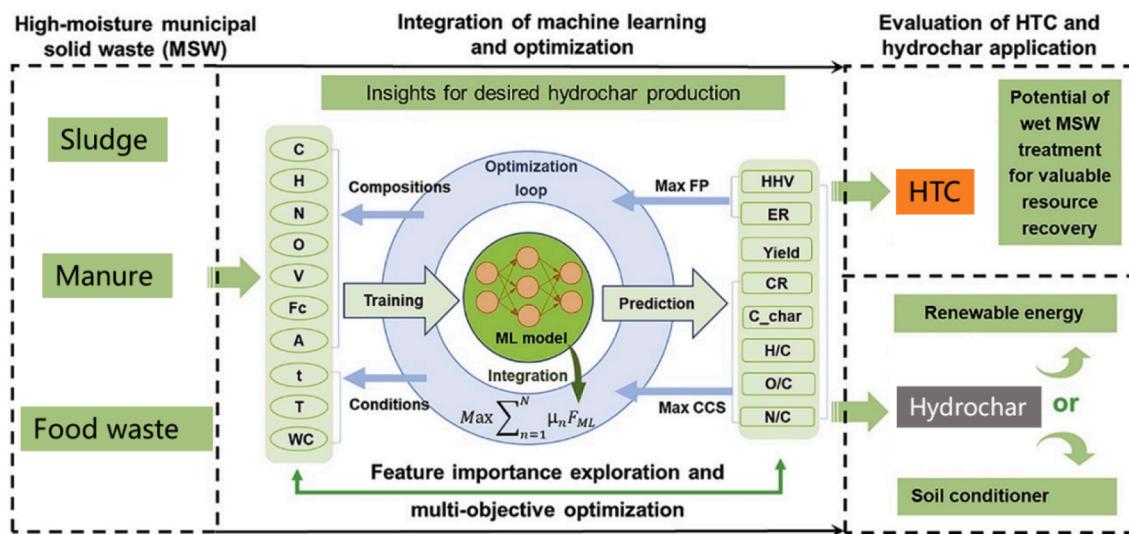


Fig. 3. ML-aided prediction, interpretation, and optimization for hydrochar from hydrothermal carbonization. Figure modified from ref. (Li et al., 2021d). Copyright Elsevier 2021.

simultaneously predict multiple characteristics of hydrochar from HTC of the municipal solid waste, including yield, HHV, carbon content, carbon recovery, energy recovery, N/C, H/C, and O/C (Li et al., 2021d). These results revealed that all three ML models had satisfactory performance (average R^2 above 0.88) for predicting eight targets together, and DNN obtained the highest expression with $R^2 = 0.91$. From feature analysis based on multi-task models, the critical input feature to predict H/C and O/C was temperature, while N content of the biomass was the most significant to N/C prediction (Table 6). Meanwhile, the multi-objective optimization was integrated with ML models to optimize the desired properties of hydrochar and experimental conditions. In another study of the same group, RF and SVM models were developed and compared to predict the fuel properties (yield, energy recovery, and energy densification) of hydrochar and pyrochar. The SVM showed better performance (average R^2 0.88–0.91) than RF (average R^2 0.75–0.79). The N content of biomass had an essential effect on hydrochar (Li et al., 2020a), and the N content differences among sewage sludge, food waste, manure, and other biomass impact hydrochar significantly. For various dataset with different data numbers and features, the performance accuracy difference. The reasons for the difference among the prediction accuracy of these three models may be as follows: First, the dataset is different with various input and output features; Secondly, SVM seems to show better prediction accuracy than DNN in small sample data; Thirdly, the models and algorithms of DNN, SVM, and RF are different, resulting in different prediction accuracy (see supplementary material).

4. Hydrothermal liquefaction

4.1. Hydrothermal liquefaction process

HTL is used to produce biocrude (bio-oil) under moderate operating conditions (240–375 °C and 5–25 MPa). Interactions between subcritical water and biomass in HTL helps to break the chemical bonds of the biomass to obtain bio-oil of high calorific value (>30 MJ/kg) (Leng et al., 2020b). In subcritical water, HTL involves a lot of complex chemical reactions and structural transformations. The major compositions of biomass feedstocks (e.g., protein, lipid, carbohydrate, lignin, cellulose, and hemicellulose) are depolymerized, decomposed, and recombined to produce bio-oil, hydrochar, aqueous phases, and gas during HTL (Brindhadevi et al., 2021). The contents of alkali and alkaline earth oxides, hydroxides, and carbonates among biomass ash components could explain the alkalinity of biomass (Vassilev et al.,

2013). For most cellulosic biomass, the main hydrothermal product was 5-(Hydroxy-methyl)-furfural (HMF) rather than carboxylic acids leading to relatively high bio-oil yields under weak alkaline conditions (Yin et al., 2011). And the bio-oil yield decreased with an increase alkalinity due to the generation of polar carboxylic acids. However, the bio-oil yield from lignin-rich biomass increased with strong alkaline conditions, because high alkalinity can greatly promote the degradation of lignin from biomass (Castro Garcia et al., 2022). Bio-oil comprises organic constituents, such as fatty acids, aromatic hydrocarbons, phenolics, nitrogen-containing heterocycles, amides, and esters (Aierzhati et al., 2019). The extraction solvent and procedure for phase separating bio-oil and aqueous phase could affect bio-oil yield (Leng et al., 2020c, b). The modest increase of temperature (range from 300 °C to 350 °C) and pressure (~10–20 MPa) could obtain maximum bio-oil yield from biomass, and a stable bio-oil yield would produce with further increasing pressure (Mathimani and Mallick, 2019; Mishra et al., 2022). The bio-oil yield is generally lower than 60 % and mostly ranges from 20 % to ~40 % (Leng et al., 2020b). Aqueous phase yield from HTL of biomass, which represents carbon yield in the aqueous phase, is also high (Fan et al., 2020), and values were reported in the range of 7–48 % (Shafizadeh et al., 2022). According to relevant experiment works, the ranges of gas yield from HTL were 0.14 % to 41.10 % (Shafizadeh et al., 2022). The main gas compositions from HTL are CO₂ and CO, while CH₄ and H₂ are the minorities.

4.2. ML-aided HTL

Different ML models have been used to predict the quantity and quality of bio-oil from HTL of biomass. The common targets were yield, elemental compositions (O and N content), and HHV. The most employed ML algorithms were GBR and RF. Fig. 4 shows a typical procedure for ML-aided prediction, mechanism interpretation, and optimization for bio-oil from HTL, and more cases are listed in Table 3.

The bio-oil yield was frequently predicted due to the higher data availability for bio-oil yield than compositions and properties. The predictive performances of yield by various ML models differed sharply, with R^2 ranging from 0.78 to 0.98 (Cheng et al., 2022; Leng et al., 2020b; Zhou et al., 2022). The influence of input features (proximate and ultimate analysis of biomass) on bio-oil yield can be visually presented by PDP based on the optimal RF model (Ullah et al., 2021). Tree-based ML models (i.e., DT, RF, and GBR) were applied to predict bio-oil yield, N content, and energy recovery by considering biomass compositions and HTL conditions as input features (Li et al., 2021c). The RF models

Table 6

The main inputs and outputs of four-aqueous products from hydrothermal treatment.

Outputs ^a	Input features and feature importance ranking															Reference		
	Elemental compositions					Element ratios			Proximate compositions ^c			Biochemical compositions ^d			Hydrothermal conditions ^e			
	C	H	N	O	S	H/C	O/C	N/C	Fc	V	A	P	L	CA	T	RT	SC	
Hydrochar																		
Yield_char	1 ^b	10	8	3	5	—	—	—	4	6	7	—	—	—	2	11	9	(Mu et al., 2022)
N/C_char	2	3	1	5	8	—	—	—	4	6	7	—	—	—	9	11	10	
Ash_char	4	5	6	1	9	—	—	—	3	7	2	—	—	—	8	11	10	
HHV_char	2	4	8	3	10	—	—	—	5	9	7	—	—	—	1	6	11	
DHD_char	3	4	5	8	11	—	—	—	6	9	2	—	—	—	1	7	10	
DCD_char	7	6	3	2	10	—	—	—	9	4	8	—	—	—	1	5	11	
N_char	✓	✓	✓	✓	✓	—	—	—	✓	✓	✓	—	—	—	✓	✓	✓	(Djandja et al., 2021)
P_char	2	10	5	3	—	—	—	—	1	7	4	—	—	—	8	9	6	(Djandja et al., 2022)
H/C_char	2	4	7	3	—	—	—	—	6	8	5	—	—	—	1	9	10	(Li et al., 2021d)
O/C_char	4	3	6	2	—	—	—	—	5	9	7	—	—	—	1	8	10	
C_char	1	3	7	8	—	—	—	—	5	4	2	—	—	—	6	10	9	
ER_char	7	4	1	9	—	—	—	—	5	6	8	—	—	—	2	10	3	
CR_char	6	7	2	10	—	—	—	—	4	5	8	—	—	—	1	9	3	
Bio-oil																		
Yield_oil	✓	✓	✓	✓	✓	—	—	—	2	4	1	7	5	6	3	(Cheng et al., 2022)		
Yield_oil	6	9	10	11	—	5	13	—	—	8	3	4	2	1	12	7	(Katongtung et al., 2022)	
HHV_oil	9	6	8	10	—	5	13	—	—	7	3	2	4	1	12	11		
Yield_oil, N_oil, ER_oil	5	7	9	10	11	—	—	—	—	6	8	1	4	2	3	—	(Li et al., 2021c)	
HHV_oil	✓	✓	✓	✓	✓	—	—	—	—	✓	—	—	—	✓	✓	✓	(Shafizadeh et al., 2022)	
O_oil	10	8	7	6	13	9	3	11	—	5	14	12	4	1	2	15	(Zhang et al., 2021)	
N-H_oil	10	13	5	11	8	9	14	1	—	12	2	3	4	7	6	15	(Leng et al., 2022c)	
Syngas																		
Yield_gas	✓	✓	✓	✓	✓	—	—	—	✓	—	—	—	✓	✓	✓	(Shafizadeh et al., 2022)		
H ₂	3	8	4	7	—	—	—	—	6	—	—	—	1	5	2	(Li et al., 2021b)		
CO ₂	5	4	6	3	—	—	—	—	8	—	—	—	1	7	2			
CH ₄	5	4	6	3	—	—	—	—	8	—	—	—	1	7	2			
CO	5	2	8	3	—	—	—	—	6	—	—	—	4	1	7			
H ₂	7	3	—	5	—	—	—	—	6	—	—	—	2	4	1	(Zhao et al., 2021)		
Aqueous phase																		
Yield_aqueous	✓	✓	✓	✓	✓	—	—	—	✓	—	—	—	✓	✓	✓	(Shafizadeh et al., 2022)		
pH	11	10	2	3	4	—	—	—	7	9	8	5	1	6	12	(Leng et al., 2022d)		
TN	8	11	2	6	10	—	—	—	—	12	1	9	7	4	5	3		
TOC	10	5	4	11	6	—	—	—	—	7	9	2	8	3	1	12		
TP	10	11	5	7	8	—	—	—	—	12	6	3	9	2	4	1		
pH, TN, TP	11	12	1	6	5	—	—	—	—	8	9	10	7	3	4	2		
Glucose	7	4	8	9	—	—	—	—	3	1	2	—	—	5	6	—	(Wei-Hsin Chen et al., 2022)	

Low

Feature importance

High

—: Not consider the input.

✓: Considered as the input but not conduct the rank of feature importance.

^a : HHV: higher heating value; DHD: dehydration degree; DCD: decarboxylation degree; ER: energy recovery; CR: carbon recovery; N—H: nitrogen heterocycles; TN: total nitrogen; TOC: total organic carbon; TP: total ph.^b : Number means the rank of relative importance.^c : Fc: fixed carbon; V: volatiles; A: ash.^d : P: protein; L:lipid; CA: carbohydrate.^e : T: temperature; RT: residence time; SC: solid content.

expressed optimal performance (average R^2 0.80) among these algorithms for predicting three targets simultaneously. To assess the model's reliability, reverse optimization schemes based on RF models were designed and verified experimentally to guide desired bio-oil production. The GPR models for predicting the yields of all products (hydrochar, bio-oil, aqueous phase, and syngas involved in the HTL system) and bio-oil elemental compositions and calorific value were developed by inputting elemental compositions and HTL conditions (Shafizadeh et al., 2022). The feature importance and effects analysis were also conducted to determine the formation mechanisms of various products. The maximum bio-oil yield can reach 48.7–53.5 % from the optimal biomass compositions (C, H, O, N, S, and ash) and HTL parameters (dry matter, temperature, residence time, and pressure) optimized by the ML model. Furthermore, RF models were developed with different input features for predicting nitrogen heterocycles in bio-oil from HTL of biomass (Leng et al., 2022c). The predictive performance was improved (R^2 from 0.59 to 0.82) after considering the bio-oil yield and N content as inputs, and the N content of bio-oil was the most important parameter

to nitrogen heterocycles of bio-oil.

Generally, the HHV of bio-oil obtained from HTL of algae during was similar to that of conventional diesel (Chen et al., 2022a; Eboibi et al., 2014). Thus, algae, including both microalgae and macroalgae, is regarded as a promising renewable biomass for HTL to generate bio-oil with high HHV (Kumar et al., 2020). The GBR and RF algorithms were used to predict yield, N content, and O content of bio-oil from HTL of algal biomass (Zhang et al., 2021). The forward and reverse optimizations for high-quality bio-oil (e.g., high yield but low N and O contents) were conducted to design experiments for the *Chlorella* based on GBR models. Due to the limited datasets and redundant features for ML training, the errors between experiments and prediction values were relatively large (>35 % for some bio-oil yield and N content) from reversed optimization. However, the overall errors for these three targets are acceptable (<8%) from forward optimization. This way, ML-aided HTL can be promising to guide experimental studies to engineer bio-oil. Recently, a ML-based mechanism study reported that the RF model could effectively predict and produce bio-oil from lignin

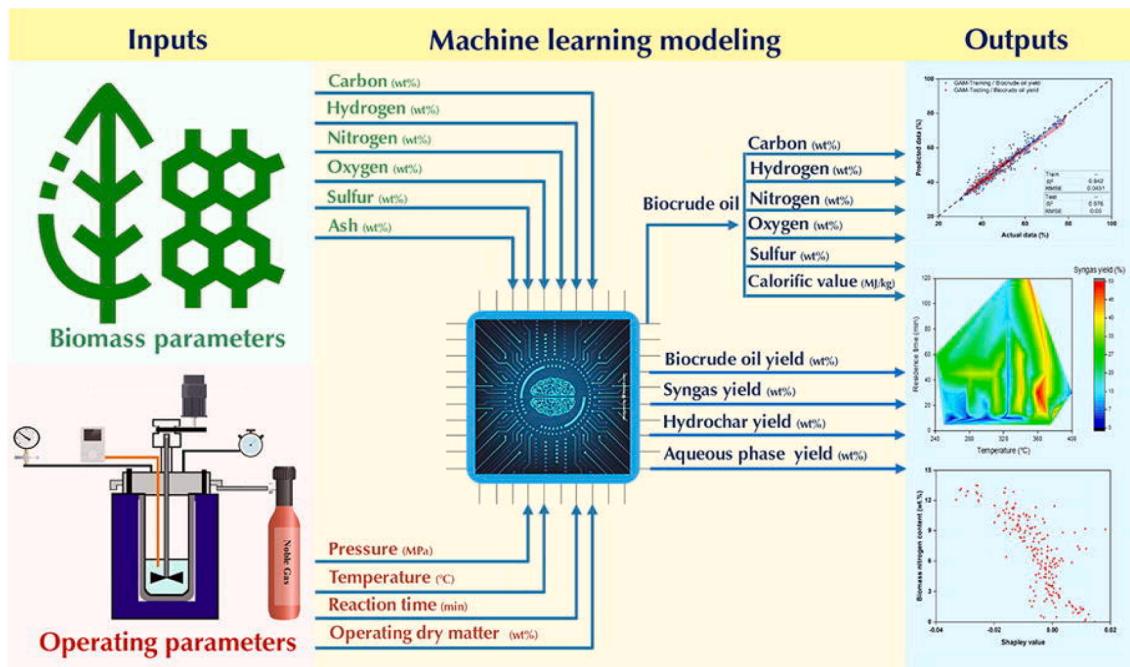


Fig. 4. ML-aided prediction, optimization, and interpretation for bio-oil from hydrothermal liquefaction. Figure cited from ref. (Shafizadeh et al., 2022). Copyright Elsevier 2022.

depolymerization during HTL considering solvent and catalyst (Castro Garcia et al., 2022). The solvent-related properties and temperature were more important than the catalyst to bio-oil yield. Compared to biomass compositions, the reaction solvent (mostly water), extracting solvent (mostly dichloromethane), and extraction solvent and procedure had little effect on bio-oil yield (Li et al., 2021c).

5. Hydrothermal gasification (supercritical water gasification)

5.1. Hydrothermal gasification process

HTG, or supercritical water (e.g., 374 °C and 22 MPa) gasification (SCWG), is a promising and clean process for syngas production under severe hydrothermal conditions (temperature > 380 °C and pressure > 25 MPa) (Leng et al., 2020a). During HTG, wet biomass and wastes can be converted to gaseous fuel. The main compositions of syngas from HTG are CO₂, CO, H₂, and C₁–C₄ hydrocarbons (Shahbeik et al., 2022; Zhang et al., 2010). The involved stoichiometric reaction equations in HTG were summarized from biomass model compounds (e.g., lignin, carbohydrate, phenols, alcohols, lipids, protein, and amino acids) (He et al., 2014). In detail, water promoted the hydrolysis of these macromolecules by acting as a solvent and catalyzing the derivatives to generate H₂ or CH₄ (Lee et al., 2015). For lowering the HTG operating conditions and improving syngas productivity, applying either homogeneous or heterogeneous catalysts to HTG was discussed frequently (Kruse, 2009; Shahbeik et al., 2022). For example, the most used homogeneous catalysts in HTG of biomass were Na₂CO₃, KHCO₃, K₂CO₃, NaOH, and KOH. In contrast, metal catalysts (e.g., Ni, Pt, Ru, Rh, Pd, and Ir) were commonly used as heterogeneous catalysts (Guo et al., 2010). The heterogeneous catalysts expressed higher hydrogen selectivity than homogeneous catalysts (He et al., 2014). Besides, the formation of tar would be restrained by the introduction of catalysts, which influenced the chemical properties of syngas.

5.2. ML-aided HTG

ML has been applied to study HTG performance, particularly the prediction of H₂ productivity. For example, predicting H₂ production

from sewage sludge by proximate compositions, ultimate compositions, and HTG conditions was accurate (R^2 0.997) using the Gaussian performance regression model (Haq et al., 2022). The RF algorithm could accurately predict H₂ yield (R^2 0.98) for HTG of the agricultural waste and municipal solid waste using biomass compositions and HTG parameters (Zhao et al., 2021). Temperature, solid contents, pressure, and time are the top four critical features for H₂ yield, though these orders may be slightly different in different studies (Table 6). The PDP indicated that H₂ yield increased with the increasing temperature, which showed a strong positive relationship between H₂ production and HTG temperature. However, the effect of pressure on H₂ yield was complex, and high pressure seemed to promote H₂ production. Increasing reaction time could raise H₂ yield only when residence time was within 10 min. Still, the H₂ production would decrease when residence time was longer than 10 min, possibly due to the methanation reactions (Zhao et al., 2021).

Furthermore, ML-aided process parameters optimization was employed for generating maximum H₂ with various biomass. Despite limited studies concerning ML-aided catalytic HTG, NN models were used for maximal H₂ by screening catalyst and optimizing HTG process parameters (Li et al., 2021a). Among all solutions of the ML model optimization, Fe-compound catalyst (FeCl₃ or Fe) might be an optimal choice for promoting H₂ production among different catalyst types at high temperatures and low solid contents (Li et al., 2021a). However, the optimal schemes and conclusions should be validated by experimentalists to enhance the model's credibility.

The joint prediction of various gas compositions by ML can better simulate the gas phase products from HTG. A GBR model with tuned hyper-parameters was built for the prediction of H₂, CH₄, CO, and CO₂ with testing R^2 > 0.90 (Li et al., 2020b). ML-based optimization could realize the directional adjustment and control of syngas compositions, such as H₂-rich syngas (Fig. 5). Besides, a multi-task NN model was proposed for predicting syngas compositions (CO, CO₂, CH₄, and H₂) from HTG of food wastes, and R^2 as high as 0.99 and mean squared error (MSE) value below 0.30 were obtained (Shenbagaraj et al., 2021). However, the aforementioned ML models only developed relationship between inputs and syngas compositions, while the ML-based feature analysis and multicomponent optimization were not conducted.

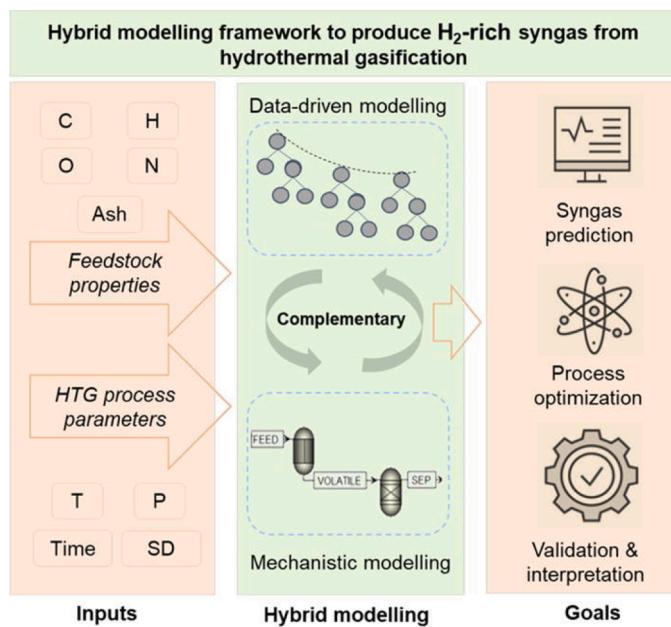


Fig. 5. ML-aided prediction and mechanism interpretation for syngas from hydrothermal gasification. Figure cited from ref. (Li et al., 2021b). Copyright Elsevier 2021.

Nevertheless, the feature effects based on multi-task models were analyzed by Shapley additive explanation (SHAP) analysis (Lundberg and Lee, 2017) in another study for the four main gas compositions (Li et al., 2021a). The three most critical features were temperature, carbon content, and N content, contributing about 70 % importance among all input variables. The yield of the tar and four main gas compositions has been predicted and optimized by ANN models for clean gas from biomass gasification (Kargbo et al., 2021), which could be a reference for further application of ML in the HTG field.

6. Research needs and future directions

6.1. ML-aided valorization of aqueous phase byproduct from hydrothermal processes

Since water acts as a reaction solvent in HTT, the abundant wastewater (process byproduct) from the aqueous phase is inevitably produced (Xu et al., 2022). The aqueous phase was generated in a large amount, about 1–20 times more than the weight of dry biomass feedstocks (Leng et al., 2021b, 2020a). The valorization of aqueous phase byproducts is becoming a research hotspot for resource recycling and a challenge for commercial viability of HTT (Watson et al., 2020). Many

effective methods have been proposed to treat and utilize the valuable component in aqueous phase wastewater (Leng et al., 2021c). However, the complex nature of the aqueous phase hinders its valorization. ML has been applied for the prediction of pH, chemical oxygen demand (COD), total organic carbon (TOC), total phosphorus (TP), and total nitrogen (TN) of the aqueous phase within testing R^2 higher than 0.85 (Leng et al., 2022d). Fig. 6 shows ML-aided prediction and mechanism interpretation of aqueous phase wastewater properties from HTT. ML was also proved to be an efficient method to valorize valuable chemicals from HTT aqueous phase. For instance, maximum sugar yield can be optimized and explained by ML combined with the orthogonal experimental design from hydrolysis of oil palm mesocarp fibers (Lee et al., 2022). ML-aided glucose production and optimization in HTT were conducted by NN models and experiments (Wei-Hsin Chen et al., 2022). The quantitative correlations have been revealed between biomass compositions & HTT conditions and these wastewater indicators (Tables 5–6), which can guide the selection of suitable biomass and HTT parameters.

Furthermore, it is essential to consider more properties or compositions of the aqueous phase in ML modelling, such as K, Na, phosphate, and heavy metal (Leng et al., 2021b, 2020a). More valuable chemicals can be prepared effectively with the guidance of ML. The experiments or testing for more properties or compositions of the aqueous phase are different to achieve, and can inevitably cause more statistical errors. Generally, picking the aqueous phase evenly, increasing the number of parallel measurements or experiments (at least 3 times) and adding controlled trials are effective ways to reduce statistical errors. More characteristics of the aqueous phase are valuable to ML models for comprehensively prediction and understanding the relationships between inputs and outputs. Despite more statistical errors may be generated, the errors could be ignored generally by ML training models with the data of the aqueous phase due to their strong fitting capabilities. Finally, the considered properties or compositions of the aqueous phase (major descriptors can be seen in Table 1) should be agreed with the requirement of the researchers, which means that not all properties need to be measured.

Most of the HTC-related studies would not several limited bio-oil and aqueous phase, the properties of liquid phase would be measured as a whole. Thus, the accurate ML models for predicting the properties of liquid phase in HTC should be developed with same measurement methods (measure bio-oil and aqueous phase as a whole). In addition, ML prediction of aqueous phase byproduct is promising and meaningful to understand the HTT mechanisms because many substance compositions are first hydrolyzed to the aqueous phase, then converted to hydrochar, bio-oil, or gas products.

6.2. ML-aided hydrothermal reaction mechanisms

Traditionally, the hydrothermal reaction mechanisms are revealed

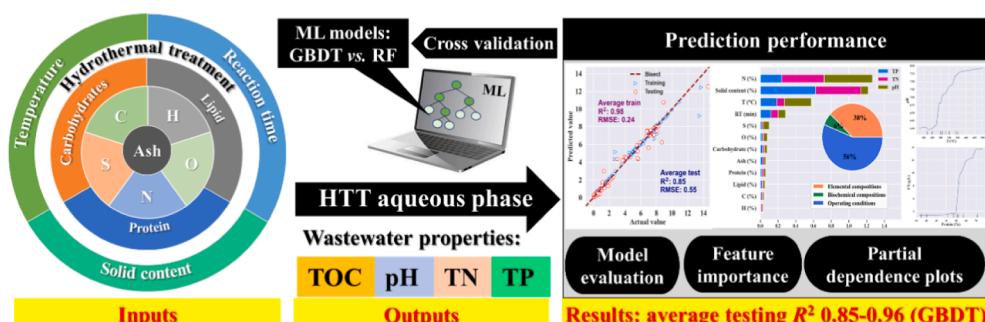


Fig. 6. ML-aided prediction and mechanism interpretation for aqueous phase from hydrothermal treatment. Figure cited from ref. (Leng et al., 2022d). Copyright Elsevier 2022.

through lots of experiments and characterizations. Generally, experiments were carried out with specific biomass and limited operating conditions in general. Biomass and model chemicals have different reaction pathways and mechanisms to generate various products from HTT. For example, the carbohydrate was main participant to hydrochar (Obeid et al., 2020), the protein and carbohydrate interact for promoting Maillard reactions to produce aromaticity and nitrogenous heterocyclic functional groups (Lu et al., 2022), and lipid and lignin expressed an antagonistic effect on the bio-oil yield (Lu et al., 2018). Therefore, there is still a lack of universal laws to understand the reaction mechanisms during different hydrothermal processes in terms of biomass compositions and quality of obtained products. The explanation results based on accurate ML models can effectively aid the study of product formation and reaction mechanism during HTT of biomass. For example, feature importance analysis can show the relative importance of different input variables on the output, and the nonlinear function relationships between inputs and targets can be explicitly drawn with PDP. In detail, the molecular modeling data (e.g., biomass model chemicals, catalysts, solvents, and additives), as well as hydrothermal conditions, can be treated as input features to figure out the biomass hydrothermal conversion behavior and their effects on the yields, compositions, or properties of bio-oil, hydrochar, syngas, and aqueous phases with ML models. Hence, real-time monitoring of HTT products and intermediates is essential to achieve a comprehensive understanding of the reaction pathways during the HTT of biomass, which should be conducted in the subsequent work for thorough data.

Moreover, ML studies can combine with density functional theory (DFT) to reduce computational cost and analysis time for a rational pathway (Butler et al., 2018). In detail, the data can be obtained from different molecular numbers and structures during the HTT process (Liu et al., 2021b). The descriptors of these data could be targeted as inputs or outputs to ML models for training. The physicochemical properties and structural parameters of biomass models calculated by DFT can be inputs. The physical and chemical properties and applied property parameters obtained by DFT can be outputs. In detail, the structures and atomic numbers of biomass could be geometric optimized as inputs. The DFT results could be used as outputs, such as van der Waals surface electrostatic potential, bond strength of different functional groups, bond energy, molecular vibrations and orbitals, electric density and so on.

As the different kinds of biomass feedstocks have different compositions, the gaps of certain biomass for generating intermediates and products (hydrochar, bio-oil, gas, and aqueous phase) in HTT vary sharply. For example, the protein and ash contents of sludge are higher than lignocellulosic biomass, while carbohydrate (here containing cellulose, hemicellulose, and lignin) is the major compound of lignocellulosic biomass (see supplementary material). Thus, protein decompositions were main reactions in HTT of sludge, and the bio-oil and hydrochar from sludge usually have low yields and quality (e.g., N content is high). The lignocellulosic biomass could be converted into various chemicals and biofuels due to high carbohydrate contents. The predictive performance ($R^2 > 0.85$ for bio-oil yield and N content) (Zhang et al., 2021) of ML models from HTT of algae was better than that ($R^2 \sim 0.80$ for bio-oil yield and N content) (Li et al., 2021c) of all considered biomass types (i.e., food waste, manure, algae, sludge, and lignocellulose). The mainly reason is that the different and complex reaction pathways of various biomass limited the accuracy of ML, while ML could show satisfactory performance for specific group of feedstocks (i.e., algae). In summary, the existing data and features are not enough to train an accurate ML model or conduct data-driven mechanism interpretation with all biomass data. The ML models are encouraged to train models for accuracy prediction with the specific biomass type or model compound, because of the similar compositions and reaction pathways of these biomass or model compounds.

6.3. Development of machine learning technique

Single ML model may not be capable of simulating the reaction process due to the complexity of different hydrothermal processes. The combination of ML models with advanced statistical technologies and optimization algorithms has excellent potential in biomass conversions. Thus, researchers turned attention to hybrid models to simulate and understand the biomass thermal conversion process. The hybrid ML models generally show better abilities of accuracy and generalizability than traditional models. Some hybrid ML models have been developed in the biomass thermal conversion processes. For example, a genetic algorithm based features selection before training RF models was employed for bio-oil yield prediction (Ullah et al., 2021); a response surface methodology was integrated into ANN models for optimizing process parameters and maximum bio-oil production (Gupta et al., 2022); and Adaboost was coupled with an ANN algorithm to describe relationships between the thermal treatment process and product yield precisely (Hu et al., 2022). Therefore, using these advanced hybrid ML models in HTT of biomass is full of great possibilities.

Moreover, deep learning (DL) has been an ideal choice for its diversification and accuracy of prediction compared to traditional ML methods. The relationships among AI, ML, and DL can be shown in Fig. 1a. DL, or advanced ML, has been widely used in other biomass conversion technologies. For example, a DL network was successfully applied to predict the lower heating value of the syngas from biomass gasification (Kartal and Özveren, 2020), with R^2 reaching 0.99. The R^2 of CO, CO₂, CH₄, H₂, and HHV predictions for gas from biomass gasification were higher than 0.85 by MLP models (Elmaz et al., 2020b). DL models have numerous advantages, especially in image identification (Wang et al., 2017), but few DL-based models have been built in hydrothermal processes. There are still some problems to be solved when using DL models in this field, such as the requirement of big data, the high computational cost, and the difficulty in model interpretation.

6.4. Increase model generalizability and interpretability

In general, 80 % of the entire data is considered as a training dataset, and the remaining 20 % is considered a testing dataset during the ML model development. Different training and testing datasets ratios can be tried to increase model predictive ability and generalizability, but too few testing data can result in overfitting. The training process and results with different datasets ratios should be shown and discussed in detail. Besides, the R^2 and RMSE are simply used for evaluating ML models in most studies. More statistical indicators (i.e., mean square error (MSE), mean absolute error (MAE), mean absolute percentage error (MAPE), root mean squared relative error (RMSPE)) could be applied to comprehensively assess ML models' performance.

The data numbers, descriptors, and distributions should be evaluated carefully to achieve good generalizability. If the distributions of data are unsatisfactory (e.g., data features of limited or biased coverages), bad generalizability of the models may appear. Creating highly generalizable models suitable across a wide range of feedstocks and various thermochemical parameters & ranges should be promoted in the future. Recently, a synthetic minority oversampling (SMOTE) algorithm (Chawla et al., 2002) was introduced to ML models for dealing with imbalanced data to get accurate predictive performance for plant-scale biogas production (Zhang et al., 2022). The input data of HTT descriptors can be filled with synthetic samples from the SMOTE algorithm, then unevenly distributed data can be effectively improved.

Although ML models generally express high performance, their "black box" nature is the main drawback. To further explain the developed ML models, open the "black box" of ML. Interpretation methods can be adopted and compared, such as PDP, Individual Conditional Expectation (ICE), Permutated Feature Importance, Global Surrogate, Local Surrogate (LIME), Shapley Value, and Gini importance (Ribeiro et al., 2016). These interpretable methods depend on different

principles, which could be selected and compared as needed for various ML models. All of those interpretation methods can be used to explain the behavior and predictions of trained machine learning models. But the interpretation methods might not work well in the following cases (Molnar, 2019): 1) if model models interactions (e.g., when a random forest is used); 2) if features strongly correlate with each other; 3) if the model does not correctly model causal relationships; 4) if parameters of the interpretation method are not set correctly.

6.5. Data and model sharing

It is challenging for researchers to collect the original hydrothermal data (containing important descriptors) from all related literature. Data processing and feature cleaning are time-consuming and laborious. It is necessary to share the continuous dataset in open-source communities. Besides, the developed ML models with different algorithms are recommended to share the codes in the public platform (e.g., GitHub project, <https://github.com/>). The shared data and models can be assessed, used, and improved by others for a better understanding of the hydrothermal process of biomass. Compared with local application software that should be downloaded (Shafizadeh et al., 2022; Ullah et al., 2021), the online website's graphical user interfaces (Leng et al., 2022d) are more user-friendly. Therefore, such communities should be built to promote the development of HTT.

6.6. ML-aided industrialization of hydrothermal processes

The ML is a strong and effective tool for prediction, optimization, and interpretation in HTT of biomass. For example, accurate prediction models could be developed with abundant data, data-driven schemes could be obtained based on ML for experimentalists, and explainable relationships could be established between input and outputs for scientists. With accurate prediction performance, the final purpose of ML models is to guide actual experiments (Ai et al., 2022). The as-obtained models can be applied to various functions. Firstly, the prediction functions of ML models are primarily used for predicting the yield, compositions, and characteristics of HTT products. Without conducting complex experiments, the results can be easily obtained with given input variables by ML. Moreover, the optimization functions of ML models are necessary and valuable for optimal output results within some restricted conditions. Therein, the forward optimization is to screen desired solutions with predictive results of ML models. For example, the prediction values from trained ML models can be presented by inputting one specific biomass (with measured compositions) and each variable of operating conditions (with set steps in the range). The optimum product property is taken as the standard for choosing optimal conditions within iteration results from ML. The reverse optimization means optimum input parameters combinations can be determined through orientation-basis target functions (e.g., maximum or minimum yield, compositions, and properties of the hydrothermal products) using optimization algorithms, such as the heuristic optimization method (i.e., PSO and GA). The optimization of process conditions (Fang et al., 2022) for preferable targets is practical and instructive. Still, limited studies considered the hydrothermal process optimization, let alone experiment validations.

Most studies focused on single-target predictions for one specific yield, composition, or property of hydrothermal products. Only a few works studied the prediction of multiple targets (Alabdrabnabi, 2021; Castro Garcia et al., 2022), and their results were obtained from multiple single-target prediction models (Shafizadeh et al., 2022). Single target prediction cannot describe the HTT performance adequately. To achieve biomass biorefinery, all products from HTT of biomass should be valorized or disposed of (Watson et al., 2020). Hence, multi-target ML models are preferable for simultaneously predicting multi-phase product yield, compositions, and properties, which are worth further investigation.

In the engineering industry, a wide variety of solvents and catalysts

could be applied to improve the properties of products from HTT. However, the descriptors of solvents or catalysts are rarely considered to build ML models in this field. In general, solvent properties can be described by critical pressure, critical temperature, molecular mass, density, solubility, polarity, and so on. And catalyst characteristics include catalyst ratio, catalyst size, catalyst BET surface area, catalyst pore size, catalyst total pore volume, molecular mass, valency, ionization energy, electron affinity, conductivity, etc. Therefore, the screening of optimal solvents and catalysts for HTT apart from other common operating conditions (temperature, residence time, and solid content) can be realized based on these ML models.

Moreover, the developed models are encouraged to be put into industrialized applications. The data needs to be processed before it can be trained in the ML models (Fig. 1), and visual platform system should be developed stably with ML models. Different models may have different data process methods, which resulting the complexity of training processes. Still, there is no comprehensive unified data and machine learning platform to meet the requirements and applications for hydrothermal treatment of biomass. Therefore, the application of machine learning is still a long way off. One way is to simplify and embed ML algorithms to chips, and then target products' monitoring and regulation can be realized in real-time. The data from the factories were generally intermittent and time-sensitive. In this way, the long short-term memory (LSTM) network (Wu et al., 2021) can be used for training these data from sensor measurements. In the future, suitable ML models could be developed for predicting and optimizing products with various types of industrial data.

7. Conclusions

ML has demonstrated great potential to predict the yield, compositions, and properties of hydrochar, bio-oil, syngas, and aqueous phases obtained from the HTT of biomass. Nevertheless, ML prediction is just the first step with a low technology readiness level to realize ML-aided optimization, monitoring, and controlling of HTT. Although the predictive performances of most ML models were accurate ($R^2 > 0.85$), it is still challenging to interpret and compare various models with no completely-satisfactory model being developed. Therefore, it is imperative to improve ML predictive performance, mechanistic interpretation, process optimization, data sharing, and model application for HTT of biomass towards high-quality production.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data that has been used is confidential.

Acknowledgements

We appreciate the financial support from the National Key Research and Development Program of China (2021YFE0104900), Hunan Provincial Natural Science Foundation of China (2022JJ20064), Open Sharing Fund for the Large-scale Instruments and Equipment of Central South University, and the Science and Technology Innovation Program of Hunan Province (2021RC4005).

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.biotech.2022.128547>.

References

- Ai, Z., Zhang, W., Yang, L., Chen, H., Xu, Z., Leng, L., Li, H., 2022. Investigation and prediction of co-pyrolysis between oily sludge and high-density polyethylene via in-situ DRIFTS, TGA, and artificial neural network. *J. Anal. Appl. Pyrol.* 166, 105610 <https://doi.org/10.1016/j.jaat.2022.105610>.
- Aierzhati, A., Stublein, M.J., Wu, N.E., Kuo, C.-T., Si, B., Kang, X., Zhang, Y., 2019. Experimental and model enhancement of food waste hydrothermal liquefaction with combined effects of biochemical composition and reaction conditions. *Bioresour. Technol.* 284, 139–147. <https://doi.org/10.1016/j.biotech.2019.03.076>.
- Alabdralnabi, A., 2021. Utilization of Machine Learning to Predict Bio-Oil and Biochar Yields from CoPyrolysis of Biomass with Waste Polymers. *Fuel* 328, 125303. <https://doi.org/10.1016/j.fuel.2022.125303>.
- Ascher, S., Sloan, W., Watson, I., You, S., 2022a. A comprehensive artificial neural network model for gasification process prediction. *Appl. Energy* 320, 119289. <https://doi.org/10.1016/j.apenergy.2022.119289>.
- Ascher, S., Watson, I., You, S., 2022b. Machine learning methods for modelling the gasification and pyrolysis of biomass and waste. *Renew. Sustain. Energy Rev.* 155, 111902 <https://doi.org/10.1016/j.rser.2021.111902>.
- Becker, R., Dorgerloh, U., Paulke, E., Mumme, J., Nehls, I., 2014. Hydrothermal carbonization of biomass: Major organic components of the aqueous phase. *Chem. Eng. Technol.* 37, 511–518. <https://doi.org/10.1002/ceat.201300401>.
- Brindhadevi, K., Anto, S., Rene, E.R., Sekar, M., Mathimani, T., Chi, T.L.N., Pugazhendhi, A., 2021. Effect of reaction temperature on the conversion of algal biomass to bio-oil and biochar through pyrolysis and hydrothermal liquefaction. *Fuel* 285, 119106. <https://doi.org/10.1016/j.fuel.2020.119106>.
- Butler, K.T., Davies, D.W., Cartwright, H., Isayev, O., Walsh, A., 2018. Machine learning for molecular and materials science. *Nature* 559, 547–555. <https://doi.org/10.1038/s41586-018-0337-2>.
- Castro Garcia, A., Shuo, C., Cross, J.S., 2022. Machine learning based analysis of reaction phenomena in catalytic lignin depolymerization. *Bioresour. Technol.* 345, 126503 <https://doi.org/10.1016/j.biotech.2021.126503>.
- Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P., 2002. SMOTE: Synthetic Minority Over-sampling Technique. *J. Artif. Intell. Res.* 16, 321–357. <https://doi.org/10.1613/jair.953>.
- Chen, Jiefeng, Ding, L., Wang, P., Zhang, W., Li, J., A. Mohamed, B., Chen, Jie, Leng, S., Liu, T., Leng, L., Zhou, W., 2022. The Estimation of the Higher Heating Value of Biochar by Data-Driven Modeling. *J. Renew. Mater.* 10, 1555–1574. 10.32604/jrm.2022.018625.
- Chen, W.-H., Lo, H.-J., Aniza, R., Lin, B.-J., Park, Y.-K., Kwon, E.E., Sheen, H.-K., Grafilo, L.A.D.R., 2022c. Forecast of glucose production from biomass wet torrefaction using statistical approach along with multivariate adaptive regression splines, neural network and decision tree. *Appl. Energy* 324, 119775. <https://doi.org/10.1016/j.apenergy.2022.119775>.
- Chen, J., Wang, P., Ding, L., Yu, T., Leng, S., Chen, J., Fan, L., Li, J., Wei, L., Li, J., Leng, L., Zhou, W., 2021. The comparison study of multiple biochar stability assessment methods. *J. Anal. Appl. Pyrol.* 156, 105070 <https://doi.org/10.1016/j.jaat.2021.105070>.
- Chen, J., Zhang, J., Pan, W., An, G., Deng, Y., Li, Y., Hu, Y., Xiao, Y., Liu, T., Leng, S., Chen, J., Li, J., Peng, H., Leng, L., Zhou, W., 2022a. A novel strategy to simultaneously enhance bio-oil yield and nutrient recovery in sequential hydrothermal liquefaction of high protein microalgae. *Energy Convers. Manage.* 255, 115330 <https://doi.org/10.1016/j.enconman.2022.115330>.
- Cheng, F., Belden, E.R., Li, W., Shahabuddin, M., Paffenroth, R.C., Timko, M.T., 2022. Accuracy of predictions made by machine learned models for biocrude yields obtained from hydrothermal liquefaction of organic wastes. *Chem. Eng. J.* 442, 136013 <https://doi.org/10.1016/j.cej.2022.136013>.
- Djandja, O.S., Duan, P.-G., Yin, L.-X., Wang, Z.-C., Duo, J., 2021. A novel machine learning-based approach for prediction of nitrogen content in hydrochar from hydrothermal carbonization of sewage sludge. *Energy* 232, 121010. <https://doi.org/10.1016/j.energy.2021.121010>.
- Djandja, O.S., Salami, A.A., Wang, Z.-C., Duo, J., Yin, L.-X., Duan, P.-G., 2022. Random forest-based modeling for insights on phosphorus content in hydrochar produced from hydrothermal carbonization of sewage sludge. *Energy* 245, 123295. <https://doi.org/10.1016/j.energy.2022.123295>.
- Eboobi, B.E., Lewis, D.M., Ashman, P.J., Chinnasamy, S., 2014. Effect of operating conditions on yield and quality of biocrude during hydrothermal liquefaction of halophytic microalgae Tetraselmis sp. *Bioresour. Technol.* 170, 20–29. <https://doi.org/10.1016/j.biotech.2014.07.083>.
- Elmaz, F., Büyükkakir, B., Yücel, Ö., Mutlu, A.Y., 2020a. Classification of solid fuels with machine learning. *Fuel* 266. <https://doi.org/10.1016/j.fuel.2020.117066>.
- Elmaz, F., Yücel, Ö., Mutlu, A.Y., 2020b. Predictive modeling of biomass gasification with machine learning-based regression methods. *Energy* 191. <https://doi.org/10.1016/j.energy.2019.116541>.
- Fan, L., Zhang, H., Li, J., Wang, Y., Leng, L., Li, J., Yao, Y., Lu, Q., Yuan, W., Zhou, W., 2020. Algal biorefinery to value-added products by using combined processes based on thermochemical conversion: A review. *Algal Res.* 47, 101819 <https://doi.org/10.1016/j.algal.2020.101819>.
- Fang, Y., Ma, L., Yao, Z., Li, W., You, S., 2022. Process optimization of biomass gasification with a Monte Carlo approach and random forest algorithm. *Energy Convers. Manage.* 264, 115734 <https://doi.org/10.1016/j.enconman.2022.115734>.
- Funke, A., Ziegler, F., 2010. Hydrothermal carbonization of biomass: A summary and discussion of chemical mechanisms for process engineering. *Biofuels. Bioprod. Biorefining* 4, 160–177. <https://doi.org/10.1002/bbb.198>.
- Gao, K., Mei, G., Piccialli, F., Cuomo, S., Tu, J., Huo, Z., 2020. Julia language in machine learning: Algorithms, applications, and open issues. *Comput. Sci. Rev.* 37, 100254 <https://doi.org/10.1016/j.cosrev.2020.100254>.
- Goldstein, A., Kapelner, A., Bleich, J., Pitkin, E., 2015. Peeking Inside the Black Box: Visualizing Statistical Learning With Plots of Individual Conditional Expectation. *J. Comput. Graph. Stat.* 24, 44–65. <https://doi.org/10.1080/10618600.2014.907095>.
- Guo, Y., Wang, S.Z., Xu, D.H., Gong, Y.M., Ma, H.H., Tang, X.Y., 2010. Review of catalytic supercritical water gasification for hydrogen production from biomass. *Renew. Sustain. Energy Rev.* 14, 334–343. <https://doi.org/10.1016/j.rser.2009.08.012>.
- Gupta, S., Patel, P., Mondal, P., 2022. Biofuels production from pine needles via pyrolysis: Process parameters modeling and optimization through combined RSM and ANN based approach. *Fuel* 310, 122230. <https://doi.org/10.1016/j.fuel.2021.122230>.
- Haq, Z.U., Ullah, H., Khan, M.N.A., Naqvi, S.R., Ahsan, M., 2022. Hydrogen production optimization from sewage sludge supercritical gasification process using machine learning methods integrated with genetic algorithm. *Chem. Eng. Res. Des.* 184, 614–626. <https://doi.org/10.1016/j.cherd.2022.06.020>.
- He, M., Cao, Y., Xu, Z., You, S., Ruan, R., Gao, B., Wong, K.-H., Tsang, D.C.W., 2022a. Process water recirculation for catalytic hydrothermal carbonization of anaerobic digestate: Water-Energy-Nutrient Nexus. *Bioresour. Technol.* 361, 127694 <https://doi.org/10.1016/j.biotech.2022.127694>.
- He, C., Chen, C.L., Giannis, A., Yang, Y., Wang, J.Y., 2014. Hydrothermal gasification of sewage sludge and model compounds for renewable hydrogen production: A review. *Renew. Sustain. Energy Rev.* 39, 1127–1142. <https://doi.org/10.1016/j.rser.2014.07.141>.
- He, M., Xu, Z., Hou, D., Gao, B., Cao, X., Ok, Y.S., Rinklebe, J., Bolan, N.S., Tsang, D.C.W., 2022b. Waste-derived biochar for water pollution control and sustainable development. *Nat. Rev. Earth Environ.* 3, 444–460. <https://doi.org/10.1038/s43017-022-00306-8>.
- He, M., Zhu, X., Dutta, S., Khanal, S.K., Lee, K.T., Masek, O., Tsang, D.C.W., 2022c. Catalytic co-hydrothermal carbonization of food waste digestate and yard waste for energy application and nutrient recovery. *Bioresour. Technol.* 344, 126395 <https://doi.org/10.1016/j.biotech.2021.126395>.
- Hoekman, S.K., Broch, A., Robbins, C., 2011. Hydrothermal Carbonization (HTC) of Lignocellulosic Biomass. *Energy Fuel* 25, 1802–1810. <https://doi.org/10.1021/ef101745n>.
- Hu, Z., Yuan, Y., Li, X., Tu, Z., Donovan Dacres, O., Zhu, Y., Shi, L., Hu, H., Liu, H., Luo, G., Yao, H., 2022. Yield prediction of “Thermal-dissolution based carbon enrichment” treatment on biomass wastes through coupled model of artificial neural network and AdaBoost. *Bioresour. Technol.* 343, 126083 <https://doi.org/10.1016/j.biotech.2021.126083>.
- Jordan, M.I., Mitchell, T.M., 2015. Machine learning: Trends, perspectives, and prospects. *Science* (80-). 349, 255–260. 10.1126/science.aaa8415.
- Juneja, M., Nagar, S.K., 2016. Particle swarm optimization algorithm and its parameters: A review, in: 2016 International Conference on Control, Computing, Communication and Materials (ICCCCM). IEEE, pp. 1–5. 10.1109/ICCCCM.2016.7918233.
- Kan, T., Streozov, V., Evans, T.J., 2016. Lignocellulosic biomass pyrolysis: A review of product properties and effects of pyrolysis parameters. *Renew. Sustain. Energy Rev.* 57, 126–140. <https://doi.org/10.1016/j.rser.2015.12.185>.
- Kargbo, H.O., Zhang, J., Phan, A.N., 2021. Optimisation of two-stage biomass gasification for hydrogen production via artificial neural network. *Appl. Energy* 302, 117567. <https://doi.org/10.1016/j.apenergy.2021.117567>.
- Kartal, F., Özveren, U., 2020. A deep learning approach for prediction of syngas lower heating value from CFB gasifier in Aspen plus®. *Energy*. <https://doi.org/10.1016/j.energy.2020.118457>.
- Katongtong, T., Onsree, T., Tippayawong, N., 2022. Machine learning prediction of biocrude yields and higher heating values from hydrothermal liquefaction of wet biomass and wastes. *Bioresour. Technol.* 344, 126278 <https://doi.org/10.1016/j.biotech.2021.126278>.
- Khan, M., Raza Naqvi, S., Ullah, Z., Ali Ammar Taqvi, S., Nouman Aslam Khan, M., Farooq, W., Taqi Mehran, M., Juchelková, D., Štěpanec, L., 2023. Applications of machine learning in thermochemical conversion of biomass-A review. *Fuel* 332, 126055. <https://doi.org/10.1016/j.fuel.2022.126055>.
- Kruse, A., 2009. Hydrothermal biomass gasification. *J. Supercrit. Fluids* 47, 391–399. <https://doi.org/10.1016/j.supflu.2008.10.009>.
- Kruse, A., Funke, A., Titirici, M.-M., 2013. Hydrothermal conversion of biomass to fuels and energetic materials. *Curr. Opin. Chem. Biol.* 17, 515–521. <https://doi.org/10.1016/j.cbpa.2013.05.004>.
- Kumar, M., Sun, Y., Rathour, R., Pandey, A., Thakur, I.S., Tsang, D.C.W., 2020. Algae as potential feedstock for the production of biofuels and value-added products: Opportunities and challenges. *Sci. Total Environ.* 716, 137116 <https://doi.org/10.1016/j.scitotenv.2020.137116>.
- LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. *Nature* 521, 436–444. <https://doi.org/10.1038/nature14539>.
- Lee, S.H., Li, J., Wang, X., Yang, K.-L., 2022. Online-learning-aided optimization and interpretation of sugar production from oil palm mesocarp fibers with analytics for industrial applications. *Resour. Conserv. Recycl.* 180, 106206 <https://doi.org/10.1016/j.resconrec.2022.106206>.
- Lee, I.-G., Nowacka, A., Yuan, C.-H., Park, S.-J., Yang, J.-B., 2015. Hydrogen production by supercritical water gasification of valine over Ni/activated charcoal catalyst modified with Pt, Pd. *Int. J. Hydrogen Energy* 40, 12078–12087. <https://doi.org/10.1016/j.ijhydene.2015.07.112>.
- Leng, E., He, B., Chen, J., Liao, G., Ma, Y., Zhang, F., Liu, S., 2021a. Prediction of three-phase product distribution and bio-oil heating value of biomass fast pyrolysis based

- on machine learning. *Energy* 236, 121401. <https://doi.org/10.1016/j.energy.2021.121401>.
- Leng, L., Huang, H., Li, H., Li, J., Zhou, W., 2019. Biochar stability assessment methods: A review. *Sci. Total Environ.* 647, 210–222. <https://doi.org/10.1016/j.scitotenv.2018.07.402>.
- Leng, S., Leng, L., Chen, L., Chen, J., Chen, J., Zhou, W., 2020c. The effect of aqueous phase recirculation on hydrothermal liquefaction/carbonization of biomass: A review. *Bioresour. Technol.* 318, 124081 <https://doi.org/10.1016/j.biortech.2020.124081>.
- Leng, L., Yang, L., Chen, J., Hu, Y., Li, H., Li, H., Jiang, S., Peng, H., Yuan, X., Huang, H., 2021. Valorization of the aqueous phase produced from wet and dry thermochemical processing biomass: A review. *J. Clean. Prod.* 294, 126238 <https://doi.org/10.1016/j.jclepro.2021.126238>.
- Leng, L., Xiong, Q., Yang, L., Li, H., Zhou, Y., Zhang, W., Jiang, S., Li, H., Huang, H., 2021b. An overview on engineering the surface area and porosity of biochar. *Sci. Total Environ.* 763, 144024 <https://doi.org/10.1016/j.scitotenv.2020.144024>.
- Leng, L., Yang, L., Leng, S., Zhang, W., Zhou, Y., Peng, H., Li, H., Hu, Y., Jiang, S., Li, H., 2021e. A review on nitrogen transformation in hydrochar during hydrothermal carbonization of biomass containing nitrogen. *Sci. Total Environ.* 756, 143679 <https://doi.org/10.1016/j.scitotenv.2020.143679>.
- Leng, L., Liu, R., Xu, S., Mohamed, B.A., Yang, Z., Hu, Y., Chen, J., Zhao, S., Wu, Z., Peng, H., Li, H., Li, H., 2022a. An overview of sulfur-functional groups in biochar from pyrolysis of biomass. *J. Environ. Chem. Eng.* 10, 107185 <https://doi.org/10.1016/j.jece.2022.107185>.
- Leng, L., Zhou, W., 2018. Chemical compositions and wastewater properties of aqueous phase (wastewater) produced from the hydrothermal treatment of wet biomass: A review. *Energy Sources Part A Recover. Util. Environ. Eff.* 40, 2648–2659. <https://doi.org/10.1080/15567036.2018.1495780>.
- Leng, L., Zhang, W., Leng, S., Chen, J., Yang, L., Li, H., Jiang, S., Huang, H., 2020a. Bioenergy recovery from wastewater produced by hydrothermal processing biomass: Progress, challenges, and opportunities. *Sci. Total Environ.* 748, 142383 <https://doi.org/10.1016/j.scitotenv.2020.142383>.
- Leng, L., Zhang, W., Peng, H., Li, H., Jiang, S., Huang, H., 2020b. Nitrogen in bio-oil produced from hydrothermal liquefaction of biomass: A review. *Chem. Eng. J.* 401, 126030 <https://doi.org/10.1016/j.cej.2020.126030>.
- Leng, L., Yang, L., Lei, X., Zhang, W., Ai, Z., Yang, Z., Zhan, H., Yang, J., Yuan, X., Peng, H., Li, H., 2022b. Machine learning predicting and engineering the yield, N content, and specific surface area of biochar derived from pyrolysis of biomass. *Biochar* 4, 63. <https://doi.org/10.1007/s42773-022-00183-w>.
- Leng, L., Zhang, W., Chen, Q., Zhou, J., Peng, H., Zhan, H., Li, H., 2022c. Machine learning prediction of nitrogen heterocycles in bio-oil produced from hydrothermal liquefaction of biomass. *Bioresour. Technol.* 362, 127791 <https://doi.org/10.1016/j.biortech.2022.127791>.
- Leng, L., Zhang, W., Liu, T., Zhan, H., Li, J., Yang, L., Li, J., Peng, H., Li, H., 2022d. Machine learning predicting wastewater properties of the aqueous phase derived from hydrothermal treatment of biomass. *Bioresour. Technol.* 358, 127348 <https://doi.org/10.1016/j.biortech.2022.127348>.
- Li, H., Ai, Z., Yang, L., Zhang, W., Yang, Z., Peng, H., Leng, L., 2023. Machine learning assisted predicting and engineering specific surface area and total pore volume of biochar. *Bioresour. Technol.* 369, 128417 <https://doi.org/10.1016/j.biortech.2022.128417>.
- Li, J., Pan, L., Suvarna, M., Tong, Y.W., Wang, X., 2020b. Machine Learning Prediction of Syngas Composition of Hydrothermal Gasification from Wet Organic Wastes, in: International Conference on Applied Energy 2020. p. 199.
- Li, J., Pan, L., Suvarna, M., Tong, Y.W., Wang, X., 2020. Fuel properties of hydrochar and pyrochar: Prediction and exploration with machine learning. *Appl. Energy* 269, 115166. <https://doi.org/10.1016/j.apenergy.2020.115166>.
- Li, J., Pan, L., Suvarna, M., Wang, X., 2021a. Machine learning aided supercritical water gasification for H₂-rich syngas production with process optimization and catalyst screening. *Chem. Eng. J.* 426, 131285 <https://doi.org/10.1016/j.cej.2021.131285>.
- Li, J., Zhu, X., Li, Y., Tong, Y.W., Ok, Y.S., Wang, X., 2021d. Multi-task prediction and optimization of hydrochar properties from high-moisture municipal solid waste: Application of machine learning on waste-to-resource. *J. Clean. Prod.* 278, 123928 <https://doi.org/10.1016/j.jclepro.2020.123928>.
- Li, J., Li, L., Tong, Y.W., Wang, X., 2022a. Understanding and optimizing the gasification of biomass waste with machine learning. *Green Chem. Eng.* <https://doi.org/10.1016/j.gce.2022.05.006>.
- Li, Y., Liu, H., Xiao, K., Liu, X., Hu, H., Li, X., Yao, H., 2019. Correlations between the physicochemical properties of hydrochar and specific components of waste lettuce: Influence of moisture, carbohydrates, proteins and lipids. *Bioresour. Technol.* 272, 482–488. <https://doi.org/10.1016/j.biortech.2018.10.066>.
- Li, Y., Gupta, R., You, S., 2022c. Machine learning assisted prediction of biochar yield and composition via pyrolysis of biomass. *Bioresour. Technol.* 359, 127511 <https://doi.org/10.1016/j.biortech.2022.127511>.
- Li, J., Suvarna, M., Pan, L., Zhao, Y., Wang, X., 2021b. A hybrid data-driven and mechanistic modelling approach for hydrothermal gasification. *Appl. Energy* 304, 117674. <https://doi.org/10.1016/j.apenergy.2021.117674>.
- Li, J., Zhang, W., Liu, T., Yang, L., Li, H., Peng, H., Jiang, S., Wang, X., Leng, L., 2021c. Machine learning aided bio-oil production with high energy recovery and low nitrogen content from hydrothermal liquefaction of biomass with experiment verification. *Chem. Eng. J.* 425, 130649 <https://doi.org/10.1016/j.cej.2021.130649>.
- Li, J., Suvarna, M., Li, L., Pan, L., Pérez-Ramírez, J., Ok, Y.S., Wang, X., 2022b. A review of computational modeling techniques for wet waste valorization: Research trends and future perspectives. *J. Clean. Prod.* 367, 133025 <https://doi.org/10.1016/j.jclepro.2022.133025>.
- Li, L., Wang, Y., Xu, J., Flora, J.R.V., Hoque, S., Berge, N.D., 2018. Quantifying the sensitivity of feedstock properties and process conditions on hydrochar yield, carbon content, and energy content. *Bioresour. Technol.* 262, 284–293. <https://doi.org/10.1016/j.biortech.2018.04.066>.
- Lin, Y., Chen, X., 2014. Simulated Annealing Algorithm Improved BP Learning Algorithm. *Appl. Mech. Mater.* 517, 734–737. <https://doi.org/10.4028/www.scientific.net/AMM.513-517.734>.
- Liu, T., Gao, X., Mofrad, A.Z., Kudo, S., Asano, S., Hayashi, J., 2021c. Leaching Char with Acidic Aqueous Phase from Biomass Pyrolysis: Removal of Alkali and Alkaline-Earth Metallic Species and Uptakes of Water-Soluble Organics. *Energy Fuel* 35, 12237–12251. <https://doi.org/10.1021/acs.energyfuels.1c01889>.
- Liu, T., Jiao, H., Yang, L., Zhang, W., Hu, Y., Guo, Y., Yang, L., Leng, S., Chen, J., Chen, J., Peng, H., Leng, L., Zhou, W., 2022. Co-hydrothermal carbonization of cellulose, hemicellulose, and protein with aqueous phase recirculation: Insight into the reaction mechanisms on hydrochar formation. *Energy* 123965. <https://doi.org/10.1016/j.energy.2022.123965>.
- Liu, S., Jin, H., Yang, Y., Yu, L., 2021b. Molecular dynamic investigation on nitrogen migration during hydrogen production by indole gasification in supercritical water. *J. Mol. Liq.* 324, 114769 <https://doi.org/10.1016/j.molliq.2020.114769>.
- Liu, H., Zeng, X., Huang, J., Yuan, X., Wang, Q., Ma, L., 2021a. Dietary fiber extracted from pomelo fruitlets promotes intestinal functions, both *in vitro* and *in vivo*. *Carbohydr. Polym.* 252, 117186 <https://doi.org/10.1016/j.carbpol.2020.117186>.
- Lu, J., Liu, Z., Zhang, Y., Savage, P.E., 2018. Synergistic and Antagonistic Interactions during Hydrothermal Liquefaction of Soybean Oil, Soy Protein, Cellulose, Xylose, and Lignin. *ACS Sustain. Chem. Eng.* 6, 14501–14509. <https://doi.org/10.1021/acssuschemeng.8b03156>.
- Lu, J., Watson, J., Liu, Z., Wu, Y., 2022. Elemental migration and transformation during hydrothermal liquefaction of biomass. *J. Hazard. Mater.* 423, 126961 <https://doi.org/10.1016/j.jhazmat.2021.126961>.
- Lundberg, S.M., Lee, S.-I., 2017. A Unified Approach to Interpreting Model Predictions, in: 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA.
- Marzbali, M.H., Kundu, S., Halder, P., Patel, S., Hakeem, I.G., Paz-Ferreiro, J., Madapusi, S., Surapaneni, A., Shah, K., 2021. Wet organic waste treatment via hydrothermal processing: A critical review. *Chemosphere* 279, 130557. <https://doi.org/10.1016/j.chemosphere.2021.130557>.
- Mathimani, T., Mallick, N., 2019. A review on the hydrothermal processing of microalgal biomass to bio-oil – knowledge gaps and recent advances. *J. Clean. Prod.* 217, 69–84. <https://doi.org/10.1016/j.jclepro.2019.01.129>.
- Maulana Kusdhyany, M.I., Lyth, S.M., 2021. New insights into hydrogen uptake on porous carbon materials via explainable machine learning. *Carbon N. Y.* 179, 190–201. <https://doi.org/10.1016/j.carbon.2021.04.036>.
- Mihajlović, M., Petrović, J., Maletić, S., Isakovski, M.K., Stojanović, M., Lopčić, Z., Trifunović, S., 2018. Hydrothermal carbonization of *Miscanthus × giganteus*: Structural and fuel properties of hydrochars and organic profile with the ecotoxicological assessment of the liquid phase. *Energy Convers. Manage.* 159, 254–263. <https://doi.org/10.1016/j.enconman.2018.01.003>.
- Mishra, R.K., Kumar, V., Kumar, P., Mohanty, K., 2022. Hydrothermal liquefaction of biomass for bio-crude production: A review on feedstocks, chemical compositions, operating parameters, reaction kinetics, techno-economic study, and life cycle assessment. *Fuel* 316, 123377. <https://doi.org/10.1016/j.fuel.2022.123377>.
- Molnar, C., 2019. Limitations of Interpretable Machine Learning Methods. *Munich University Press, Munich*.
- Mu, L., Wang, Z., Wu, D., Zhao, L., Yin, H., 2022. Prediction and evaluation of fuel properties of hydrochar from waste solid biomass: Machine learning algorithm based on proposed PSO-NN model. *Fuel* 318, 123644. <https://doi.org/10.1016/j.fuel.2022.123644>.
- Natekin, A., Knoll, A., 2013. Gradient boosting machines, a tutorial. *Front. Neurorob.* 7 <https://doi.org/10.3389/fnbot.2013.00021>.
- Obeid, R., Lewis, D.M., Smith, N., Hall, T., van Eyk, P., 2020. Reaction kinetics and characterisation of species in renewable crude from hydrothermal liquefaction of monomers to represent organic fractions of biomass feedstocks. *Chem. Eng. J.* 389, 124397 <https://doi.org/10.1016/j.cej.2020.124397>.
- Oliveira, A.L.I., Braga, P.L., Lima, R.M.F., Cornélio, M.L., 2010. GA-based method for feature selection and parameters optimization for machine learning regression applied to software effort estimation. *Inf. Softw. Technol.* 52, 1155–1166. <https://doi.org/10.1016/j.infsof.2010.05.009>.
- Ouadah, A., Zemmouchi-Ghomari, L., Salhi, N., 2022. Selecting an appropriate supervised machine learning algorithm for predictive maintenance. *Int. J. Adv. Manuf. Technol.* 119, 4277–4301. <https://doi.org/10.1007/s00170-021-08551-9>.
- Palansooriya, K.N., Li, J., Dissanayake, P.D., Suvarna, M., Li, L., Yuan, X., Sarkar, B., Tsang, D.C.W., Rinklebe, J., Wang, X., Ok, Y.S., 2022. Prediction of Soil Heavy Metal Immobilization by Biochar Using Machine Learning. *Environ. Sci. Tech. acs.est.1c08302* <https://doi.org/10.1021/acsest.1c08302>.
- Perera, S.M.H.D., Wickramasinghe, C., Samarasiri, B.K.T., Narayana, M., 2021. Modeling of thermochemical conversion of waste biomass – a comprehensive review. *Biofuel Res. J.* 8, 1481–1528. <https://doi.org/10.18331/BRJ2021.8.4.3>.
- Peterson, A.A., Vogel, F., Lachance, R.P., Fröling, M., Antal Jr., M.J., Tester, J.W., 2008. Thermochemical biofuel production in hydrothermal media: A review of sub- and supercritical water technologies. *Energ. Environ. Sci.* 1, 32. <https://doi.org/10.1039/b810100k>.
- Qi, C., Guo, L., Ly, H.B., Le, H.V., Pham, B.T., 2021. Improving pressure drops estimation of fresh cemented paste backfill slurry using a hybrid machine learning method. *Miner. Eng.* 163, 106790 <https://doi.org/10.1016/j.mineng.2021.106790>.
- Ragauskas, A.J., Beckham, G.T., Biddy, M.J., Chandra, R., Chen, F., Davis, M.F., Davison, B.H., Dixon, R.A., Gilna, P., Keller, M., Langan, P., Naskar, A.K., Saddler, J.

- N., Tschaplinski, T.J., Tuskan, G.A., Wyman, C.E., 2014. Lignin Valorization: Improving Lignin Processing in the Biorefinery. *Science* 80-, 344. <https://doi.org/10.1126/science.1246843>.
- Rasam, S., Talebkiekhah, F., Talebkiekhah, M., Salimi, A., Moraveji, M.K., 2021. Physico-chemical properties prediction of hydrochar in macroalgae *Sargassum horneri* hydrothermal carbonisation. *Int. J. Environ. Anal. Chem.* 101, 2297–2318. <https://doi.org/10.1080/03067319.2019.1700973>.
- Ribeiro, M.T., Singh, S., Guestrin, C., 2016. "Why Should I Trust You?" Explaining the Predictions of Any Classifier, in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, New York, NY, USA, pp. 1135–1144. 10.1145/2939672.2939778.
- Seo, M.W., Lee, S.H., Nam, H., Lee, D., Tokmurzin, D., Wang, S., Park, Y.K., 2022. Recent advances of thermochemical conversion processes for biorefinery. *Bioresour. Technol.* 343, 126109 <https://doi.org/10.1016/j.biotech.2021.126109>.
- Serrano, E., Blas, J.G., Carretero, J., Abella, M., Desco, M., 2017. Medical Imaging Processing on a Big Data Platform Using Python: Experiences with Heterogeneous and Homogeneous Architectures. In: 2017 17th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGRID). IEEE, pp. 830–837. <https://doi.org/10.1109/CCGRID.2017.56>.
- Shafizadeh, A., Shahbeig, H., Nadian, M.H., Mobli, H., Dowlati, M., Gupta, V.K., Peng, W., Lam, S.S., Tabatabaei, M., Aghbashlo, M., 2022. Machine learning predicts and optimizes hydrothermal liquefaction of biomass. *Chem. Eng. J.* 445, 136579 <https://doi.org/10.1016/j.cej.2022.136579>.
- Shahbeik, H., Peng, W., Panahi, K.S., Dehhaghi, M., Guillemin, G.J., Fallahi, A., Amiri, H., Rehan, M., Raikwar, D., Latine, H., Pandalone, B., Khoshnevisan, B., Sonne, C., Vaccaro, L., Nizami, A.S., Gupta, V.K., Lam, S.S., Pan, J., Luque, R., Sels, B., Tabatabaei, M., Aghbashlo, M., 2022. Synthesis of liquid biofuels from biomass by hydrothermal gasification: A critical review. *Renew. Sustain. Energy Rev.* 167, 112833 <https://doi.org/10.1016/j.rser.2022.112833>.
- Shenbagaraj, S., Sharma, P.K., Sharma, A.K., Raghav, G., Kota, K.B., Ashokkumar, V., 2021. Gasification of food waste in supercritical water: An innovative synthesis gas composition prediction model based on Artificial Neural Networks. *Int. J. Hydrogen Energy* 46, 12739–12757. <https://doi.org/10.1016/j.ijhydene.2021.01.122>.
- Shi, N., Liu, Q., He, X., Wang, G., Chen, N., Peng, J., Ma, L., 2019. Molecular Structure and Formation Mechanism of Hydrochar from Hydrothermal Carbonization of Carbohydrates. *Energy Fuel* 33, 9904–9915. <https://doi.org/10.1021/acs.energyfuels.9b02174>.
- Sigmund, G., Gharaso, M., Hüffer, T., Hofmann, T., 2020. Deep Learning Neural Network Approach for Predicting the Sorption of Ionizable and Polar Organic Pollutants to a Wide Range of Carbonaceous Materials. *Environ. Sci. Tech.* 54, 4583–4591. <https://doi.org/10.1021/acs.est.9b06287>.
- Tang, Q., Chen, Y., Yang, H., Liu, M., Xiao, H., Wang, S., Chen, H., Raza Naqvi, S., 2021. Machine learning prediction of pyrolytic gas yield and compositions with feature reduction methods: Effects of pyrolysis conditions and biomass characteristics. *Bioresour. Technol.* 339, 125581 <https://doi.org/10.1016/j.biotech.2021.125581>.
- Tuck, C.O., Pérez, E., Horváth, I.T., Sheldon, R.A., Poliakoff, M., 2012. Valorization of Biomass: Deriving More Value from Waste. *Science* (80-) 337, 695–699. <https://doi.org/10.1126/science.1218930>.
- Ullah, Z., Khan, M., Raza Naqvi, S., Farooq, W., Yang, H., Wang, S., Vo, D.V.N., 2021. A comparative study of machine learning methods for bio-oil yield prediction – A genetic algorithm-based features selection. *Bioresour. Technol.* 335, 125292 <https://doi.org/10.1016/j.biotech.2021.125292>.
- Vassilev, S.V., Baxter, D., Andersen, L.K., Vassileva, C.G., 2013. An overview of the composition and application of biomass ash. Part 2. Potential utilisation, technological and ecological advantages and challenges. *Fuel* 105, 19–39. <https://doi.org/10.1016/j.fuel.2012.10.001>.
- Velvizhi, G., Balakumar, K., Shetti, N.P., Ahmad, E., Kishore Pant, K., Aminabhavi, T.M., 2022. Integrated biorefinery processes for conversion of lignocellulosic biomass to value added materials: Paving a path towards circular economy. *Bioresour. Technol.* 343, 126151 <https://doi.org/10.1016/j.biotech.2021.126151>.
- Wang, Z., Song, C., Chen, T., 2017. Deep learning based monitoring of furnace combustion state and measurement of heat release rate. *Energy* 131, 106–112. <https://doi.org/10.1016/j.energy.2017.05.012>.
- Wang, Z., Peng, X., Xia, A., Shah, A.A., Huang, Y., Zhu, X., Zhu, X., Liao, Q., 2021. The role of machine learning to boost the bioenergy and biofuels conversion. *Bioresour. Technol.* 343, 126099 <https://doi.org/10.1016/j.biotech.2021.126099>.
- Watson, J., Wang, T., Si, B., Chen, W.T., Aierzhati, A., Zhang, Y., 2020. Valorization of hydrothermal liquefaction aqueous phase: pathways towards commercial viability. *Prog. Energy Combust. Sci.* 77, 100819 <https://doi.org/10.1016/j.pecsc.2019.100819>.
- Wu, Z., Rincon, D., Luo, J., Christofides, P.D., 2021. Machine learning modeling and predictive control of nonlinear processes using noisy data. *AIChE J.* 67, 19–22. <https://doi.org/10.1002/aic.17164>.
- Xiang, W., Zhang, X., Chen, J., Zou, W., He, F., Hu, X., Tsang, D.C.W., Ok, Y.S., Gao, B., 2020. Biochar technology in wastewater treatment: A critical review. *Chemosphere* 252, 126539. <https://doi.org/10.1016/j.chemosphere.2020.126539>.
- Xiu, S., Shahbazi, A., 2012. Bio-oil production and upgrading research: A review. *Renew. Sustain. Energy Rev.* 16, 4406–4414. <https://doi.org/10.1016/j.rser.2012.04.028>.
- Xu, S., Chen, J., Peng, H., Leng, S., Li, H., Qu, W., Hu, Y., Li, H., Jiang, S., Zhou, W., Leng, L., 2021. Effect of biomass type and pyrolysis temperature on nitrogen in biochar, and the comparison with hydrochar. *Fuel* 291, 120128. <https://doi.org/10.1016/j.fuel.2021.120128>.
- Xu, Z.-X., Ma, X.-Q., Zhou, J., Duan, P.-G., Zhou, W.-Y., Ahmad, A., Luque, R., 2022. The influence of key reactions during hydrothermal carbonization of sewage sludge on aqueous phase properties: A review. *J. Anal. Appl. Pyrol.* 167, 105678 <https://doi.org/10.1016/j.jaap.2022.105678>.
- Yan (Harry), Y., Borhani, T., Subraveti, G., Pai, N., Prasad, V., Rajendran, A., Nkulikiyinka, P., Asibor, J.O., Zhang, Z., Shao, D., Wang, L., Zhang, W., Yan, Y., Ampomah, W., You, J., Wang, M., Anthony, E., Manovic, V., Clough, P.T., 2021. Harnessing the power of machine learning for carbon capture, utilisation, and storage (CCUS) – A state-of-the-art review. *Energ. Environ. Sci.* <https://doi.org/10.1039/D1EE02395K>.
- Yang, L., Shami, A., 2020. On hyperparameter optimization of machine learning algorithms: Theory and practice. *Neurocomputing* 415, 295–316. <https://doi.org/10.1016/j.neucom.2020.07.061>.
- Yin, S., Mehrotra, A.K., Tan, Z., 2011. Alkaline hydrothermal conversion of cellulose to bio-oil: Influence of alkalinity on reaction pathway change. *Bioresour. Technol.* 102, 6605–6610. <https://doi.org/10.1016/j.biotech.2011.03.069>.
- Yu, S., Dong, X., Zhao, P., Luo, Z., Sun, Z., Yang, X., Li, Q., Wang, L., Zhang, Y., Zhou, H., 2022. Decoupled temperature and pressure hydrothermal synthesis of carbon sub-micron spheres from cellulose. *Nat. Commun.* 13, 3616. <https://doi.org/10.1038/s41467-022-31352-x>.
- Zhai, Y., Peng, C., Xu, B., Wang, T., Li, C., Zeng, G., Zhu, Y., 2017. Hydrothermal carbonisation of sewage sludge for char production with different waste biomass: Effects of reaction temperature and energy recycling. *Energy* 127, 167–174. <https://doi.org/10.1016/j.energy.2017.03.116>.
- Zhang, W., Li, J., Liu, T., Leng, S., Yang, L., Peng, H., Jiang, S., Zhou, W., Leng, L., Li, H., 2021. Machine learning prediction and optimization of bio-oil production from hydrothermal liquefaction of algae. *Bioresour. Technol.* 342, 126011 <https://doi.org/10.1016/j.biotech.2021.126011>.
- Zhang, Y., Li, L., Ren, Z., Yu, Y., Li, Y., Pan, J., Lu, Y., Feng, L., Zhang, W., Han, Y., 2022. Plant-scale biogas production prediction based on multiple hybrid machine learning technique. *Bioresour. Technol.* 363, 127899 <https://doi.org/10.1016/j.biotech.2022.127899>.
- Zhang, L., Xu (Charles), C., Champagne, P., 2010. Energy recovery from secondary pulp/paper-mill sludge and sewage sludge with supercritical water treatment. *Bioresour. Technol.* 101, 2713–2721. <https://doi.org/10.1016/j.biotech.2009.11.106>.
- Zhao, S., Li, J., Chen, C., Yan, B., Tao, J., Chen, G., 2021. Interpretable machine learning for predicting and evaluating hydrogen production via supercritical water gasification of biomass. *J. Clean. Prod.* 316, 128244 <https://doi.org/10.1016/j.jclepro.2021.128244>.
- Zhou, X., Zhao, J., Chen, M., Wu, S., Zhao, G., Xu, S., 2022. Effects of hydration parameters on chemical properties of biocrudes based on machine learning and experiments. *Bioresour. Technol.* 350, 126923 <https://doi.org/10.1016/j.biotech.2022.126923>.
- Zhu, L.T., Chen, X.Z., Ouyang, B., Yan, W.C., Lei, H., Chen, Z., Luo, Z.H., 2022. Review of Machine Learning for Hydrodynamics, Transport, and Reactions in Multiphase Flows and Reactors. *Ind. Eng. Chem. Res.* 61, 9901–9949. <https://doi.org/10.1021/acs.iecr.2c01036>.
- Zhu, X., Wang, X., Ok, Y.S., 2019. The application of machine learning methods for prediction of metal sorption onto biochars. *J. Hazard. Mater.* 378, 120727 <https://doi.org/10.1016/j.jhazmat.2019.06.004>.
- Zhu, X., Wan, Z., Tsang, D.C.W., He, M., Hou, D., Su, Z., Shang, J., 2021. Machine learning for the selection of carbon-based materials for tetracycline and sulfamethoxazole adsorption. *Chem. Eng. J.* 406, 126782 <https://doi.org/10.1016/j.cej.2020.126782>.