Biofuel Research Journal 37 (2023) 1786-1809

BRTeam BIOFUEL RESEARCH TEAM

Biofuel Research Journal

Journal homepage: www.biofueljournal.com

Review Paper

Machine-learning-aided thermochemical treatment of biomass: a review

Hailong Li[1], Jiefeng Chen[1], Weijin Zhang[1], Hao Zhan[1], Chao He[2], Zequn Yang[1], Haoyi Peng[1], Lijian Leng[1,*]

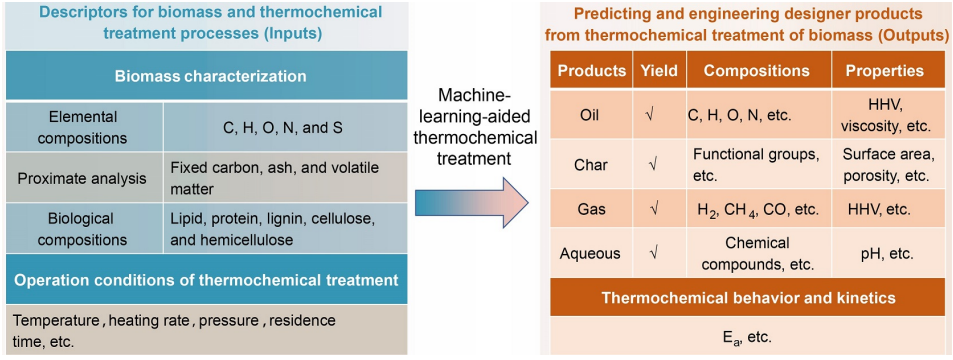
[1]*School of Energy Science and Engineering, Central South University, Changsha, Hunan 410083, People's Republic of China*.

[2]*Faculty of Engineering and Natural Sciences, Tampere University, Tampere, Finland.*

HIGHLIGHTS

- Machine learning (ML) schemes and descriptors of input and output features were examined.
- ML can predict yields, compositions, and properties of oil/char/gas/aqueous phases.
- ML predictive performance for wet and dry thermochemical processes is different.
- Improving ML predictive performance, generalizability, and application are future needs.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 17 November 2022
Received in revised form 11 January 2023
Accepted 11 February 2023
Published 1 March 2023

Keywords:
Artificial intelligence
Pyrolysis
Hydrothermal conversion
Bio-oil
Char
Syngas

ABSTRACT

Thermochemical treatment is a promising technique for biomass disposal and valorization. Recently, machine learning (ML) has been extensively used to predict yields, compositions, and properties of biochar, bio-oil, syngas, and aqueous phases produced by the thermochemical treatment of biomass. ML demonstrates great potential to aid the development of thermochemical processes. The present review aims to 1) introduce the ML schemes and strategies as well as descriptors of the input and output features in thermochemical processes; 2) summarize and compare the up-to-date research in both ML-aided wet (hydrothermal carbonization/liquefaction/gasification) and dry (torrefaction/pyrolysis/gasification) thermochemical treatment of biomass (i.e., predicting the yields, compositions, and properties of oil/char/gas/aqueous phases as well as thermal conversion behavior or kinetics); and 3) identify the gaps and provide guidance for future studies concerning how to improve predictive performance, increase generalizability, aid mechanistic and application studies, and effectively share data and models in the community. The development of biomass thermochemical treatment processes is envisaged to be greatly accelerated by ML in the near future.

* Corresponding author at:
E-mail address: lljchs@126.com

Please cite this article as: Li H., Chen J., Zhang W., Zhan H., He C., Yang Z., Peng H., Leng L. Machine-learning-aided thermochemical treatment of biomass: a review. Biofuel Research Journal 37 (2023) 1786-1809. DOI: 10.18331/BRJ2023.10.1.4

Contents

Abbreviations

AI	Artificial intelligence	ML	Machine learning
ANFIS	Artificial neuro-fuzzy inference system	MLP-NN	Multilayer perceptron neural network
ANN	Artificial neural network	MLR	Multilinear regression
CEC	Cation exchange capacity	PDP	Partial dependence plots
CFD	Computational fluid dynamics	PSO	Particle swarm optimization
COD	Chemical oxygen demand	R^2	Coefficient of determination
DAF	Dry-ash-free	RF	Random forest
DT	Decision tree	RMSE	Root mean square error
GBR	Gradient boosting regression	SCWG	Supercritical water gasification
GUI	Graphical user interface	SHAP	Shapley additive explanations analysis
HHV	Higher heating value	SSA	Specific surface area
HTC	Hydrothermal carbonization	SVM	Support vector machine
HTL	Hydrothermal liquefaction	TOC	Total organic carbon
LCA	Life cycle assessment	TN	Total nitrogen
LHV	Lower heating value	TP	Total phosphorus

1. Introduction

Biomass is a kind of renewable energy with huge reserves, possessing environment-friendly and carbon-neutral properties. Using biomass is one of the effective ways to alleviate future energy needs while mitigating the greenhouse effect and environmental pollution crises (Leng et al., 2019b; Wang et al., 2021a; Chen et al., 2022c). Thermochemical processes, such as torrefaction, pyrolysis, gasification, and hydrothermal carbonization/liquefaction/gasification, are important methods to valorize biomass or biowastes (carbon-neutral materials), such as forestry waste (wood, woody biomass, etc.), agricultural wastes (straw, husks, grasses, etc.), sewage sludge, animal manure, food waste, and algae (Peterson et al., 2008; Wang et al., 2017; Leng et al., 2021b; Perera et al., 2021; Xu et al., 2022). During the thermochemical treatment of biomass, valuable oil and gas products can be generated, which can be used for the production of biofuels and biochemicals to replace fossil resources (Tuck et al., 2012; Ragauskas et al., 2014; Leng et al., 2023). In addition, versatile carbonaceous materials are also generated, which can be used to replace fossil resources through the production of char fuel (Liu et al., 2015; Leng et al., 2021a, 2020b; Li et al., 2023) while char carbon (black carbon) could also be used in soli for sequestrating carbon and mitigating climate change (Leng et al., 2019b and c). Many variables affect the thermochemical treatment performance, and investigations on the thermochemical reaction behavior and mechanisms, as well as the optimization of the as-produced gas/oil/char products by conventional experimental methods, are time-consuming and labor-intensive, which can be overcome by incorporating machine learning (ML) technology.

The term “artificial intelligence (AI)” was introduced by the American scientist John McCarthy at a conference at Dartmouth College in 1956. However, AI was not very popular until the end of the 20th century and the beginning of the 21st century, when computer science and AI algorithms were fully developed (Xu et al., 2021b). Currently, AI, particularly ML, has been developed and widely used in various areas, including the recognition of speech and visual objects, detection of objects, and prediction of the yields, compositions, and properties of products, as well as the reaction

Please cite this article as: Li H., Chen J., Zhang W., Zhan H., He C., Yang Z., Peng H., Leng L. Machine-learning-aided thermochemical treatment of biomass: a review. Biofuel Research Journal 37 (2023) 1786-1809. DOI: 10.18331/BRJ2023.10.1.4

behavior of chemical reactions, e.g., thermochemical reactions. ML is believed to be as popular as, if not more popular than, thermodynamic equilibrium, kinetics, and computational fluid dynamics (CFD) to model highly complex processes efficiently (Ascher et al., 2022b). For example, ML reduced the computational expense of detailed kinetic models by four orders of magnitude and predicted the outputs of the detailed kinetic models with very high accuracy for new data (Hough et al., 2017). However, ML has only recently been widely applied to aid the thermochemical treatment of biomass.

The number of studies available in the existing literature, journal distribution, and funding agencies focused on the ML-aided thermochemical treatment of biomass based on the dataset of the Web of Science are shown in **Figure 1** and **Table 1**. The number of published articles increased over time from 2013 to 2022. Bioresource Technology, with 28 published studies, makes up 14.43% of all related publications, making it the most popular journal in the field of ML-aided thermochemical treatment of biomass. The National Natural Science Foundation of China, the National Key Research and Development Program of China, and the National Research Foundation of Korea are the top funding agencies for scientific research projects in this field (**Table 1**). Before 2018, only a few studies were published on the thermochemical treatment of biomass using ML (**Fig. 1**). However, in the last three years (2020–2022), extensive research has been conducted to predict the yields, compositions, and properties of char, oil, gas, and aqueous phases from the thermochemical treatment of biomass. Some researchers have published reviews exclusive to individual thermochemical technologies (**Table 2**). ML and statistical approaches for biomass torrefaction have been reviewed (Manatura et al., 2023). In another review published earlier, the applications of AI-based modeling for bioenergy systems, including thermochemical treatment processes, were reviewed, covering the articles published between 2005

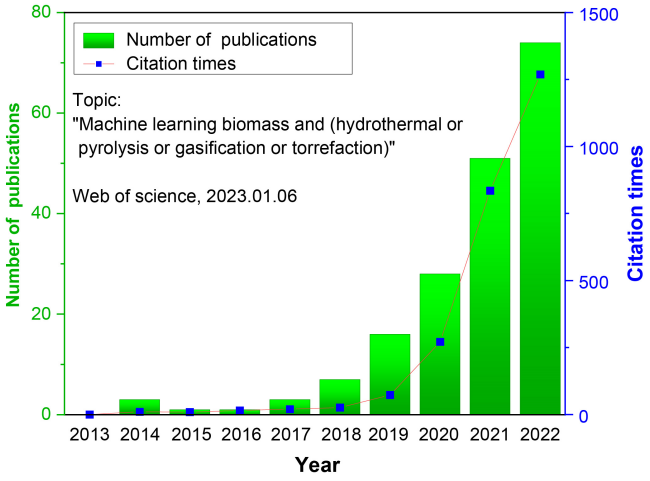


Fig. 1. The number of published documents on ML-aided thermochemical treatment of biomass from 2013 to 2022.

Table 1.
Top 10 journals with the largest number of publications and top 10 funding agencies of publications on ML-aided biomass thermochemical treatment.

Rank	Top 10 Journals			Top 10 Funding Agencies		
	Journal Name	Number of Publications	Contribution (%)	Institution	Number of Publications	Contribution (%)
1	Bioresource Technology	28	14.43	National Natural Science Foundation of China	35	18.04
2	Fuel	20	10.31	National Key Research and Development Program of China	10	5.16
3	Energy	12	6.19	National Research Foundation of Korea	7	3.61
4	Renewable Energy	10	5.16	Chiang Mai University	5	2.58
5	Energy Conversion and Management	9	4.63	Ministry of Science, ICT and Future Planning (Republic of Korea)	5	2.58
6	Chemical Engineering Journal	7	3.61	Fundamental Research Funds for the Central Universities	4	2.06
7	Applied Energy	6	3.09	Higher Institution Centre of Excellence (HICoE), Institute of Tropical Aquaculture and Fisheries (AKUATROP) (Ministry of Higher Education Malaysia)	4	2.06
8	International Journal of Hydrogen Energy	6	3.09	Ministry of Science and ICT (Republic of Korea)	4	2.06
9	Journal of Cleaner Production	5	2.60	National Research Council of Thailand (NRCT)	4	2.06
10	Biomass Conversion and Biorefinery	4	2.06	National Science Foundation (NSF)	4	2.06

Table 2.
Comparison of the present review and published reviews on ML-aided thermochemical treatment of biomass.

Dry Thermochemical Treatment			Wet thermochemical treatment			Reference
Torrefaction	Slow/Fast Pyrolysis	Gasification	Hydrothermal Carbonization (HTC)	Hydrothermal Liquefaction (HTL)	Supercritical Water Gasification (SCWG)	
√	-	-	-	-	-	Manatura et al. (2023)
-	√	√	-	-	-	Liao and Yao (2021)
-	-	-	√	√	√	Li et al. (2022a)
-	-	√	-	-	√	Umenweke et al. (2022)
-	-	-	√	√	√	Zhang et al. (2023)
√	√	√	√	√	√	**This Review**

Please cite this article as: Li H., Chen J., Zhang W., Zhan H., He C., Yang Z., Peng H., Leng L. Machine-learning-aided thermochemical treatment of biomass: a review. Biofuel Research Journal 37 (2023) 1786-1809. DOI: 10.18331/BRJ2023.10.1.4

and 2019, but only a limited number of cases related to thermochemical treatment processes were included, mainly cases of biomass pyrolysis and gasification (Liao and Yao, 2021). Several other reviews have been recently published concerning either ML-aided hydrothermal treatment (Li et al., 2022a; Umenweke et al., 2022; Zhang et al., 2023) or gasification/pyrolysis of biomass (Ascher et al., 2022a). However, to the best of our knowledge, no review has been published that covers and compares the ML-aided wet and dry thermochemical treatments of biomass. In addition, the input and output features of ML models and the predictive performance and interpretation of the studied ML models have not been systematically reviewed.

The present review aims to summarize and compare the up-to-date research on both ML-aided wet and dry thermochemical treatments of biomass and provide guidance for future studies. In the second section of the review, biomass and wet/dry thermochemical processes are introduced, with particular emphasis on the descriptors of the variables used in previous ML studies. The third section summarizes the ML schemes and popular algorithms used in this area. In the fourth section, the application of ML for predicting the yields, compositions, and properties (structural characteristics were also included as properties) of products from wet and dry thermochemical treatments of biomass are compared and discussed. Finally, challenges and strategies to bridge them are provided before the conclusions are presented.

2. Thermochemical technologies for biomass treatment

2.1. Biomass characterization

Biomass is mainly composed of organic matter; the yield of biomass volatile matter is generally more than 50%, and the major elements in biomass, presented in order of weight percentage from high to low, are C, O, H, and N (Vassilev et al., 2010; Leng et al., 2021c). The variation in the ash yield of biomass is significant, ranging from near 0 to 50% or even higher, and the composition of ash is even more complex, constituting dozens of inorganic components, if not hundreds (Vassilev et al., 2012 and 2013). The remainder, moisture ignored, is fixed carbon, which is less abundant than volatile matter and ash (Vassilev et al., 2010; Leng et al., 2021c).

Biomass types are commonly classified according to their biological compositions. Lignin, cellulose, and hemicellulose are the major components of most traditional lignocellulosic biomasses, namely forestry and agricultural biomasses, including wood and woody biomass, husks, straw, and grasses, and the contents of these three components vary depending on the specific biomass type (Vassilev et al., 2010). For lignocellulosic biomass, other biological components are generally not characterized, and extractives are determined by subsequent leaching with various solvents (Vassilev et al., 2012). For biomass resources originating from microorganisms and animals, as well as fruit wastes, lipids, proteins, and carbohydrates are the dominant components (Ge et al., 2021). In algal biomass (microalgae and macroalgae), sewage sludge, and food waste, non-fibrous carbohydrates also exist (Vassilev et al., 2010 and 2012).

From a biochemical perspective, lignin is a polymer of phenolic units (p-hydroxyphenyl, guaiacyl, and syringyl), cellulose is a polymer of glucose, hemicellulose is made of various sugars (including glucose, xylose, and mannose), and proteins are polymers of 20 α-amino acids (Yang et al., 2022b). By contrast, lipids are composed of triglycerides, whose building block, i.e., fatty acids, can vary significantly depending on the lipid sources (Sawangkeaw and Ngamprasertsith, 2013; Leng et al., 2020a; Leng et al., 2023). Generally, lipids with relatively short carbon chains and highly saturated fatty acids are fat, more commonly found in biomass of animal origin, whereas lipids with long carbon chains and low-saturation levels are oil, more commonly seen in the biomass of plant and microbial origin (Sawangkeaw and Ngamprasertsith, 2013; Leng et al., 2020a). During thermochemical treatment, lignin tends to yield biochar, and lipids yield bio-oil, while the other biochemical components contribute differently to char, oil, gas, and aqueous phases, depending on the processing conditions (Leng et al., 2021b). It should be noted that both elemental and biological compositions can be expressed either on a dry basis or a dry-ash-free (DAF) basis. The numerical and category descriptors for biomass are listed in **Table 3**.

The descriptors for biomass are related to each other, and some biomass descriptors can be predicted by other biomass descriptors using ML. For example, the elemental composition of biomass can be predicted by proximate analysis (Ghugare et al., 2017; Olatunji et al., 2019) or infrared spectroscopy (Tao et al., 2020), and the cellulose, hemicellulose, and lignin content of biomass can be predicted by ultimate analysis (Xing et al., 2019; Kartal and Özveren, 2021). Although biomass, as a material, can be characterized by many other structure and property descriptors, such as mechanical properties, these descriptors were seldom used as variables for ML. The main reason is that only a few studies have reported such descriptors.

2.2. Dry thermochemical treatment

Dry thermochemical treatment processes mainly refer to torrefaction, slow/fast pyrolysis, and gasification generally used to process biomass with low moisture content, such as lignocellulosic biomass. These three processes are differentiated by the process parameters (mainly the temperature and heating rate) and their major products. Torrefaction is a process that operates at lower temperatures and heating rates of < 300 °C and < 20 °C/min, respectively (Dai et al., 2019), and is generally used to pretreat biomass for combustion or facilitate the following processes to yield a solid product (da Silva et al., 2018). Slow pyrolysis also occurs at low heating rates, but the temperatures are generally higher than those of torrefaction, i.e., 300–700 °C, with biochar as the dominant product (Liu et al., 2015). Owing to the low heating rates and mild reaction conditions, the residence time in torrefaction and slow pyrolysis ranges from several hours to days. Similarly, fast pyrolysis is generally processed at 300–700 °C but at much higher heating rates, i.e., ranging from 100 °C/min to more than 1000 °C/s, with bio-oil as the major product (Wang et al., 2017). Gasification is a process that involves high heating rates and high working temperatures, ranging from 800 to more than 1000 °C, with syngas being the dominant product (Molino et al., 2016). Owing to the high heating rates and strong reaction conditions, the residence time in fast pyrolysis and gasification is generally within several seconds.

Apart from the temperature, heating rate, and residence times mentioned above, other process parameters can be used to describe these processes and may have considerable effects on biomass thermal treatment performance. For example, the purge gas for pyrolysis is normally inert, and an oxygen-deficient atmosphere is used in commercial facilities. In torrefaction, an oxidative atmosphere such as air is also commonly used to enhance the energy density of torrefaction char (da Silva et al., 2018). While reactive atmospheres, such as air, water, and O_2, are required during gasification to enhance the efficient breaking of chemical bonds. When microwave pyrolysis is used, the microwave power replaces the temperature and acts as the dominant descriptor of this process (Mari Selvam and Balasubramanian, 2022). Other important descriptors include particle size, purge gas flow rate, reactor type and characteristics (e.g., bed materials, heating source), and catalyst. The descriptors for the dry thermochemical treatment processes are shown in **Table 3**.

2.3. Wet thermochemical treatment

Wet thermochemical treatment processes are generally used to treat biomass with high moisture content, such as algae, sludge, manure, and food waste, so that the biomass moisture can be used as a reaction solvent, thus removing the need for energy-intensive pre-drying. The wet thermochemical treatment processes include i) hydrothermal carbonization (HTC), also called wet torrefaction, which works at low temperatures (180–260 °C) and low pressures (2–5 MPa), and the main product is hydrochar (also called biochar) (Peng et al., 2017; Zhai et al., 2017); ii) hydrothermal liquefaction (HTL), which is conducted at temperatures of 250–400 °C and pressures of 5–20 MPa, with bio-oil being the dominant product (Huang et al., 2013; Huang and Yuan, 2015); and iii) supercritical water gasification (SCWG), or hydrothermal gasification, which processes biomass at temperatures of 380–650 °C and pressures of 20–40 MPa to produce syngas rich in H_2, CO, and CH_4 (He et al., 2014; Su et al., 2015).

In addition to temperature and pressure during hydrothermal treatment, other parameters, such as residence time, solid content (solid loading), moisture content (water content), heating rate, reaction solvent, extraction procedure and solvent, and catalyst (shown in **Table 3**), are important descriptors for these processes and may have considerable effects on the hydrothermal treatment. For example, HTC requires a long heating time (several hours to a day) to reach the target temperature, owing to the use of conventional reactors with heating rates lower than 10 °C/min, whereas only several minutes are required for the reaction time of HTL and SCWG,

Please cite this article as: Li H., Chen J., Zhang W., Zhan H., He C., Yang Z., Peng H., Leng L. Machine-learning-aided thermochemical treatment of biomass: a review. Biofuel Research Journal 37 (2023) 1786-1809. DOI: 10.18331/BRJ2023.10.1.4

Table 3.
Descriptors for biomass, thermochemical treatment processes, and products.

Item	Numerical Variables	Category Variables
Biomass	• Elemental composition [a] • Proximate analysis • Biological composition	• Biomass type • Ash composition
Dry thermochemical treatment process	• Temperature • Heating rate • Residence time • Particle size • Purge gas flow rate, etc.	• Purge gas type • Reactor type and reactor parameter descriptors • Heating source • Reaction mode (batch/continuous) • Catalyst, etc.
Wet thermochemical treatment process	• Temperature • Pressure • Residence time • Solid content • Heating rate, etc.	• Reaction solvent • Product extraction procedure • Product extraction solvent • Reaction mode (batch/continuous) • Catalyst, etc.

	Yield	Compositions	Properties
Product (oil phase)	• Yield	• Elemental composition [a] • Chemical composition • Water content • Metal content, etc.	• Various liquid fuel properties (calorific value, carbon/solid residue, viscosity, density, flash point, pour point, acid number, etc.) • Exergy value, pH, etc.
Product (char phase)	• Yield	• Elemental composition [a] • Proximate analysis • Functional group • Total phosphorus • Crystalline phase (inorganic) • Metal content, etc.	• Various solid fuel properties (calorific value, density, etc.) • Various material properties (surface area, total pore volume, micro-/meso-pore volume, average pore size, cation/anion exchange capacity, aromaticity, pH, etc.) • Exergy value
Product (gas phase)	• Yield	• Gas composition (H_2, CH_4, CO, CO_2, C_2H_n, etc.)	• Various gas fuel properties (calorific value, tar content, etc.) • Exergy value
Product (aqueous phase)	• Yield	• Total organic carbon/nitrogen/ phosphorus • Chemical composition • Metal content, etc.	• Various wastewater properties (chemical oxygen demand, pH, etc.)

[a] Elemental composition refers to the contents of C, H, O, N, and S; ash is generally included as a part of the elemental composition when used as input during machine modeling because O is calculated using ash.

although the same heating time is required. However, the heating rate of HTL and SCWG can be up to hundreds of centimeters per minute in fast-heating reactors, which can be achieved by immersing the reactors in a pre-heated sand bath (Akiya and Savage, 2002; Peterson et al., 2008).

In addition, for both wet and dry thermochemical processes, parameters can be integrated and presented as new parameters, e.g., the reaction severity index, which is generally the integration of temperature and residence time by functions (Leng et al., 2021c). Moreover, other thermochemical treatment processes, such as hydrolysis and "thermal-dissolution-based carbon enrichment", have also been proposed that share similar process descriptors as those mentioned above (Hu et al., 2022).

2.4. Product characterization

According to previous reviews, the differences in the compositions and properties of oil/char/gas/aqueous phase products between wet and dry thermochemical treatments are small (Kambo and Dutta, 2015; Leng et al., 2018c and 2021b), and many descriptors for these products from wet and dry processes are the same. As shown in **Table 3**, there are three branches of descriptors for the products from the thermochemical treatment of biomass: yields, compositions, and properties. While the yields of bio-oil, biochar, and gas phases are mainly calculated based on the weight ratio of these products to the weight of the dry (more commonly used) or DAF-based biomass, the yield of the aqueous phase is generally calculated as the carbon recovery rate in the aqueous phase because weighing organics in the aqueous phase is difficult (Leng et al., 2018d).

The properties and compositions of the four phases vary significantly. As an oil for engine use, bio-oil fuel properties such as calorific value (mainly higher heating value, HHV), lower heating value (LHV), carbon/solid residue, viscosity, density, flash point, pour point, and acid number are required and have been reported frequently (Kan et al., 2016; Leng et al., 2018c). Other properties, such as pH, molecular weight, lubricity, and boiling range, have also been used to describe bio-oil (Kanaujia et al., 2014; Kan et al., 2016). These properties depend on the composition. Elemental compositions (C, H, O, N, and S), atomic ratios (O/C, H/C, and (O+N)/C), chemical compounds (relative or absolute contents of various chemicals), and water content are commonly reported. The water content of bio-oil has mainly been reported by references dealing with biomass pyrolysis, and the value is approximately 10–30%, constituting a large part of the bio-oil (Leng et al., 2018c). However, water content has only occasionally been reported for bio-oil from HTL because HTL-bio-oil is typically dewatered during the solvent extraction and separation procedure, and the value is generally much lower if reported (Leng et al., 2018c). N and S in bio-oil can be detrimental because they cause pollution upon combustion, but it can be beneficial if N/S-rich chemicals or materials are targeted (Leng et al., 2020c and 2020e).

Biochar can also be used as a fuel for boilers; thus, calorific values such as HHV and LHV are also generally reported, and many equations are available for their calculation (Chen et al., 2022c). However, using biochar as a carbon material, replacing activated carbon and other materials, is more promising than using it as fuel. Material properties such as specific surface area (SSA), porosity, aromaticity, electrical conductivity, and

Please cite this article as: Li H., Chen J., Zhang W., Zhan H., He C., Yang Z., Peng H., Leng L. Machine-learning-aided thermochemical treatment of biomass: a review. Biofuel Research Journal 37 (2023) 1786-1809. DOI: 10.18331/BRJ2023.10.1.4

cation/anion exchange capacity determine the potential of biochar in various applications (**Table 4**) (Kan et al., 2016; Leng et al., 2021a). For example, SSA and porosity are the most important properties determining biochar's adsorption capacities for organic pollutants from the aqueous phase (Sigmund et al., 2020; Zhang et al., 2020) and for capturing CO_2 or storing H_2 (Maulana Kusdhany and Lyth, 2021; Yuan et al., 2021). Aromaticity is a proxy for biochar stability that determines the carbon sequestration capability of biochar in soil (Leng et al., 2019b and c). Cation exchange capacity (CEC) is the most decisive biochar property for the removal of heavy metals from the aqueous phase (Zhu et al., 2019b) (**Table 4**). The chemical compositions of biochar, such as elemental (C, H, O, N, S, O/C, H/C, and (O+N)/C) and proximate (ash, volatile matter, fixed carbon, and moisture) compositions, functional groups, and crystalline phases, can also be used to determine the application performance of biochar. For example, the ratios of O/C and H/C, as well as the ratio of fixed carbon/volatile matter, can be used to replace aromaticity as biochar stability proxies (Leng et al., 2019b; Chen et al., 2021). Biochar functional groups, such as N/O/S-containing functional groups, are very important for biochar application in pollutant removal, energy storage, catalysis, etc. (Leng et al., 2022a and b, and 2021c). **Table 5** lists the indicators for the application performance of thermochemical products in different areas and the closely related compositions and properties of the char/oil/gas/aqueous phases, which should be investigated in the future.

Gas generated by the thermochemical treatment of biomass is generally considered gas fuel. The chemical components, LHV, and tar content of the gas are the most important factors in its application (Kan et al., 2016). The enrichment of gas with H_2 and CH_4 has received the most interest because these components are the most effective and clean fuel components, and they can be used as fuel or applied to produce various chemicals (Molino et al., 2016). Gases from HTC, HTL, torrefaction, and slow pyrolysis are not frequently collected and analyzed because of the low contents of H_2 and CH_4 but the high content of CO_2 (Leng et al., 2020f).

The aqueous phase, however, is a byproduct of the thermochemical treatment of biomass, with hydrothermal treatment processes producing a large amount, i.e., 1–20 times higher than the dry biomass weight (Leng et al., 2018a, 2020d, and 2021b). Pyrolysis only produces an aqueous phase weight of 10–30% of the mass of dry biomass, and most frequently, it is mixed with bio-oil, which is why the water content in bio-oil from pyrolysis is high. The treatment and valorization of the aqueous phase, especially from hydrothermal treatment, is becoming a challenge for pathways toward commercial viability (Watson et al., 2020). Common wastewater indicators such as pH, chemical oxygen demand (COD), total organic carbon (TOC), and total nitrogen (TN) have been reported in the literature, and they help indicate the wastewater properties and facilitate the matching of suitable technologies to manage the wastewater. Total phosphorus (TP) is mainly reported in the aqueous phase from the thermochemical treatment of P-rich biomass, such as algae, sludge, manure, and food waste (Leng et al., 2019a; Chen et al., 2022b).

Chemical compounds of organics in the aqueous phase have also been characterized, and their compositions are similar to those of bio-oil; however, the contents of polar fractions are higher and nonpolar fractions are lower than those of bio-oil (Leng et al., 2021b). It is worth mentioning that biomass inorganics, such as K and Na, are distributed predominantly in the aqueous phase. In addition, only small fractions of phosphate and heavy metals are distributed in the aqueous phase (Leng et al., 2020d and 2021b), large fractions are in biochar (Leng et al., 2014 and 2018b), and trace amounts are in bio-oil (Leng et al., 2015; Yuan et al., 2015). The yields, compositions, and properties of oil/char/gas/aqueous phases have been reported in recent ML studies using descriptors for biomass and thermochemical processes introduced in *Sections 2.1–2.3* (**Fig. 2**).

3. Machine learning technologies

3.1. Machine learning schemes

A typical ML scheme is shown in **Figure 3**. ML always starts with data collection from references, using tools such as Plot Digitizer or AI tools or data from experiments or open sources, which can be more easily collected. The distribution of each descriptor should be analyzed initially to ensure

Table 4.
Machine-learning-aided predicting application performance of biochar.

Application Area	Target (Output)	Biochar Property (Input, Ranked by Feature Importance) [A]	Predictive Performance, Test Data			Reference
			Algorithms [b]	R^2	RMSE [c]	
Heavy metal removal from the aqueous phase	Heavy metal adsorption capacity	CEC > pH > C > (O+N)/C > ash > H/C > SSA > particle size	RF	0.97	0.057	Zhu et al. (2019b)
		NA [d]	FCM-BPNN	0.99	0.038	Ke et al. (2021a and b)
Heavy metal immobilization in soil	Heavy metal immobilization rate	N > C > O > pH > O/C > H > ash > (O+N)/C > SSA > H/C	RF	0.91	10.54	Palansooriya et al. (2022)
		pH > CEC > (O+N)/C > C/H > THM > SSA > ash	ANN	0.84	NA [d]	Sun et al. (2022)
CO_2 adsorption	CO_2 adsorption capacity	SSA > total pore volume > N > mesopore volume > O > C > H	GBDT	0.84	0.66	Yuan et al. (2021)
	CO_2 adsorption capacity (0 °C, 0.6–1 bar)	Ultra-micropore volume > SSA > mesopore volume > micropore volume > H > O > C >N	RF	0.96	0.266	Zhu et al. (2020)
H_2 adsorption	H_2 adsorption capacity	SSA > O > total pore volume > H > ultra-/micro-pore ratio > C > micropore volume > N	RF	0.91	0.542	Maulana Kusdhany and Lyth (2021)
Pharmaceuticals and personal care products (PPCPs) removal from the aqueous phase	PPCPs adsorption capacity	SSA > (O+N)/C > C > C–O > C=O > non-polar C	RF	0.91	0.166	Zhu et al. (2022a)
Proton-exchange membrane fuel cell polarization curve	Current density and specific power	SSA > N > mesopore ratio > S > micropore ratio	RF	NA	NA	Ding et al. (2020)
Biochar electrode-specific capacitance	Current density	SSA > C/O > C/N	DT	0.98	15.01	Yang et al. (2023)

[a] Only inputs of biochar properties were included; other inputs were not listed. CEC: cation exchange capacity; pH: pH value; SSA: specific surface area; ash: ash yield of biochar; THM: total heavy metal of biochar; C, H, O, N, S, H/C or C/H, O/C or C/O, C/N, and (O+N)/C are the elemental compositions or atomic ratios of biochar; non-polar C (NPC, including C−C/C = C and π–π* transition), C−O (e.g., phenolic, alcoholic, and etheric), and C = O (e.g., carbonyl, quinone, carboxyl, or ester) of biochar, as determined from the deconvoluted peaks in high-resolution C1s X-ray photoelectron spectroscopy.
[b] RF: random forest; FCM–BPNN: fuzzy C-means clustering algorithm (FCM) integrated with back-propagation neural network (BPNN); ANN: artificial neural network; GBDT: gradient boosting decision tree; DT: decision tree.
[c] Root mean square error.
[d] NA: not available.

Please cite this article as: Li H., Chen J., Zhang W., Zhan H., He C., Yang Z., Peng H., Leng L. Machine-learning-aided thermochemical treatment of biomass: a review. Biofuel Research Journal 37 (2023) 1786-1809. DOI: 10.18331/BRJ2023.10.1.4

Table 5.
Indicators for application performance of thermochemical products in different areas and the closely related compositions and properties of the products.

Product	Application Area	Application Performance (and the Closely Related Compositions and Properties of Products)[a]
Char	Carbon sequestration in soil	Char C stability (aromaticity, H/C, O/C, etc.)
	Greenhouse gas mitigation in soil	Methane and nitrous oxide emissions (priming effect capability, N/C, pH, etc.)
	Blast furnace coke for steelmaking	Coke reactivity index (aromaticity, H/C, O/C, etc.) Coke strength after reaction (particle size, etc.)
	Soil amendment	Crop productivity (nutrients contents, toxins, etc.) Water holding capacity (porosity, hydrophobicity) Soil fertility (nutrients, pH, etc.)
	Adsorption of organic pollutants in soil, water, or gas	Adsorption capacity (SSA, porosity, etc.)
	Adsorption of inorganics in soil, water, or gas	Adsorption capacity (CEC, functional groups, etc.)
	Catalytic conversion	Yield of chemicals such as biodiesel (S-functional groups, etc.)
	Additive in anaerobic digestion	Biogas yield or methane productivity (electric conductivity, porosity, nutrients, C/N, toxins, etc.)
	Additive in organic solid waste composting	Degradation degree or microorganism diversity (porosity, nutrients, C/N, toxins, etc.)
	Solid fuel	Slagging tendency (ash compositions, metal contents, etc.)
Oil	Biofuel additives	Cetane number (chemical compositions, etc.) Cold flow property (carbon structure, etc.)
	Bio-asphalt binder	Mechanical properties such as rutting resistance, fatigue performance, dynamic stiffness, and tensile strength (chemical compositions, etc.) Rheological property (viscosity, etc.)
	Jet fuel or kerosene components	Smoke point (chemical compositions, etc.)
	Adhesive additives	Rheological property and others (viscosity, molecular weight distribution, boiling point, etc.)
Gas	Fermentation to various chemicals	Yields of chemicals such as ethanol (toxins, inhibitors, etc.)
	Fischer-Tropsch process	Yields of hydrocarbons (H_2/CO, etc.)
Aqueous	Anaerobic digestion	Biogas yield or methane productivity (BOD, BOD/COD, etc.) Pollutant removal rate (COD, TP, TN, etc.)
	Fermentation to various chemicals	Yields of chemicals such as H_2 (toxins, inhibitors, etc.) Pollutant removal rate (COD, TP, TN, etc.)
	Algae cultivation	Algae productivity (nutrients, toxins, etc.) Pollutant removal rate (COD, TP, TN, etc.)

[a] Compositions or properties of the thermochemical products; H/C, O/C, and C/N are the elemental atomic ratios; SSA: specific surface area; CEC: cation exchange capacity; pH: pH value; COD: chemical oxygen demand; BOD: biological oxygen demand; TP: total phosphorus; TN: total nitrogen

that the collected data are typical, cover the required ranges, and exhibit a favorable normal distribution of the descriptor values. The ML model is built based on the data collected; thus, the applicability and generalization capability of the model is limited to the distribution ranges of the descriptors. When the dataset is ready, data processing is required because different descriptors have varied value scales, and normalization of all descriptors to be within -1 and 1 is generally conducted. Dimensionality reduction is sometimes required when the number of inputs is much higher than the number of outputs or when input descriptors are highly correlated, and principal component, discriminant, and independent component analyses are commonly used for such purposes (Li et al., 2021a).

After the dataset is prepared and processed, it can be used for ML modeling (**Fig. 3**). The first step of modeling is selecting suitable algorithms (see *Section 3.2*) for prediction tasks. Then, the model is trained, and the corresponding hyperparameters are tuned with selected input data, i.e., generally, 70–90% of the data collected. During hyperparameter tuning, cross-validation, which includes hold-out, leave-one-out (Bagheri et al., 2019), rolling-windows analysis (Elmaz and Yücel, 2020), and k-fold methods (Mutlu and Yucel, 2018), allowing more data to be trained and tested, is required to avoid overfitting, thus ensuring accurate prediction of the final optimum ML model (**Fig. 3**). For example, in k-fold cross-validation, the training dataset using the k-1 fold data and tested with the remaining fold; the above procedures are repeated k times to allow all training data to be used for training and validating, which is beneficial for testing with the test dataset (**Fig. 4a**). However, in some studies, cross-validation was not used, for example, ML with test but without cross-validation (Chen et al., 2018; Ozbas et al., 2019) (**Fig. 4c**) and ML with validation and test but without cross-validation (Qasem et al., 2023) (**Fig. 4d**). The average values of the coefficient of determination (R^2) and root mean square error (RMSE) from cross-validation are used to assess the predictive performance, and hyperparameters with the highest fitness (R^2) and/or lowest error (RMSE) as the optimum hyperparameters. Several other statistical criteria can be used to determine ML models' performance, accuracy, and reliability (Umenweke et al., 2022). With the optimum hyperparameters, the data used during hyperparameter tuning (i.e., 70–90% of the data collected) will be input as the training data to retrain the ML model, and the remainder (10–30%) will be used as test data to test the model. However, in some studies, the model was not further tested after cross-validation, and a model test was implemented within the cross-validation process (**Fig. 4b**) (Elmaz et al., 2020; Yapıcı et al., 2022), indicating that the models were not tested with data from out-of-dataset cases. The R^2 and RMSE of such models would be calculated again to evaluate whether the model is acceptable; if not, the

Please cite this article as: Li H., Chen J., Zhang W., Zhan H., He C., Yang Z., Peng H., Leng L. Machine-learning-aided thermochemical treatment of biomass: a review. Biofuel Research Journal 37 (2023) 1786-1809. DOI: 10.18331/BRJ2023.10.1.4

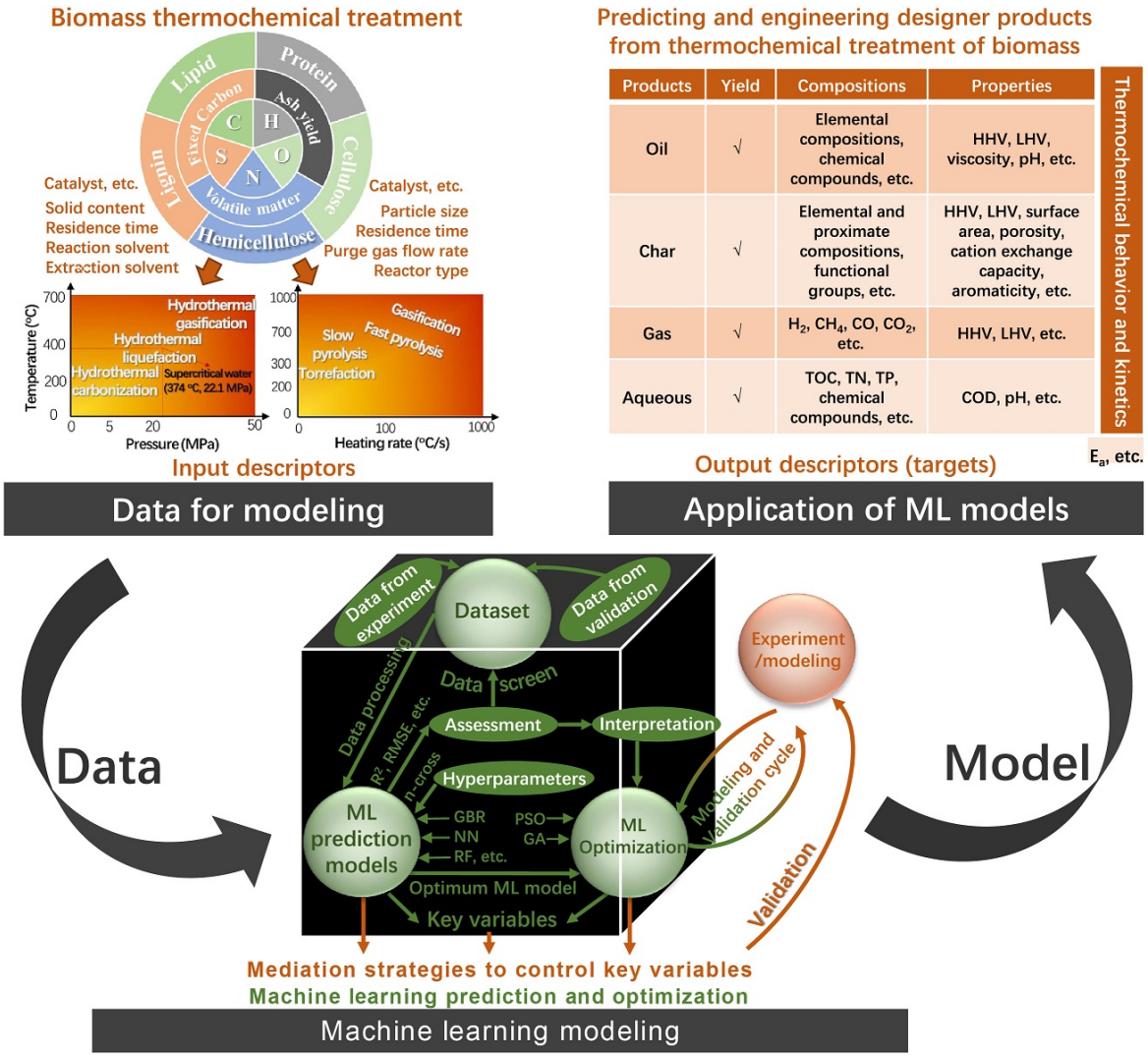


Fig. 2. Machine-learning-aided thermochemical treatment of biomass.

hyperparameters should be optimized again; otherwise, new hyperparameters should be introduced (Fig. 3).

For the optimum ML model, the model interpretation will be studied so that the working mechanisms of the "black box" model can be understood to some extent (Fig. 3). Feature importance analysis, Shapley additive explanations (SHAP) analysis, and partial dependence plots (PDP) are generally used to study the correlations between inputs and outputs. The reasonability of the model may not be favorable, even if it has an acceptable predictive performance. However, the data never lie, and the problem may be that the data collected are biased. The dataset needs to be checked carefully, and the deletion of odd data may solve the interpretation problems, or as much new data as possible should be added to allow for the reconstruction of models with both satisfactory predictive performance and reasonability. For the final optimum models, online or offline platforms can be developed to allow the models to be used by others (Fig. 3). For example, an online graphical user interface (GUI) website or offline software can be developed. With the optimum model, optimization can be conducted to obtain optimal thermochemical operating conditions, as well as biomass characteristics or mixing recipes, that will achieve the targets (generally high yield, high energy recovery, and favorable properties) , which should be the major research direction in the future.

3.2. Machine learning algorithms

Many frameworks and libraries are available for conducting ML modeling, e.g., scikit-learn, TensorFlow, PyTorch, CRAN, Keras, Weka, H2O, mlpack, and those from Amazon, IBM, and Google, among which the free Python-based library scikit-learn seems to be the most popular. The core of these libraries is similar to that of ML algorithms, and the major differences are the different working languages (Python, C++, R, Java, etc.) and operating companies or organizations. ML can be classified into three groups: supervised learning, which makes the machine learn explicitly that data with clearly defined output are required; unsupervised learning, in which the machine learns the data without any defined output; and reinforcement learning, in which the machine learns how to act within a certain environment to maximize the rewards because each learning task returns a reward (Mahmood and Wang, 2021). Supervised learning, mainly used to resolve regression and classification problems, is generally used to predict the yields, compositions, and properties of products obtained from biomass thermochemical treatment.

Many regression and classification algorithms are available, and typical algorithms are described in Figure 5. The accuracy and generalizability of

Please cite this article as: Li H., Chen J., Zhang W., Zhan H., He C., Yang Z., Peng H., Leng L. Machine-learning-aided thermochemical treatment of biomass: a review. Biofuel Research Journal 37 (2023) 1786-1809. DOI: 10.18331/BRJ2023.10.1.4

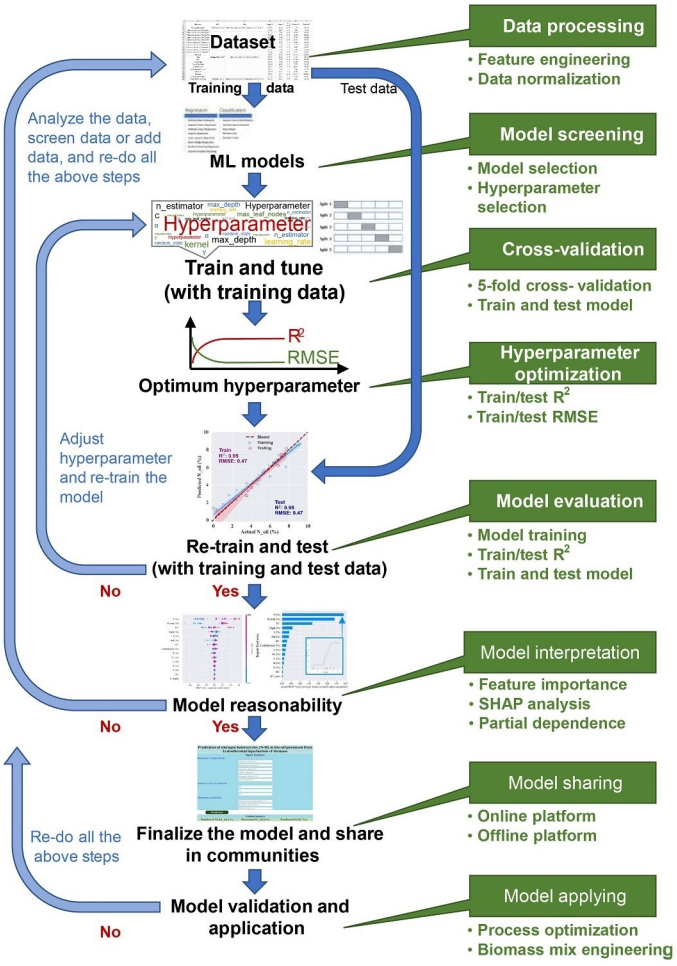


Fig. 3. Typical schemes and strategies for conducting machine learning (ML) studies.

the as-built ML model are dependent on the algorithms, and each algorithm has its own merits and weaknesses. Regression algorithms can be used to build the relationship between dependent and independent variables, and linear regression is the simplest, with an artificial neural network (ANN, **Fig. 5c**) being the most complex. An ANN is a biologically inspired algorithm that imitates the human brain's functionality and consists of an input layer for receiving input variables, one or more hidden layers for identifying nonlinearity and correlating inputs and outputs, and an output layer for representing output variables. ANNs have been widely used in this area owing to their ability to model highly nonlinear processes (Liao et al., 2019; Xia et al., 2021; Khan et al., 2022). However, ANN models are rather complex, and screening network structures, training algorithms, and activation functions are challenging. Some studies have compared ANNs with different modeling skills. For example, comparisons between standard and ordinary activation functions indicate that the former outperforms the latter in predicting hydrogen production (Ayodele et al., 2021). In addition, the ANN has limited interpretability, and few interpretations are included in published references, posing challenges to understanding its working mechanism.

Tree-based models (**Fig. 5d**), such as decision tree (DT) (Quinlan, 1986) and random forest (RF) (Pavlov, 2001), have been increasingly used in this area because of their improved interpretability over ANN and comparable predictive performance. For example, RF is an ensemble model with many DTs that uses bagging or bootstrap aggregation to achieve accurate prediction and improved generalization results compared to single-DT models (**Fig. 5d**) (Pavlov, 2001). Tree-based models are classification algorithms that can be used to identify object categories, and they are favorable for processing non-numerical variables, such as category variables. For example, DT outperformed ANN in predicting hydrogen production (Haq et al., 2022). RF showed better performance than the multilayer perceptron neural network (MLP-NN) in predicting the yields of oil, char, and gas, as well as the compositions of gas and char (Shahbeik et al., 2022). However, there are few studies related to the comparisons between ANN and tree-based models, and the lower predictive performance of ANN may be due to limited tuning strategies. For example, suitable value-assigning methods for non-classification algorithms, such as the one-hot encoding method (Ascher et al., 2022a), may offset this shortage of regression algorithms, such as ANN, and improve the predictive performance.

Other regression and classification algorithms, such as support vector machine (SVM) (Cortes and Vapnik, 1995) (**Fig. 5e**) and gradient boosting regression (GBR) (**Fig. 5f**), have also been increasingly applied to predict yields, compositions, and properties of oil, char, gas, and aqueous phases. In SVM, the training algorithm identifies the separating hyperplane to classify two classes and allows the maximization of the distance between the nearest data points and the hyperplane, which is constructed by support vectors (data points from either class closest to the hyperplane) (**Fig. 5e**). By contrast, the GBR algorithm is trained using a boosting strategy, which establishes the first tree to predict the errors, i.e., variation between the actual and initial values, followed by the calculation of new prediction values under the previous prediction values and predicting new errors from the new tree until there is no obvious decrease in residues (**Fig. 5f**). The predictive performance of these models and ANN or tree-based models varies in different studies. For example, the SVM model had a more satisfactory prediction performance than ANN for predicting the HHV of bio-oil (Chen et al., 2018); GBR outperformed RF in predicting bio-oil yield and elemental compositions (Zhang et al., 2021); the RF model showed better prediction performance than SVM and DT, and multilinear regression (MLR) had the worst performance in predicting bio-oil yield (Ullah et al., 2021). Although the predictive performance depends not only on algorithms but also on dataset characteristics, many researchers have proposed that tree-based models such as RF, SVM, and GBR may be preferable to ANN for small-number dataset problems.

4. Machine-learning-aided thermochemical treatment of biomass

4.1. Prediction of the yields of products

The yields of the four phases of wet and three phases (excluding the aqueous phase) of dry thermochemical processes have been predicted by researchers, with most cases concerning the yields of char and oil, and the predictive performance is shown in **Figure 6a** (see **Supplementary Information**). Most studies used elemental compositions and thermochemical operation conditions as inputs. Proximate analysis was included in most studies dealing with dry thermochemical processes, while atomic ratios and biochemical compositions were also often considered (see **Supplementary Information**). However, only a few studies considered solvents (reaction solvents and extraction solvents for HTL) (Li et al., 2021c), catalysts (Castro Garcia et al., 2022), and metal compositions of biomass (Gu et al., 2021). Even when these variables are included as inputs, biased prediction and interpretation may be encountered if cases including these variables are too few or have a poor distribution (biased dataset) (Li et al., 2021c). In this case, the ML models can be built case-by-case for each particular solvent or catalyst (Zhou et al., 2022b), but models built in this way have low generalizability.

Strictly speaking, the predictive performance between the wet and dry thermochemical processes should not be compared directly because of the differences in datasets, algorithms, etc.; however, the statistical summaries of the predictive performance of all cases from so many references indicate that the ML predictive performance of yields from the dry thermochemical processes is superior to that of the wet ones (**Fig. 6a**). The R^2 values, mainly around 0.85–0.94, for the dry process models are slightly higher than those for the wet ones (R^2 mainly 0.85–0.90) (**Fig. 5a**), while RMSE values for the former (lower than five) are better than those for the latter. Considering that the acceptable errors of biomass thermochemical processes are within 5% (experimental uncertainty), the models obtained for predicting yields of

Please cite this article as: Li H., Chen J., Zhang W., Zhan H., He C., Yang Z., Peng H., Leng L. Machine-learning-aided thermochemical treatment of biomass: a review. Biofuel Research Journal 37 (2023) 1786-1809. DOI: 10.18331/BRJ2023.10.1.4

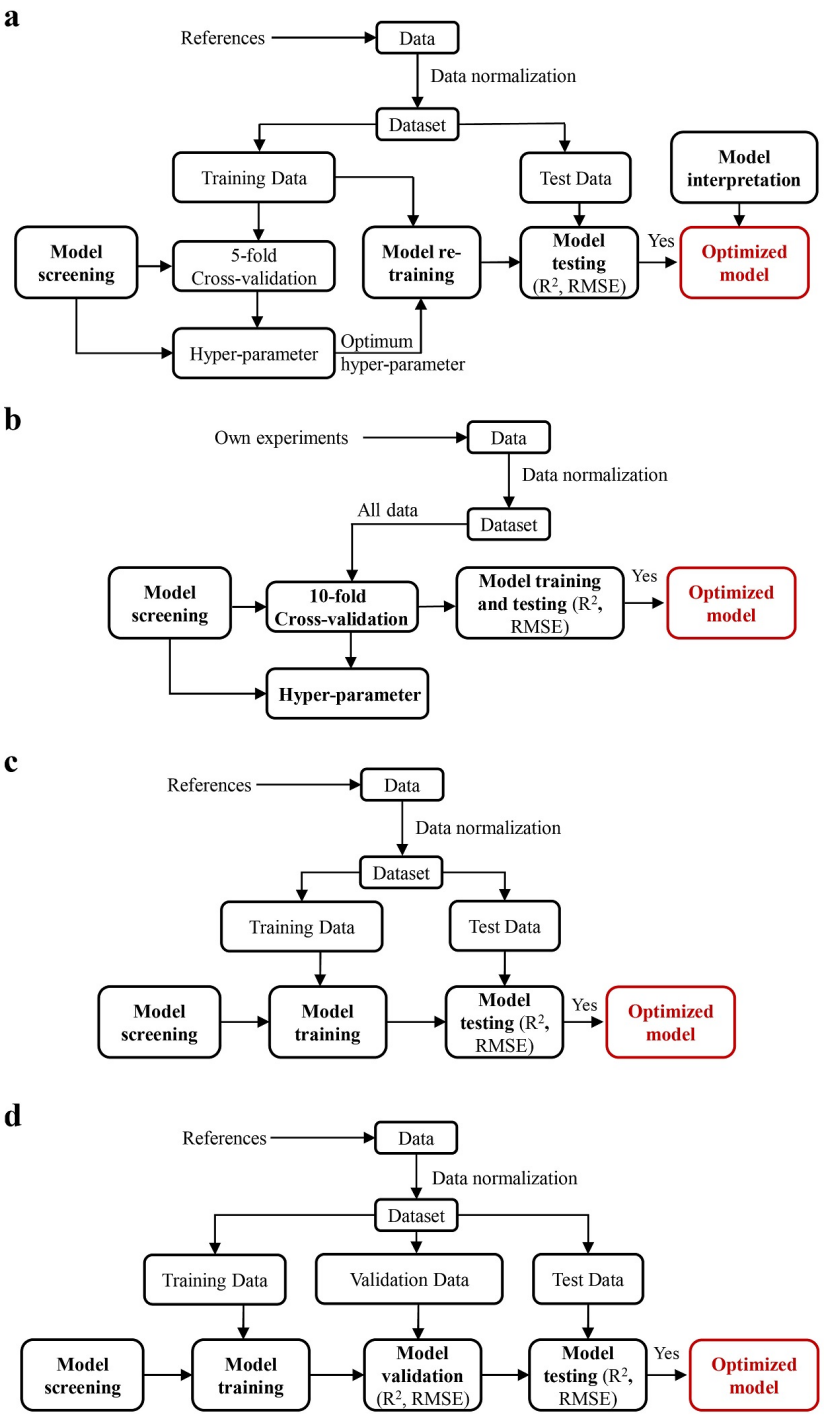


Fig. 4. Different machine learning (ML) schemes reported in the literature: **(a)** ML with cross-validation and additional test (Zhu et al., 2019a; Ullah et al., 2021; Ascher et al., 2022a; Leng et al., 2022c and d); **(b)** ML with cross-validation, but without additional test (Yapıcı et al., 2022); **(c)** ML with test, but without cross-validation (Chen et al., 2018; Ozbas et al., 2019); and **(d)** ML with validation and test, but without cross-validation (Qasem et al., 2023).

products from dry thermochemical processes may be preferable, although an RMSE of 3–8 for the wet processes is also acceptable (**Fig. 6a**).

4.2. Prediction of the compositions of products

Unlike the prediction of the yields of products, the prediction of elemental and proximate compositions of the oil or char, as well as the compositions of the gas phase, in wet thermochemical process models is better than in the dry ones, with R^2 values of approximately 0.90 and RMSE ranging from near 0.5 to 2 for the former, and R^2 values of approximately 0.85 and RMSE ranging from near 0.2 to 4 for the latter (**Fig. 6b**). Only one study considered a catalyst as an input for the prediction of compositions (Li et al., 2021a) among the cases shown in **Figure 6b**, with others including thermochemical operation conditions and biomass compositions (elemental

Please cite this article as: Li H., Chen J., Zhang W., Zhan H., He C., Yang Z., Peng H., Leng L. Machine-learning-aided thermochemical treatment of biomass: a review. Biofuel Research Journal 37 (2023) 1786-1809. DOI: 10.18331/BRJ2023.10.1.4

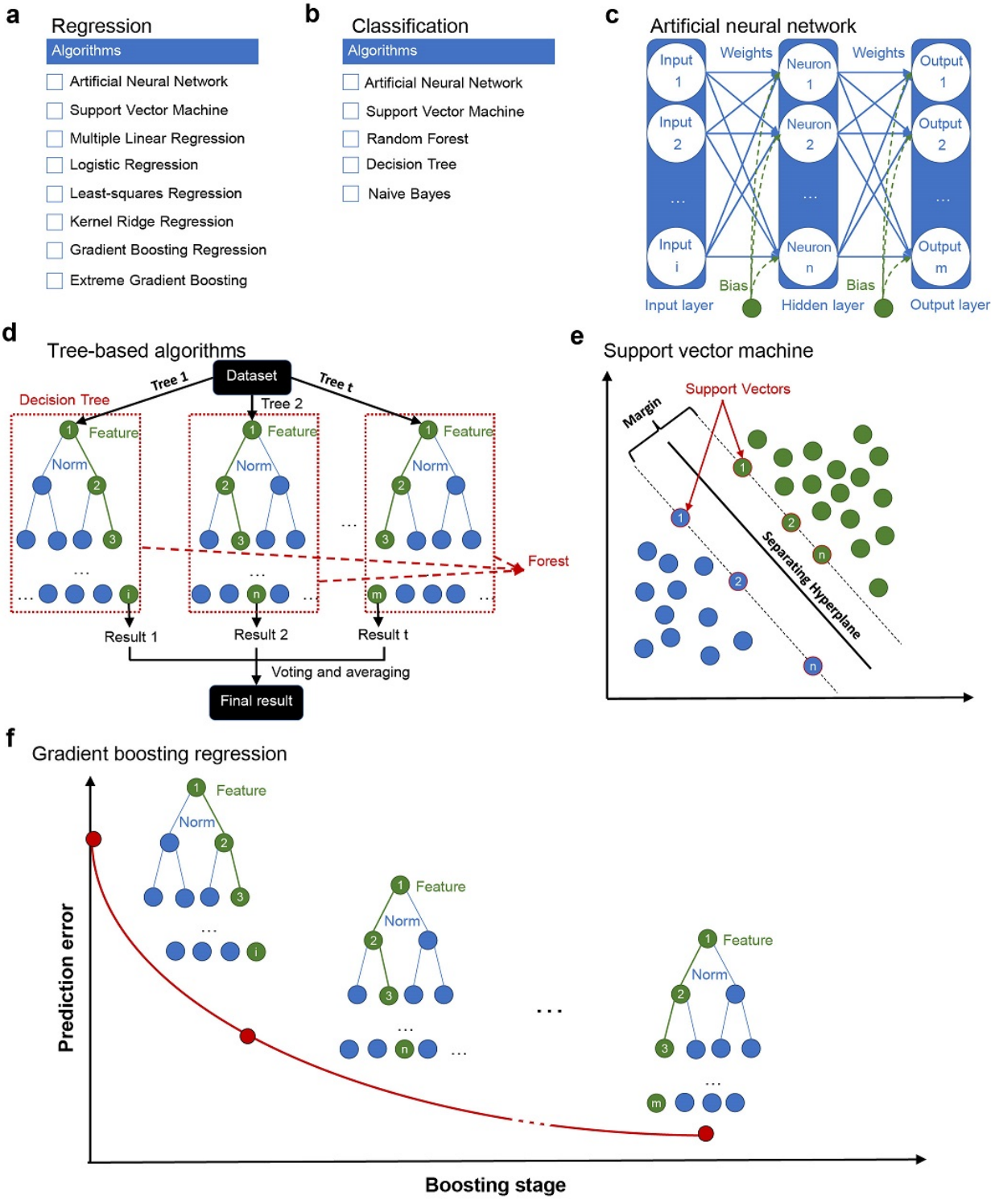


Fig. 5. Typical machine learning algorithms.

compositions/atomic ratio/biochemical composition, see **Supplementary Information**). There is also a case that used only thermochemical parameters as inputs for predicting syngas compositions in SCWG of food wastes with 40 data points by ANN (Shenbagaraj et al., 2021), which seems unreliable with low generalizability. In addition to the basic compositions mentioned above, the chemical compositions of bio-oil, such as N-heterocycles (Leng et al., 2022c), TP in char (Djandja et al., 2022), and TN, TP, and TOC in the aqueous phase (Leng et al., 2022d) have been predicted with acceptable predictive performance (**Table 6**). The yield of glucose from wet torrefaction of microalgae and sorghum distillery residue using H_2SO_4 as a catalyst was also predicted (Chen et al., 2022e). To enhance the predictive performance of N-heterocycles in bio-oil , the yield of bio-oil and the

Please cite this article as: Li H., Chen J., Zhang W., Zhan H., He C., Yang Z., Peng H., Leng L. Machine-learning-aided thermochemical treatment of biomass: a review. Biofuel Research Journal 37 (2023) 1786-1809. DOI: 10.18331/BRJ2023.10.1.4

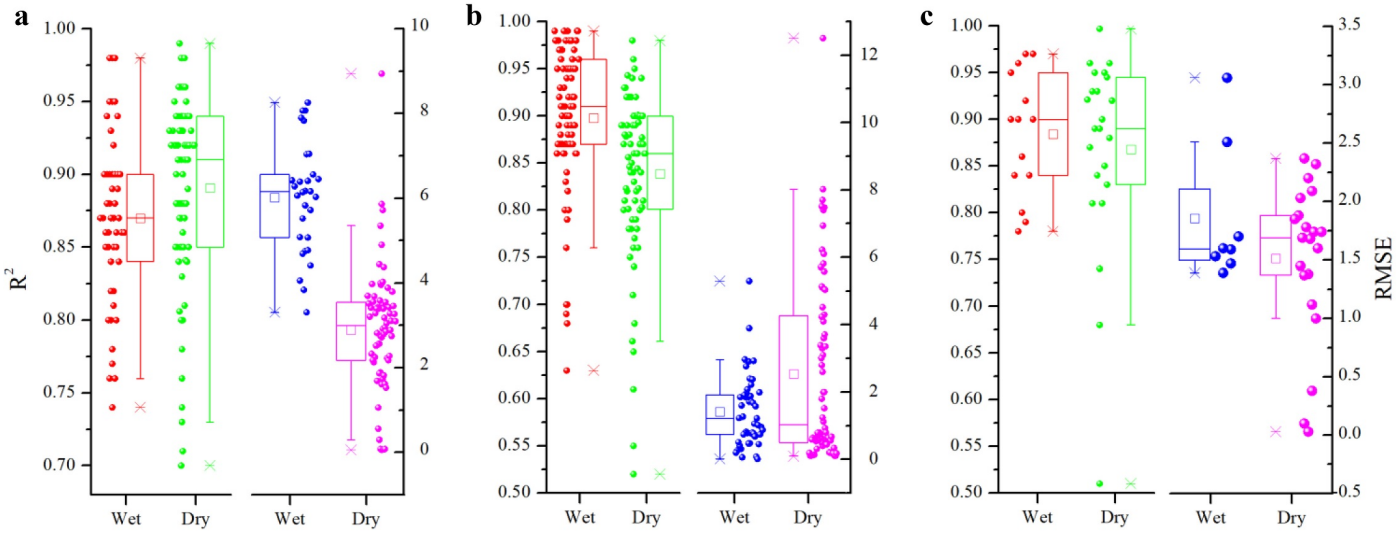


Fig. 6. (a) Predictive performance for machine learning predicting the yields with data from (Abdulsalam et al., 2020; Alabdrabalnabi, 2021; Ascher et al., 2022a; Castro Garcia et al., 2022; Cheng et al., 2022, 2020a, 2020b; Gu et al., 2021; Kartal and Özveren, 2022a; Katongtung et al., 2022; Khan et al., 2022; Leng et al., 2021d and 2022c; Li et al., 2018b, 2020, 2021c, 2021d, and 2022b; Mu et al., 2022; Onsree and Tippayawong, 2021; Pathy et al., 2020; Sezer et al., 2021; Shafizadeh et al., 2022; Shahbeik et al., 2022; Tang et al., 2020 and 2021; Ullah et al., 2021; Yang et al., 2022a; Zhang et al., 2021 and 2022; Zhu et al., 2019a); **(b)** compositions (elemental compositions, proximate analysis, and gas compositions only) with data from (Ascher et al., 2022a; Cheng et al., 2020b; Djandja et al., 2021; Haq et al., 2022; Leng et al., 2022c; Li et al., 2018b, 2021a, 2021d, and 2022b; Mu et al., 2022; Mutlu and Yucel, 2018; Shafizadeh et al., 2022; Shahbeik et al., 2022; Shenbagaraj et al., 2021; Tang et al., 2020 and 2021; Wang et al., 2021b; Yang et al., 2022a; Zhang et al., 2021 and 2022; Zhao et al., 2021; Zhu et al., 2019a); and **(c)** properties (higher/lower heating values only) with data from (Abdulsalam et al., 2020; Ascher et al., 2022a; Chen et al., 2022c; Cheng et al., 2020b; Katongtung et al., 2022; Leng et al., 2021d; Li et al., 2018b, 2020, and 2021d; Mu et al., 2022; Mutlu and Yucel, 2018; Shafizadeh et al., 2022; Shahbeik et al., 2022; Zhang et al., 2022).

Table 6.
Machine-learning-aided prediction of the compositions (excluding elemental compositions, proximate analysis, and gas compositions, which can be found in **Supplementary Information**) of thermochemical treatment products.

Biomass	Inputs [a]	Data Number	Outputs [b]	Predictive Performance, Test Data			Reference
				Algorithms [c]	R^2	RMSE [d]	
Wet thermochemical treatment (hydrothermal treatment)							
Biomass	EC, AR, OC	117	N-heterocycle_oil	RF	0.67	10.16	Leng et al. (2022c)
	EC, AR, BC, OC	91			0.59	9.62	
	EC, AR, BC, OC, Yield_oil, N_oil	91			0.75	7.99	
	EC, AR, OC, Yield_oil, N_oil	117			0.82	7.60	
Sewage sludge	PA, EC, TP, OC	109	TP_char	RF	0.93	3.88	Djandja et al. (2022)
	PA, EC, OC				0.80	6.50	
	PA, TP, OC				0.92	4.44	
	EC, TP, OC				0.92	4.11	
Biomass	EC, BC, OC	224	TN_aqueous	GBDT	0.96	1.25	Leng et al. (2022d)
				RF	0.95	1.45	
		174	TOC_aqueous	GBDT	0.91	9.06	
				RF	0.86	14.87	
		136	TP_aqueous	GBDT	0.90	0.11	
				RF	0.86	0.13	
Microalgae and sorghum distillery residue	PA, EC, OC, H_2SO_4	49	Glucose_aqueous	MARS	0.93	NA[e]	Chen et al. (2022e)
				ANN	0.99	NA	

[a] EC: elemental compositions of biomass; AR: atomic ratios of biomass; BC: biological compositions of biomass; PA: proximate analysis of biomass; OC: operating conditions of thermochemical process; Yield_oil: yield of oil; N_oil: N content of oil; TP: total phosphorus of biomass.
[b] Labeled according to "Composition_product"; N-heterocycles: relative content of N-heterocycles; TP: total phosphorus; TN: total nitrogen; TOC: total organic carbon; Glucose: relative content of glucose
[c] RF: random forest; GBDT: gradient boosting decision tree; MARS: multivariate adaptive regression splines; ANN: artificial neural network.
[d] Root mean square error
[e] NA: not available.

Please cite this article as: Li H., Chen J., Zhang W., Zhan H., He C., Yang Z., Peng H., Leng L. Machine-learning-aided thermochemical treatment of biomass: a review. Biofuel Research Journal 37 (2023) 1786-1809. DOI: 10.18331/BRJ2023.10.1.4

content of N in bio-oil were included because they are highly related to the N-heterocycles (Leng et al., 2022c) (**Table 6**). This means that one composition can be used to predict the other compositions of the thermochemical products. For example, the elemental compositions of bio-oil have been predicted by ML using functional group compositions, as characterized by Fourier-transform infrared spectroscopy (Chen et al., 2022a). There are many other chemical components of the oil and aqueous phases, e.g., hydrocarbons, phenols, acids, ketones, alcohols, amines, nitriles, and furans (Leng et al., 2018b, 2020c, and 2020e; Hoang Pham et al., 2021; Mohamed et al., 2022; Zhou et al., 2022a) and many functional groups of char, such as N/O/S-containing functional groups (Leng et al., 2020b and 2022a; Xu et al., 2021a), which can also be predicted by ML but have received limited interest thus far.

4.3. Prediction of the properties of products

The prediction of the HHV or LHV of the oil, char, or gas is the most popular among researchers, and the predictive performance of the dry and wet thermochemical processes is similar, with an R^2 and RMSE of approximately 0.90 and 1.5, respectively (**Fig. 6c**), which is comparable to, if not better than, the performance of ML predicting the HHV of various municipal solid wastes including biomass, as shown in a previous review (Bagheri et al., 2019). It should be noted that the caloric value scopes of gas are much smaller than those of oil and char; therefore, the RMSE for the prediction of the caloric values of gas is generally lower (Mutlu and Yucel, 2018; Ascher et al., 2022a). Other properties predicted include energy recovery and energy density (Li et al., 2020) which are highly related to caloric values and yields of products, pH of oil (Chong et al., 2022) and aqueous phases (Leng et al., 2022d), aromaticity (Cao et al., 2021), SSA of char (Liao et al., 2019; Leng et al., 2022b; Palansooriya et al., 2022), and viscosity of oil (Zhang et al., 2022), which have been modeled with acceptable predictive performance (**Table 7**). However, there are still many other properties, such as density, flash point, pour point, acid number of bio-oil, porosity (total pore volume, micro-/meso-pore volume, and average pore size), and cation/anion exchange capacity of biochar, that are important for the application of these products (Ippolito et al., 2020) and have not yet been reported; thus, they are worthy of further investigation.

Table 7.
Machine-learning-aided prediction of the properties (excluding caloric value, which can be seen in supplementary material) of thermochemical treatment products.

Biomass	Inputs [a]	Data Number	Outputs [b]	Predictive Performance, Test Data			Reference
				Algorithms [c]	R^2	RMSE [d]	
Wet thermochemical treatment (hydrothermal treatment)							
Biomass	PA, EC, OC	248	ER_char	SVR	0.92	6.22	Li et al. (2020)
			ED_char		0.89	0.09	
Biomass	PA, EC, AR, OC	248	ER_char	DNN	0.88	7.60	Li et al. (2021d)
Biomass	EC, BC, OC	257	pH_aqueous	GBDT	0.93	0.36	Leng et al. (2022d)
				RF	0.87	0.49	
Dry thermochemical treatment							
Biomass	EC, AR, BC, OC	101	Viscosity_oil	RF	0.72	0.014	Zhang et al. (2022)
	EC, OC				0.76	0.012	
	BC, OC				0.73	0.009	
	PA, OC				0.72	0.013	
Biomass	PA, EC, OC	128	pH_oil	RSML	NA [e]	NA	Chong et al. (2022)
Biomass	EC, AR	98	Aromaticity_char	MW	0.91	0.05	Cao et al. (2021)
				GP	0.95	0.04	
Biomass	EC, OC	253	SSA_char	RF	0.83	33.84	Leng et al. (2022b)
				GBR	0.90	26.03	
	PA, OC	183		RF	0.88	38.22	
				GBR	0.84	45.42	
	BC, OC	75		RF	0.76	68.78	
				GBR	0.86	53.17	
	EC, EA, OC	183		RF	0.86	41.62	
				GBR	0.93	29.89	
	EC, EA, BC, OC	68		RF	0.85	51.85	
				GBR	0.91	39.65	
Biomass	EC_char, AR_char, pH_char, OC	131	SSA_char	RF	0.98	16.09	Palansooriya et al. (2022)
Biomass	PA, OC, AC	155	SSA_char	ANN	0.94	NA	Liao et al. (2019)
	EC, OC, AC				0.93	NA	
	PA, EC, OC, AC				0.92	NA	
Biomass	PA, EC, OC	165	ER_char	SVR	0.79	9.40	Li et al. (2020)
			ED_char		0.88	0.09	
Biomass	PA, EC, OC, HHV	329	Char_exergy value	ANN	0.797	NA	Kartal and Özveren (2022b)
Biomass	EC, OC	NA	Gas_exergy value	ANN	0.999	NA	Sezer and Özveren (2021)

[a] PA: proximate analysis of biomass; EC: elemental compositions of biomass; OC: operating conditions of thermochemical process; AR: atomic ratios of biomass; BC: biological compositions of biomass; EC_char: elemental composition of char; AR_char: atomic ratios of char; pH_char: pH of char; HHV: higher heating value.
[b] Labeled according to "Composition_product"; ER: energy recovery; ED: energy density; SSA: specific surface area.
[c] SVR: support vector regression; DNN: deep neural network; GBDT: gradient boosting decision tree; RF: random forest; RSML: rough set machine learning; MW: Mazumdar–Wang; GP: genetic programming; GBR: gradient boosting regression; ANN: artificial neural network.
[d] Root mean square error.
[e] NA: not available.

Please cite this article as: Li H., Chen J., Zhang W., Zhan H., He C., Yang Z., Peng H., Leng L. Machine-learning-aided thermochemical treatment of biomass: a review. Biofuel Research Journal 37 (2023) 1786-1809. DOI: 10.18331/BRJ2023.10.1.4

4.4. Other predictions

(i) Simultaneous prediction of yields, compositions, and properties

Most current studies have performed single-task predictions, with a few reporting on multi-task predictions for simultaneous prediction of the yields of two or more product phases (Alabdrabalnabi, 2021; Leng et al., 2021d; Castro Garcia et al., 2022) as well as on the yield of one product phase and its compositions and/or properties (Li et al., 2021a and 2021c; Leng et al., 2022c). However, no study has been conducted on the yields of multiple product phases and their compositions and/or properties. Although many targets were predicted in some studies, they were achieved using the single-target prediction mode (Ascher et al., 2022a; Shafizadeh et al., 2022; Shahbeik et al., 2022). In these multitask predictions, a predictive performance comparable to single-target predictions was achieved. According to the biorefinery concept, all product phases from the thermochemical treatment of biomass should be utilized or disposed of to valorize the biomass resource fully (Fan et al., 2020; Watson et al., 2020). In addition, although it is possible to use multiple single-target models for the predictions of the required targets, it would be difficult to mediate the targets among many models when applying ML models. Therefore, multi-task predictions for simultaneous prediction and mediation of the yields of multiple product phases and their compositions and properties within one model are favorable and encouraged. However, the availability of data limits its implementation.

(ii) Prediction of the thermochemical behavior and kinetics

Other studies have reported the prediction of thermochemical conversion behavior or kinetics, for example, the degree of dehydration and decarboxylation of char during HTC (Mu et al., 2022), remaining mass of biomass during pyrolysis (Zhang et al., 2019), thermal degradation rate (Lee et al., 2023), torrefaction severity index (Chen et al., 2022d), activation energy (E_a) of biomass pyrolysis (Kartal and Özveren, 2022c; Wang et al., 2022), combustion index during biomass combustion (Sezer et al., 2022), and exergy of char or gas from biomass pyrolysis (Sezer et al., 2021; Sezer and Özveren 2021; Kartal and Özveren, 2022b) (**Table 8**), which can be useful for understanding the thermal degradation behavior and reaction kinetics of biomass or thermochemical process performance, and more studies should be conducted in this area.

4.5. Interpretation of the ML models

ML is very popular because it is not just a "black box" for predictions; it can also be interpreted to help understand the reaction mechanisms and engineer the thermochemical processes. First, the factors affecting the target can be ranked according to their importance level against the target through feature importance analysis (**Fig. 3**). Additionally, PDP and SHAP method analyses can be used to indicate the correlations between variables and targets. In PDP, the effect of any one input or the mutual influences of any two input variables on the predicted target can be plotted to show the linear, monotonous, or more complex connections between the input and output features. SHAP analysis can also be used to analyze feature importance levels as well as influence trends. In addition, SHAP assigns a value for each data point of the input feature to indicate its importance to the target, and the accumulation of SHAP values of all data points results in a full interpretation of the studied feature. These three interpretation methods have been widely used in the reviewed studies.

Table 9 lists the top features for predicting the yields, compositions, and properties of the thermochemical products reported in the reviewed articles. The yield of bio-oil from HTL is dominated by temperature and lipids (contents of C and H are highly related to lipids (Leng et al., 2022c)) (**Table 9**). Temperature is also the most important factor for bio-oil yield, as well as char yield from pyrolysis, but the other important features can vary in different studies. C and H are important to HHV, as they are the major elements that can be combusted to release heat. N can be very significant to the yields, compositions, and properties of bio-oil from HTL and pyrolysis because it participates in reactions to yield oil components, e.g., through the Maillard reaction (Leng et al., 2020c and 2020e). It seems that N has a more significant effect on the yield of char during HTC than on that of oil during HTL. The more prominent role of N on oil and char in wet thermochemical processes than in dry processes is because biomass with a higher N content is more commonly used in wet processes. The prediction of biochar SSA indicates that the parameters during activation are much more significant than the pyrolysis parameters (Liao et al., 2019), which is why biochar requires activation before it can be used to replace activated carbon from fossil resources. There are many other interesting and useful pieces of information in the ranking lists shown in **Table 9**, but these will not be detailed due to limited space.

Table 8.
Machine-learning-aided prediction of thermochemical conversion behavior or kinetics.

Biomass	Inputs [a]	Data number	Outputs [b]	Predictive Performance, Test Data			Reference
				Algorithms [c]	R^2	RMSE [d]	
Wet thermochemical treatment (hydrothermal treatment)							
Biomass	PA, EC, OC	296	DHD_char	NN	0.88	NA [e]	Mu et al. (2022)
			DCD_char		0.84		
			DHD_char	PSO-NN	0.85		
			DCD_char		0.91		
Dry thermochemical treatment							
Biomass	Biomass type, EC, AR, E_a model	281	E_a	RF	0.99	8.47	Wang et al. (2022)
Biomass, coal, plastics, blends, etc.	EC, OC, particle size, thermal degradation data	NA	E_a	ANN	0.964	NA	Kartal and Özveren (2022c)
Biomass	Thermal degradation data	6721	Combustion index	ANN	0.85-0.99	0.90-13.4	Sezer et al. (2022)
Biomass	Biomass type, OC	1626	Severity index	ANN	0.985	NA	Chen et al. (2022d)
				MARS	0.978		
Cattle manure	OC	86200	Remaining mass	GBR	0.998	0.82	Zhang et al. (2019)
				RF	0.999	0.43	

[a] PA: proximate analysis of biomass; EC: elemental compositions of biomass; OC: operating conditions of the thermochemical process; AR: atomic ratios of biomass; E_a model: activation energy calculation models.
[b] DHD_char: dehydration degree of char; DCD_char: decarboxylation degree of char; E_a: activation energy.
[c] NN: neural network; PSO: particle swarm optimization; RF: random forest; ANN: artificial neural network; MARS: multivariate adaptive regression splines.
[d] Root mean square error.
[e] NA: not available.

Please cite this article as: Li H., Chen J., Zhang W., Zhan H., He C., Yang Z., Peng H., Leng L. Machine-learning-aided thermochemical treatment of biomass: a review. Biofuel Research Journal 37 (2023) 1786-1809. DOI: 10.18331/BRJ2023.10.1.4

Table 9.
Machine-learning-aided prediction of thermochemical conversion behavior or kinetics.

Product	Wet Thermochemical Treatment [a,b]	Dry Thermochemical Treatment [a,b]
Oil	**Yield**: Lipid > T > RT (Li et al., 2021c); T > Lipid > RT (Zhang et al., 2021); C > H > T (Shafizadeh et al., 2022); T > carbohydrate > protein (Katongtung et al., 2022); O/C > H/C > ash (Cheng et al., 2020b) **HHV**: H > T > N (Shafizadeh et al., 2022); T > lipid > protein (Katongtung et al., 2022) **ER**: Lipid > T > RT (Li et al., 2021c) **C**: O > C > ash (Shafizadeh et al., 2022); O/C > ash > T (Cheng et al., 2020b) **H**: H > O > C (Shafizadeh et al., 2022) **O**: T > RT > O/C (Zhang et al., 2021); N > T > O (Shafizadeh et al., 2022) **N**: N > T > RT (Li et al., 2021c); T > protein > RT (Zhang et al., 2021); N > C > ash (Shafizadeh et al., 2022) **S**: S > RT > T (Shafizadeh et al., 2022) **N-heterocycle**: N/C > protein > lipid (Leng et al., 2022c)	**Yield**: FC > V > ash (Tang et al., 2020); T > cellulose > hemicellulose (Leng et al., 2021d); Log PS > Log HR > HR (Ge et al., 2021); T > H > PS (Zhang et al., 2022); ash > FC > T (Shahbeik et al., 2022) **HHV**: C > H > T (Leng et al., 2021d); cellulose > hemicellulose > PS (Zhang et al., 2022) **Viscosity**: N > PS > C (Zhang et al., 2022) **H**: H > N > PS (Tang et al., 2020) **H/C**: PS > lignin > hemicellulose (Zhang et al., 2022) **O/C**: O > C > T (Zhang et al., 2022)
Char	**Yield**: N > T > C (Li et al., 2020); T > N > SC (Li et al., 2021d); N > T > C (Shafizadeh et al., 2022); C > T > O (Mu et al., 2022); T > O/C > H/C (Cheng et al., 2020b) **ER**: N > C > SC (Li et al., 2020); N > T > SC (Li et al., 2021d) **ED**: T > O > FC (Li et al., 2020); **HHV**: C > T > H (Li et al., 2020); C > H > ash (Li et al., 2021d); T > C > O (Mu et al., 2022); T > H/C > ash (Cheng et al., 2020b) **C**: C > ash > H (Li et al., 2021d); T > H/C > ash (Cheng et al., 2020b) **TP**: total phosphorus > O > VM (Djandja et al., 2022) **N/C**: N > C >H (Mu et al., 2022) **Ash**: O > ash > FC (Mu et al., 2022)	**Yield**: T > ash > hemicellulose (Zhu et al., 2019a); T > C > HR (Li et al., 2020); T > C > H/C (ash) (Pathy et al., 2020); T > RT > VM (Onsree and Tippayawong, 2021); T > lignin > PS (Leng et al., 2021d); T > RT > HR (Khan et al., 2022); T > ash > N (Shahbeik et al., 2022) **SSA**: H/C_char > pH_char > T (Palansooriya et al., 2022); T_AC > steam > RT_AC (Liao et al., 2019) **ER**: T > H > O (Li et al., 2020) **ED**: H > O > ash (Li et al., 2020) **HHV**: C > ash > T (Li et al., 2020); FC > ash > C (Chen et al., 2022c); T > O/C > ash (Cheng et al., 2020a) **C**: T > HR > C (Zhu et al., 2019a); T > VM > ash (Kartal and Özveren, 2022a) **H**: T > C > O (Kartal and Özveren, 2022a) **O**: C > FC > ash (Kartal and Özveren, 2022a) **N**: N/C > T > ash (Cheng et al., 2020a) **H/C**: T > HR > FC (Shahbeik et al., 2022) **O/C**: N > O > C (Shahbeik et al., 2022) **H/N**: T > N > HR (Shahbeik et al., 2022)
Gas	**Yield**: T > pressure > O (Shafizadeh et al., 2022); H/C > T > SC (Cheng et al., 2020b) **H_2**: T > SC > C (non-catalytic); SC > T > catalyst dose (catalytic) (Li et al., 2021a); SC > T > H (Zhao et al., 2021); T > SC > P (Haq et al., 2022; Li et al., 2021b)	**Yield**: T > gas feeding rate > cellulose (Leng et al., 2021d); T > potassium > FC (Gu et al., 2021); T > H > C (Shahbeik et al., 2022) **H_2**: T > HR > S (Shahbeik et al., 2022) **CH_4**: T > S > HR (Shahbeik et al., 2022) **CO_2**: T > N > H (Shahbeik et al., 2022) **CO**: T > S > C (Shahbeik et al., 2022)
Aqueous	**Yield**: T > P > S (Shafizadeh et al., 2022); SC > O/C > ash (Cheng et al., 2020b) **pH**: T > N > O (Leng et al., 2022d) **TN**: protein > N > SC (Leng et al., 2022d) **TP**: SC > T > lipid (Leng et al., 2022d) **TOC**: RT > lipid > T (Leng et al., 2022d)	NA [c]
Reaction behavior	**DHD**: T > ash > C (Mu et al., 2022) **DCD**: T > O > N (Mu et al., 2022)	**E_a**: C > H/C > S (Wang et al., 2022)

[a] Bold front text represents the yields, compositions, and properties of oil/char/gas/aqueous phases, and the abbreviations can be found in **Tables 6**, **7**, and **8**.
[b] Normal text is the elemental composition (C, H, O, N, S), atomic ratio (H/C, O/C, N/C), biological composition (cellulose, lignin, hemicellulose, protein, lipid, carbohydrate), and proximate analysis (ash, fixed carbon (FC), volatile matter (VM)) of biomass or the operating condition, such as temperature (T), retention time (RT), heating rate (HR), particle size (PS), and solid content (SC), of thermochemical processes unless specified otherwise. H/C_char: H/C of char; pH_char: pH of char; T_AC: temperature of activation process; RT_AC: retention time of activation process.
[c] NA: not available.

4.6. Application of the ML models

The ML models obtained are mainly used to aid the thermochemical treatment of biomass. First, they can be used to predict the yields, compositions, and properties without performing experiments, which is the major focus of current studies. Additionally, ML models can be used to provide raw data for subsequent applications. For example, input data in life cycle assessment (LCA) and economic analysis studies can be obtained from ML predictions of the yields, compositions, and properties of char/oil/gas/aqueous phases, LCA (Cheng et al., 2020a and b), which can facilitate a more comprehensive LCA (Cheng et al., 2020a, 2020b). Second, ML can be applied to solve optimization problems, e.g., finding the optimal processing

Please cite this article as: Li H., Chen J., Zhang W., Zhan H., He C., Yang Z., Peng H., Leng L. Machine-learning-aided thermochemical treatment of biomass: a review. Biofuel Research Journal 37 (2023) 1786-1809. DOI: 10.18331/BRJ2023.10.1.4

parameters for a given biomass (with known compositions, which are used as inputs) to produce target products, which is considered forward optimization. In this forward optimization, iterations of processing parameters with set step sizes yield corresponding targets, from which functions can be used to find preferable solutions. In addition, optimization of the biomass mixing ratio can be achieved by evaluating the results obtained from different biomass mixtures. ML models can also reverse design not only processing parameters but also biomass compositions (or biomass mixing recipes), which is considered reverse optimization. Through these methods, ML models can be applied to aid experimental studies to effectively determine the optimal solutions of thermochemical treatment processes and verify the validity of the models. For example, forward optimization of bio-oil production from HTL of specified algae using the iteration method and reverse optimization using the particle swarm optimization method with model compounds to obtain preferable bio-oil were conducted with experimental verification, and the results were satisfactory (**Fig. 7**) (Li et al., 2021c; Zhang et al., 2021).

However, the optimization and experimental verification were only conducted in limited studies, with several studies reporting the optimization without verification (Leng et al., 2022c; Li et al., 2022a), and most other studies solely reporting the predictive performance (and model interpretation). Further studies in this area are encouraged to put these models into practice. It should be noted, however, that verification may not necessarily have to be accomplished through experiments; it can also be achieved by other methods, such as process modeling in Aspen Plus and other computational modelings (Li et al., 2021b and 2022a; Ullah et al., 2022).

crucial for the ML model's performance. The accuracy of the data was doubtful in some studies, such as inconsistency in calculation formulas (some on a dry basis while others on a DAF basis), feature engineering, scientific processing of textual data, etc. Furthermore, few articles have focused on the variables derived from thermochemical treatment-related processes during the construction of ML models based on biomass thermochemical conversion, such as the variables in the pretreatment of biomass, bio-oil separation, catalytic conversion, etc.

(ii) Modeling process. Different researchers used varied ML schemes during the construction of the ML model (**Fig. 3**) to obtain the optimized models trained and tested with the highest R^2 and/or lowest RMSE. However, schemes used in some studies are problematic. For example, models trained without cross-validation may have the problem of overfitting. Moreover, more indicators (e.g., generalizability), in addition to R^2 and RMSE, could be introduced to evaluate the model for better predictive performance. There are many hyperparameters for ML models, but only one or two of them were tuned in most studies, with other hyperparameters being unknown (whether they were tuned or the default).

(iii) Model application. Few studies focused on ML model optimization for enhanced thermochemical treatment. On the other hand, some studies only reported the predictive performance but with no interpretation of the ML model, particularly research based on ANN. In addition, the exploration of thermochemical conversion mechanisms based on ML model interpretation is rare.

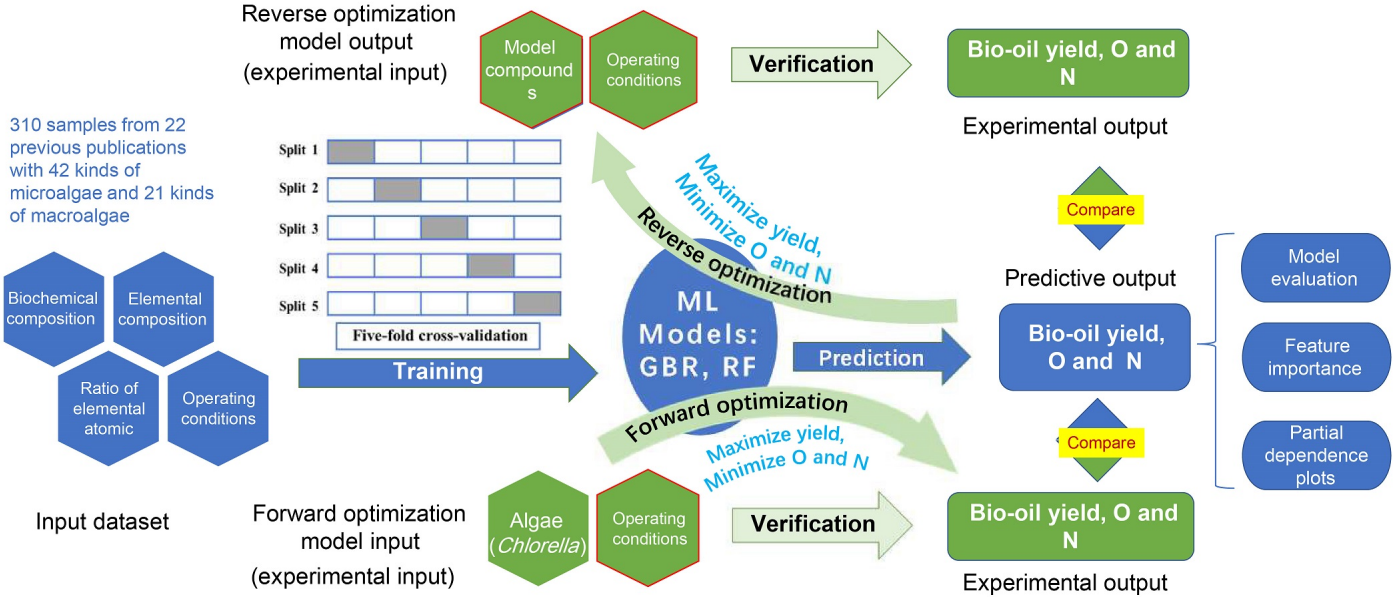


Fig. 7. Machine learning prediction and optimization of bio-oil production from hydrothermal liquefaction of algae with experiment verification (Zhang et al., 2021). Copyright© Elsevier 2021. Licence Number: 5496900753721.

5. Limitations and implications

5.1. Limitations of current studies

ML model, trained with experimental data, can predict the output accurately, and the application of ML to aid the thermochemical treatment of biomass has been receiving growing attention in recent years. As discussed in the previous parts of this work, many authors have published papers about the application of ML for predicting the yields, compositions, and properties of products from wet and dry thermochemical treatments of biomass. However, there are some limitations associated with these studies:

(i) Data: Data unavailability is common, and data details, including input and output datasets, are missing in many studies. The accuracy of data is also

5.2. Practical implications of this review

This review summarized and compared the up-to-date research in both machine-learning-aided wet and dry thermochemical treatment of biomass. In addition, the ML schemes, as well as strategies and descriptors of the input and output features in thermochemical processes, were also introduced. This study would make a significant practical contribution to the research and application work of the thermochemical treatment of biomass.

First, researchers can find the state-of-the-art, major, and strongly influenced journals and funding agencies of the ML-aided thermochemical treatment of biomass from this review. The summary and comparison of characterizations of biomass, technologies, and products would help

Please cite this article as: Li H., Chen J., Zhang W., Zhan H., He C., Yang Z., Peng H., Leng L. Machine-learning-aided thermochemical treatment of biomass: a review. Biofuel Research Journal 37 (2023) 1786-1809. DOI: 10.18331/BRJ2023.10.1.4

interested researchers to have a deeper understanding of the field of biomass thermochemical conversion.

Second, ML schemes and algorithms, which would be very useful for new researchers interested in carrying out studies on ML-aided thermochemical conversion of biomass, were introduced in this review. Moreover, the discussion about the application of ML for predicting the yields, compositions, and properties of products from wet and dry thermochemical treatments of biomass can provide new inspiration and guidance for researchers. Moreover, it would be meaningful and helpful for the rapid development of the research area.

Third, the limitations of the present review on the ML-aided thermochemical treatment of biomass have also been overviewed, and the major challenges and perspectives were put forward, which could shed light on bridging the major gaps between the studies and real-world needs.

6. Major challenges and perspectives

6.1. Improving the predictive performance of models

Predictive performance is the top priority of ML studies because it is the basis of interpretation, optimization, and application (monitoring, controlling, etc.). Data availability, inconsistency, and accuracy are vital for predictive performance. Including more cases to enlarge the dataset size (data number of several hundred or more is preferable) and introducing new input descriptors, such as subunit compositions of lignin, cellulose/hemicellulose, and protein (as detailed in *Section 2.1*); image (Ögren et al., 2018) and color of products (Li et al., 2018a); category descriptors for biomass or thermochemical processes (Ascher et al., 2022a); biomass ash compositions (Yan et al., 2020); and molecular simulation results from model biomass (Freitas et al., 2022) are effective approaches to increase the data dimension. Interpolation of missing data in a dataset is meaningful for ensuring data availability; interpolation by algorithms directly (Sun et al., 2022) or by building another ML model (Palansooriya et al., 2022) are both effective methods. Additionally, redundant input variables identified during ML modeling or feature analysis can be removed from the dataset for better prediction performance. Data consistency requires the data to be generated under the same or highly comparable conditions. For example, the consistency of the calculation equations for the input and output variables is the first thing to note. Researchers are advised to carefully check the calculation of the elemental compositions of biomass and char, as well as the yields of all products if they are calculated on a dry, DAF, or other bases. However, thermochemical reactors and their configurations in different studies may have large differences, and category descriptors for reactors should be listed as variables to overcome such inconsistencies. Currently, data are collected indiscriminately from references for most studies, and the accuracy of the data is not considered. Future studies may explore effective methods to exploit only accurate data for ML modeling. For example, data from modeling, e.g., data from Aspen Plus process modeling (Sezer and Özveren 2021), should be validated before use in ML modeling.

Model screening, feature selection, and model hyperparameter tuning are key to predictive performance. Current studies mainly screen ML based on evaluation metrics such as R^2 and RMSE because most researchers in this area are not from computer science, and many are not truly familiar with the working mechanisms of ML algorithms. Collaboration with peers in computer science and a screening model depending on the applicable characteristics of each ML technique corresponding to a given problem are encouraged. Suitable features and hyperparameters should be selected based on the domain expertise of the collaborators in thermochemical treatment and computer science. Additionally, optimization algorithms can be used before modeling to screen feature pairs (Ullah et al., 2021; Khan et al., 2022) or during modeling to obtain the optimal hyperparameters (Haq et al., 2022; Mu et al., 2022; Shafizadeh et al., 2022), which is user-friendly for non-expert ML users. Genetic algorithms and particle swarm optimization (PSO) are more commonly used in these optimizations than other algorithms, such as the Rao algorithm, Sine Cosine Algorithm, and grey wolf optimization (Khan et al., 2022). When tuning the hyperparameters, cross-validation should be used, and the number of folds of the cross-validation may have a considerable effect on the RMSE; for example, the test RMSE was reduced from 8.43 to 8.07 when the fold number increased from 10 to 100 (Cheng et al., 2022). However, in the reviewed articles, some studies did not use a cross-validation process, which would result in overfitting because the optimum hyperparameters were obtained most probably by chance, although running trial and error modeling several times may be beneficial for increasing accuracy (Djandja et al., 2021).

For multi-target ML, optimizing the weight percent of each target (generally treated equally in most studies) can also balance the ML to obtain preferable predictive performance for all studied targets. Finally, advanced algorithms and modeling techniques such as deep learning (Lecun et al., 2015) can be used to improve predictive performance, especially for cases with a large amount of data.

6.2. Increasing model generalizability

The ML model cannot be simply evaluated by R^2 and RMSE; other indicators, such as the model's generalizability, are also important, and trade-offs between these two should be considered. A model with good predictive performance (high R^2 and low RMSE) does not necessarily indicate high generalizability. A model built based on a specific biomass type or thermochemical parameters, such as a specific thermochemical reactor, is likely to work only within this specific condition; it may not have the generalizability to predict under other conditions. For example, even if the data numbers are higher than 1000 with a predictive model R^2 higher than 0.95, the models built based on one or two biomasses in a particular gasifier (Mutlu and Yucel, 2018; Elmaz et al., 2020; Elmaz and Yücel, 2020) cannot be used for accurately predicting other gasification processes. To obtain good generalizability, the coverage of the descriptors, amount of data in the dataset, and data distribution should be assessed carefully, with data distribution being the most important; bad data distribution, e.g., data of limited or biased coverage, would lead to poor generalizability. Creating highly generalizable models suitable across a wide range of feedstocks as well as thermochemical parameters and ranges, should be promoted in the future. Integrating dry and wet thermochemical treatment datasets to predict the yields, compositions, and properties of the char/oil/gas/aqueous phases without differentiating products from dry or wet processes may be a promising direction for testing. In addition, future ML models should be built with the extrapolative ability to explore the “unseen” space, such as the ML-aided discovery of new materials and chemicals (Butler et al., 2018), which is challenging but of high priority.

6.3. Increasing model interpretability and aiding thermochemical conversion mechanistic studies

Model interpretation results can not only be used to understand the fundamentals behind ML model-based decision-making but also be applied to guide thermochemical treatment mechanistic studies. Through feature ranking and PDP, the effects of biomass compositions and thermochemical parameters on a target can be understood in a rational manner. Mechanistic studies can be conducted for screened cases that are indicative of the connections between biomass compositions/thermochemical parameters and the yields, compositions, and properties of oil, char, gas, and aqueous phases. For some features, such as model biomass chemicals, catalysts, solvents, or additives that are composed of specific chemicals or elements, their molecular modeling data can be used directly as input features to understand how the structures, compositions, or properties of these features affect the yields, compositions, and properties of the oil, char, gas, and aqueous phases or the biomass thermochemical conversion behavior. In addition, more advanced and promising interpretation algorithms can be developed to help understand the connections between inputs and outputs. However, model complexity and interpretability should be balanced because increasing interpretability can result in higher structural complexity of ML.

6.4. Enhancing the real-world application of ML models

The ultimate target of ML studies is the real-world application of ML models. Current studies mainly concentrate on predicting the yields, compositions, and properties of thermochemical products; future studies should focus more on ML optimization. For example, thermochemical products should be engineered by integrating forward and reverse optimizations with the application performance of thermochemical products in different areas. During engineering, the ML model targets (e.g., yields, compositions, and properties of the oil/char/gas/aqueous products)

Please cite this article as: Li H., Chen J., Zhang W., Zhan H., He C., Yang Z., Peng H., Leng L. Machine-learning-aided thermochemical treatment of biomass: a review. Biofuel Research Journal 37 (2023) 1786-1809. DOI: 10.18331/BRJ2023.10.1.4

should be screened based on the effects of the compositions and properties of the oil/char/gas/aqueous products on application performance. Therefore, the properties/structures of oil/char/gas/aqueous application performance relationships should be understood first, preferably by ML (e.g., those shown in **Table 4**). The main descriptors of oil/char/gas/aqueous products determining the application performance will be used as targets in thermochemical process ML models, with biomass compositions and thermochemical parameters as inputs. Therefore, the ML models in **Table 4** can be integrated with the models presented in *Section 4* to guide the production of smart products (see **Fig. 8**). For example, the main descriptors determining CO_2 adsorption capacity, namely SSA, total pore volume, contents of N and O, and mesopore volume (Yuan et al., 2021), can be optimized within the as-built thermochemical treatment-for-biochar production prediction ML models to obtain optimal biochar production parameters, produce smart biochar, and achieve the highest CO_2 adsorption capacity. The engineering of oil/char/gas/aqueous products in other areas, such as those in **Table 5**, can also be conducted in this manner. However, no studies have yet been conducted in this area.

models. One study reported that the MLP-NN model was approximately three times faster than the artificial neuro-fuzzy inference system (ANFIS) model (Li et al., 2022b). Another concern is whether ML models are reliable and efficient enough (uncertainty quantifiable and acceptable) to guide and replace human-expert decision-making.

6.5. Promoting data and model sharing in the community

Sharing data and as-built models in published papers should be encouraged. Large-scale and high-quality databases may be built by researchers in this community, thus facilitating high-quality ML studies. Additionally, model sharing allows the models to be evaluated, used, and even rebuilt by others to promote the development of this area more effectively. Simple offline apps and online GUIs have been developed by some researchers (Li et al., 2021a; Ullah et al., 2021; Leng et al., 2022c and 2022d) and can be adopted by others.

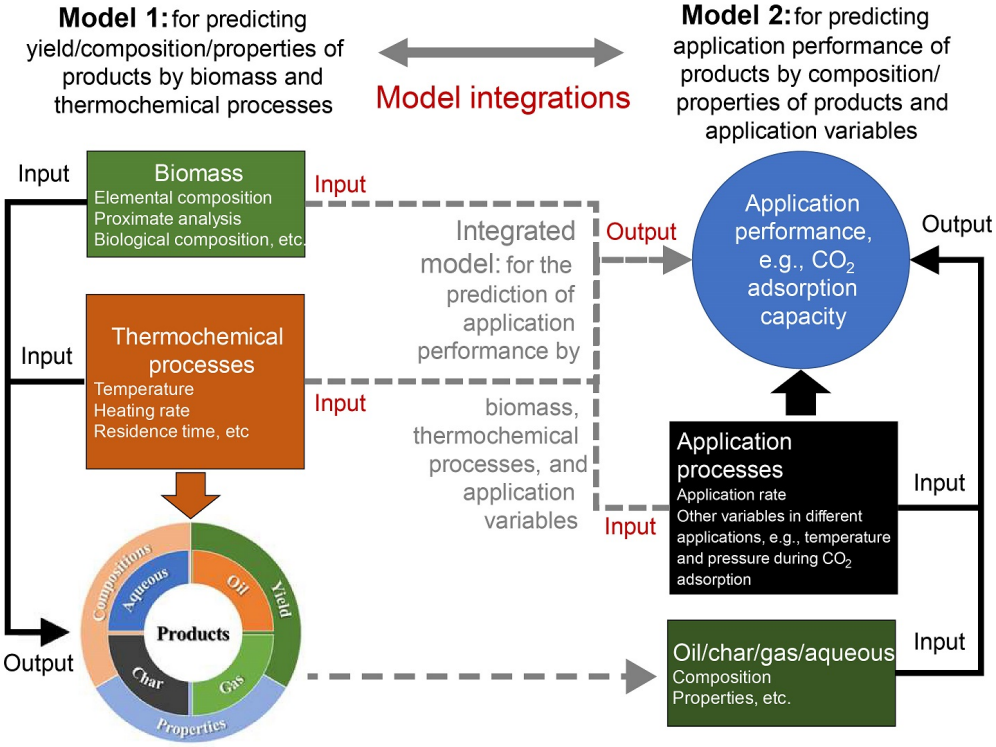


Fig. 8. Integrated models for producing smart oil/char/gas/aqueous products.

In comparison to the prediction of the exact yields, compositions, and properties, the classification of the oil/char/gas/aqueous products according to different applications, such as the classification of the slagging degree of char (Bi et al., 2023) and carbon stability level of char (Leng et al., 2019b; Chen et al., 2021), can be useful for the application of these products. ML can also be used to optimize computational parameters in other computational models, such as CFD and kinetic models, to indirectly aid thermochemical treatment. In addition to solving prediction and optimization problems, the ML model can be used for classification and control. Examples include identifying and classifying images of oil/char/gas/aqueous products from biomass thermochemical treatment processes or images from computational modeling for advanced predictions and monitoring (Zhu et al., 2022). However, few studies have been conducted in this direction.

There are other considerations when applying the ML model to the real world (Meena et al., 2021). For example, computational cost and efficiency (computation time) are vital if the model is used for online monitoring and control. Only a few studies have recorded the computation times of developed

7. Conclusions

General ML schemes and strategies were summarized in this review. Descriptors for the input and output features in the ML models for dry and wet thermochemical processes are similar, and predictive performance is preferable. The predictive performance for the yields of oil/char/gas/aqueous phases in modeling dry thermochemical processes is better than that of wet processes, while an inverse trend was observed for predicting the product compositions. The interpretation of the ML model indicates the key features affecting the yields, compositions, and properties of oil/char/gas/aqueous products, which can be useful in guiding future experimental studies on biomass thermochemical treatment. Improving predictive performance, increasing model generalizability, increasing model interpretability, aiding mechanistic studies, enhancing the real-world application of ML models in various areas, and sharing data and as-built models in the community are the frontiers of future investigations to bring ML to the next stage. In the near future, the development and research of

Please cite this article as: Li H., Chen J., Zhang W., Zhan H., He C., Yang Z., Peng H., Leng L. Machine-learning-aided thermochemical treatment of biomass: a review. Biofuel Research Journal 37 (2023) 1786-1809. DOI: 10.18331/BRJ2023.10.1.4

biomass thermochemical treatment processes are envisaged to be accelerated by ML-aided prediction of yields, compositions, and properties of oil/char/gas/aqueous products, thermochemical conversion behavior and kinetics, as well as the characterization and application performance of different biomass products in various areas, in addition to ML-aided optimization, monitoring, and control of the thermochemical processes.

Acknowledgments

The authors express appreciation to the National Key Research and Development Program of China (2021YFE0104900), Hunan Provincial Natural Science Foundation of China (2022JJ20064), National Natural Science Foundation of China (51808278; 51906247), and the Science and Technology Innovation Program of Hunan Province (2021RC4005). We are grateful to Dr. Badr Mohamed at Cairo University for the helpful comments on earlier drafts.

References

[1] Abdulsalam, J., Lawal, A.I., Setsepu, R.L., Onifade, M., Bada, S., 2020. Application of gene expression programming, artificial neural network and multilinear regression in predicting hydrochar physicochemical properties. Bioresour. Bioprocess. 7(1), 1-22.
[2] Akiya, N., Savage, P.E., 2002. Roles of water for chemical reactions in high-temperature water. Chem. Rev. 102(8), 2725-2750.
[3] Alabdrabalnabi, A., Gautam, R., Sarathy, S.M., 2021. Utilization of Machine Learning to Predict Bio-Oil and Biochar Yields from CoPyrolysis of Biomass with Waste Polymers. Fuel. 328, 125303.
[4] Ascher, S., Sloan, W., Watson, I., You, S., 2022a. A comprehensive artificial neural network model for gasification process prediction. Appl. Energy. 320, 119289.
[5] Ascher, S., Watson, I., You, S., 2022b. Machine learning methods for modelling the gasification and pyrolysis of biomass and waste. Renew. Sust. Energy Rev. 155, 111902.
[6] Ayodele, B.V., Mustapa, S.I., Kanthasamy, R., Zwawi, M., Cheng, C.K., 2021. Modeling the prediction of hydrogen production by co-gasification of plastic and rubber wastes using machine learning algorithms. Int. J. Energy Res. 45(6), 9580-9594.
[7] Bagheri, M., Esfilar, R., Golchi, M.S., Kennedy, C.A., 2019. A comparative data mining approach for the prediction of energy recovery potential from various municipal solid waste. Renew. Sust. Energy Rev. 116, 109423.
[8] Bi, Y., Chen, C., Huang, X., Wang, H., Wei, G., 2023. Discrimination method of biomass slagging tendency based on particle swarm optimization deep neural network (DNN). Energy. 262, 125368.
[9] Butler, K.T., Davies, D.W., Cartwright, H., Isayev, O., Walsh, A., 2018. Machine learning for molecular and materials science. Nature. 559(7715), 547-555.
[10] Cao, H., Milan, Y.J., Mood, S.H., Ayiania, M., Zhang, S., Gong, X., Lora, E.E.S., Yuan, Q., Garcia-Perez, M., 2021. A novel elemental composition based prediction model for biochar aromaticity derived from machine learning. Artif. Intell. Agric. 5, 133-141.
[11] Castro Garcia, A., Shuo, C., Cross, J.S., 2022. Machine learning based analysis of reaction phenomena in catalytic lignin depolymerization. Bioresour. Technol. 345, 126503.
[12] Chen, J., Wang, P., Ding, L., Yu, T., Leng, S., Chen, J., Fan, L., Li, J., Wei, L., Li, J., Lu, Q., Leng, L., Zhou, W., 2021. The comparison study of multiple biochar stability assessment methods. J. Anal. Appl. Pyrolysis. 156, 105070.
[13] Chen, C., Liang, R., Ge, Y., Li, J., Yan, B., Cheng, Z., Tao, J., Wang, Z., Li, M., Chen, G., 2022a. Fast characterization of biomass pyrolysis oil via combination of ATR-FTIR and machine learning models. Renewable Energy. 194, 220-231.
[14] Chen, J., Zhang, J., Pan, W., An, G., Deng, Y., Li, Y., Hu, Y., Xiao, Y., Liu, T., Leng, S., Chen, Jiefeng, Li, J., Peng, H., Leng, L., Zhou, W., 2022b. 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.
[15] Chen, J., Ding, L., Wang, P., Zhang, W., Li, J., A. Mohamed, B.A., Chen, J., Leng, S., Liu, T., Leng, L., Zhou, W., 2022c. The estimation of the higher heating value of biochar by data-driven modeling. J. Renew. Mater. 10, 1555-1574.
[16] Chen, W.H., Aniza, R., Arpia, A.A., Lo, H.J., Hoang, A.T., Goodarzi, V., Gao, J., 2022d. A comparative analysis of biomass torrefaction severity index prediction from machine learning. Appl. Energy. 324, 119689.
[17] Chen, W.H., Lo, H.J., Aniza, R., Lin, B.J., Park, Y.K., Kwon, E.E., Sheen, H.K., Alan, L., Grafilo, L.A.D.R., 2022e. 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.
[18] Chen, X., Zhang, H., Song, Y., Xiao, R., 2018. Prediction of product distribution and bio-oil heating value of biomass fast pyrolysis. Chem. Eng. Process.-Process Intensif. 130, 36-42.
[19] 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.
[20] Cheng, F., Luo, H., Colosi, L.M., 2020a. Slow pyrolysis as a platform for negative emissions technology: an integration of machine learning models, life cycle assessment, and economic analysis. Energy Convers. Manage. 223, 113258.
[21] Cheng, F., Porter, M.D., Colosi, L.M., 2020b. Is hydrothermal treatment coupled with carbon capture and storage an energy-producing negative emissions technology?. Energy Convers. Manage. 203, 112252.
[22] Chong, J.W., Thangalazhy-Gopakumar, S., Tan, R.R., Aviso, K.B., Chemmangattuvalappil, N.G., 2022. Estimation of fast pyrolysis bio-oil properties from feedstock characteristics using rough-set-based machine learning. Int. J. Energy Res. 46(13), 19159-19176.
[23] Cortes, C., Vapnik, V., 1995. Support-vector networks. Mach. Learn. 20, 273-297.
[24] da Silva, C.M.S., Carneiro, A.D.C.O., Vital, B.R., Figueiró, C.G., de Freitas Fialho, L., de Magalhães, M.A., Carvalho, A.G., Cândido, W.L., 2018. Biomass torrefaction for energy purposes-definitions and an overview of challenges and opportunities in Brazil. Renew. Sust. Energy Rev. 82, 2426-2432.
[25] Dai, L., Wang, Y., Liu, Y., Ruan, R., He, C., Yu, Z., Jiang, L., Zeng, Z., Tian, X., 2019. Integrated process of lignocellulosic biomass torrefaction and pyrolysis for upgrading bio-oil production: a state-of-the-art review. Renew. Sust. Energy Rev. 107, 20-36.
[26] Ding, R., Wang, R., Ding, Y., Yin, W., Liu, Y., Li, J., Liu, J., 2020. Designing ai-aided analysis and prediction models for nonprecious metal electrocatalyst-based proton-exchange membrane fuel cells. Angew. Chemie. 132(43), 19335-19345.
[27] 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.
[28] 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.
[29] Elmaz, F., Yücel, Ö., 2020. Data-driven identification and model predictive control of biomass gasification process for maximum energy production. Energy. 195, 117037.
[30] Elmaz, F., Yücel, Ö., Mutlu, A.Y., 2020. Predictive modeling of biomass gasification with machine learning-based regression methods. Energy. 191, 116541.
[31] Fan, L., Zhang, H., Li, Jingjing, Wang, Y., Leng, L., Li, Jun, 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.
[32] Freitas, R.S., Lima, Á.P., Chen, C., Rochinha, F.A., Mira, D., 2022. Towards predicting liquid fuel physicochemical properties using molecular dynamics guided machine learning models. Fuel. 329, 125415.
[33] Ge, S., Shi, Y., Xia, C., Huang, Z., Manzo, M., Cai, L., Ma, H., Zhang, S., Jiang, J., Sonne, C., Lam, S.S., 2021. Progress in pyrolysis conversion of waste into value-added liquid pyro-oil, with focus on heating source and machine learning analysis. Energy Convers.

Please cite this article as: Li H., Chen J., Zhang W., Zhan H., He C., Yang Z., Peng H., Leng L. Machine-learning-aided thermochemical treatment of biomass: a review. Biofuel Research Journal 37 (2023) 1786-1809. DOI: 10.18331/BRJ2023.10.1.4

Manage. 245, 114638.
[34] Ghugare, S.B., Tiwary, S., Tambe, S.S., 2017. Computational intelligence based models for prediction of elemental composition of solid biomass fuels from proximate analysis. Int. J. Syst. Assur. Eng. Manage. 8, 2083-2096.
[35] Gu, C., Wang, X., Song, Q., Li, H., Qiao, Y., 2021. Prediction of gas-liquid-solid product distribution after solid waste pyrolysis process based on artificial neural network model. Int. J. Energy Res. 45(9), 13786-13800.
[36] 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.
[37] 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. Sust. Energy Rev. 39, 1127-1142.
[38] Hough, B.R., Beck, D.A., Schwartz, D.T., Pfaendtner, J., 2017. Application of machine learning to pyrolysis reaction networks: reducing model solution time to enable process optimization. Comput. Chem. Eng. 104, 56-63.
[39] Hu, Z., Yuan, Y., Li, X., Tu, Z., Dacres, O.D., 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.
[40] Huang, H., Yuan, X., 2015. Recent progress in the direct liquefaction of typical biomass. Prog. Energy Combust. Sci. 49, 59-80.
[41] Huang, H.J., Yuan, X.Z., Zhu, H.N., Li, H., Liu, Y., Wang, X.L., Zeng, G.M., 2013. Comparative studies of thermochemical liquefaction characteristics of microalgae, lignocellulosic biomass and sewage sludge. Energy. 56, 52-60.
[42] Ippolito, J.A., Cui, L., Kammann, C., Wrage-Mönnig, N., Estavillo, J.M., Fuertes-Mendizabal, T., Cayuela, M.L., Sigua, G., Novak, J., Spokas, K., Borchard, N., 2020. Feedstock choice, pyrolysis temperature and type influence biochar characteristics: a comprehensive meta-data analysis review. Biochar. 2, 421-438.
[43] Kambo, H.S., Dutta, A., 2015. A comparative review of biochar and hydrochar in terms of production, physico-chemical properties and applications. Renew. Sust. Energy Rev. 45, 359-378.
[44] Kan, T., Strezov, V., Evans, T.J., 2016. Lignocellulosic biomass pyrolysis: a review of product properties and effects of pyrolysis parameters. Renew. Sust. Energy Rev. 57, 1126-1140.
[45] Kanaujia, P.K., Sharma, Y.K., Garg, M.O., Tripathi, D., Singh, R., 2014. Review of analytical strategies in the production and upgrading of bio-oils derived from lignocellulosic biomass. J. Anal. Appl. Pyrolysis 105, 55-74.
[46] Kartal, F., Özveren, U., 2022a. Prediction of torrefied biomass properties from raw biomass. Renewable Energy. 182, 578-591.
[47] Kartal, F., Özveren, U., 2022b. Investigation of the chemical exergy of torrefied biomass from raw biomass by means of machine learning. Biomass Bioenergy. 159, 106383.
[48] Kartal, F., Özveren, U., 2022c. Prediction of activation energy for combustion and pyrolysis by means of machine learning. Therm. Sci. Eng. Prog. 33, 101346.
[49] Kartal, F., Özveren, U., 2021. An improved machine learning approach to estimate hemicellulose, cellulose, and lignin in biomass. Carbohydr. Polym. Technol. Appl. 2, 100148.
[50] Katongtung, 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.
[51] Ke, B., Nguyen, H., Bui, X.N., Bui, H.B., Choi, Y., Zhou, J., Moayedi, H., Costache, R., Nguyen-Trang, T., 2021a. Predicting the sorption efficiency of heavy metal based on the biochar characteristics, metal sources, and environmental conditions using various novel hybrid machine learning models. Chemosphere. 276, 130204.
[52] Ke, B., Nguyen, H., Bui, X.N., Bui, H.B., Nguyen-Thoi, T., 2021b. Prediction of the sorption efficiency of heavy metal onto biochar using a robust combination of fuzzy C-means clustering and back-propagation neural network. J. Environ. Manage. 293, 112808.
[53] Khan, M., Ullah, Z., Mašek, O., Naqvi, S.R., Khan, M.N.A., 2022. Artificial neural networks for the prediction of biochar yield: a comparative study of metaheuristic algorithms. Bioresour. Technol. 355, 127215.
[54] Kusdhany, M.I.M., Lyth, S.M., 2021. New insights into hydrogen uptake on porous carbon materials via explainable machine learning. Carbon. 179, 190-201.
[55] Lecun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature. 521(7553), 436-444.
[56] Lee, B.H., Thieu Trinh, V., Moon, H. Bin, Lee, J.H., Kim, H.T., Lee, J.W., Jeon, C.H., 2023. Physicochemical properties and pyrolysis behavior of petcoke with artificial neural network modeling. Fuel. 331, 125735.
[57] Leng, L., Yuan, X., Huang, H., Jiang, H., Chen, X., Zeng, G., 2014. The migration and transformation behavior of heavy metals during the liquefaction process of sewage sludge. Bioresour. Technol. 167, 144-150.
[58] Leng, L., Yuan, X., Huang, H., Peng, X., Chen, H., Wang, H., Wang, L., Chen, X., Zeng, G., 2015. Distribution behavior and risk assessment of metals in bio-oils produced by liquefaction/pyrolysis of sewage sludge. Environ. Sci. Pollut. Res. 22, 18945-18955.
[59] Leng, L., Zhou, W., 2018a. Chemical compositions and wastewater properties of aqueous phase (wastewater) produced from the hydrothermal treatment of wet biomass: a review. Energy Sources, Part A. 40(22), 2648-2659.
[60] Leng, L., Leng, S., Chen, J., Yuan, X., Li, J., Li, K., Wang, Y., Zhou, W., 2018b. The migration and transformation behavior of heavy metals during co-liquefaction of municipal sewage sludge and lignocellulosic biomass. Bioresour. Technol. 259, 156-163.
[61] Leng, L., Li, H., Yuan, X., Zhou, W., Huang, H., 2018c. Bio-oil upgrading by emulsification/microemulsification: a review. Energy. 161, 214-232.
[62] Leng, L., Li, J., Wen, Z., Zhou, W., 2018d. Use of microalgae to recycle nutrients in aqueous phase derived from hydrothermal liquefaction process. Bioresour. Technol. 256, 529-542.
[63] Leng, L., Bogush, A.A., Roy, A., Stegemann, J.A., 2019a. Characterisation of ashes from waste biomass power plants and phosphorus recovery. Sci. Total Environ. 690, 573-583.
[64] Leng, L., Huang, H., Li, H., Li, J., Zhou, W., 2019b. Biochar stability assessment methods: a review. Sci. Total Environ. 647, 210-222.
[65] Leng, L., Xu, X., Wei, L., Fan, L., Huang, H., Li, Jianan, Lu, Q., Li, Jun, Zhou, W., 2019c. Biochar stability assessment by incubation and modelling: methods, drawbacks and recommendations. Sci. Total Environ. 664, 11-23.
[66] Leng, L., Li, W., Li, H., Jiang, S., Zhou, W., 2020a. Cold Flow Properties of Biodiesel and the Improvement Methods: a Review. Energy Fuels. 34(9), 10364-10383.
[67] Leng, L., Xu, S., Liu, R., Yu, T., Zhuo, X., Leng, S., Xiong, Q., Huang, H., 2020b. Nitrogen containing functional groups of biochar: an overview. Bioresour. Technol. 298, 122286.
[68] Leng, L., Yang, L., Chen, J., Leng, S., Li, H., Li, H., Yuan, X., Zhou, W., Huang, H., 2020c. A review on pyrolysis of protein-rich biomass: Nitrogen transformation. Bioresour. Technol. 315, 123801.
[69] Leng, L., Zhang, W., Leng, S., Chen, J., Yang, L., Li, H., Jiang, S., Huang, H., 2020d. Bioenergy recovery from wastewater produced by hydrothermal processing biomass: progress, challenges, and opportunities. Sci. Total Environ. 748, 142383.
[70] Leng, L., Zhang, W., Peng, H., Li, H., Jiang, S., Huang, H., 2020e. Nitrogen in bio-oil produced from hydrothermal liquefaction of biomass: a review. Chem. Eng. J. 401, 126030.
[71] Leng, S., Leng, L., Chen, L., Chen, J., Chen, J., Zhou, W., 2020f. The effect of aqueous phase recirculation on hydrothermal liquefaction/carbonization of biomass: a review. Bioresour. Technol. 318, 124081.
[72] Leng, L., Xiong, Q., Yang, L., Li, H., Zhou, Y., Zhang, W., Jiang, S., Li, H., Huang, H., 2021a. An overview on engineering the surface area and porosity of biochar. Sci. Total Environ. 763, 144204.
[73] Leng, L., Yang, L., Chen, J., Hu, Y., Li, H., Li, H., Jiang, S., Peng, H., Yuan, X., Huang, H., 2021b. Valorization of the aqueous phase

Please cite this article as: Li H., Chen J., Zhang W., Zhan H., He C., Yang Z., Peng H., Leng L. Machine-learning-aided thermochemical treatment of biomass: a review. Biofuel Research Journal 37 (2023) 1786-1809. DOI: 10.18331/BRJ2023.10.1.4

produced from wet and dry thermochemical processing biomass : a review. J. Clean. Prod. 294, 126238.
[74] Leng, L., Yang, L., Leng, S., Zhang, W., Zhou, Y., Peng, H., Li, Hui, Hu, Y., Jiang, S., Li, Hailong, 2021c. A review on nitrogen transformation in hydrochar during hydrothermal carbonization of biomass containing nitrogen. Sci. Total Environ. 756, 143679.
[75] Leng, E., He, B., Chen, J., Liao, G., Ma, Y., Zhang, F., Liu, S., Jiaqiang, E., 2021d. Prediction of three-phase product distribution and bio-oil heating value of biomass fast pyrolysis based on machine learning. Energy. 236, 121401.
[76] Leng, L., Liu, R., Xu, S., Mohamed, B.A., Yang, Z., Hu, Y., Chen, J., Zhao, S., Wu, Z., Peng, H., Li, Hui, Li, Hailong, 2022a. An overview of sulfur-functional groups in biochar from pyrolysis of biomass. J. Environ. Chem. Eng. 10(2), 107185.
[77] 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(1), 63.
[78] 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.
[79] 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.
[80] Li, H., Wang, S., Yuan, X., Xi, Y., Huang, Z., Tan, M., Li, C., 2018a. The effects of temperature and color value on hydrochars' properties in hydrothermal carbonization. Bioresour. Technol. 249, 574-581.
[81] Li, L., Wang, Y., Xu, J., Flora, J.R., Hoque, S., Berge, N.D., 2018b. Quantifying the sensitivity of feedstock properties and process conditions on hydrochar yield, carbon content, and energy content. Bioresour. Technol. 262, 284-293.
[82] 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.
[83] Li, J., Pan, L., Suvarna, M., Wang, X., 2021a. Machine learning aided supercritical water gasification for H_2-rich syngas production with process optimization and catalyst screening. Chem. Eng. J. 426, 131285.
[84] 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.
[85] 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.
[86] 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.
[87] Li, J., Suvarna, M., Li, L., Pan, L., Pérez-Ramírez, J., Ok, Y.S., Wang, X., 2022a. A review of computational modeling techniques for wet waste valorization: research trends and future perspectives. J. Clean. Prod. 367, 133025.
[88] Li, Y., Gupta, R., You, S., 2022b. Machine learning assisted prediction of biochar yield and composition via pyrolysis of biomass. Bioresour. Technol. 359, 127511.
[89] 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.
[90] Liao, M., Kelley, S.S., Yao, Y., 2019. Artificial neural network based modeling for the prediction of yield and surface area of activated carbon from biomass. Biofuels, Bioprod. Biorefin. 13(4), 1015-1027.
[91] Liao, M., Yao, Y., 2021. Applications of artificial intelligence-based modeling for bioenergy systems: a review. GCB Bioenergy. 13(5), 774-802.
[92] Liu, W.J., Jiang, H., Yu, H.Q., 2015. Development of biochar-based functional materials: toward a sustainable platform carbon material. Chem. Rev. 115(22), 12251-12285.
[93] Mahmood, A., Wang, J.L., 2021. Machine learning for high performance organic solar cells: current scenario and future prospects. Energy Environ. Sci. 14(1), 90-105.
[94] Manatura, K., Chalermsinsuwan, B., Kaewtrakulchai, N., Kwon, E.E., Chen, W.H., 2023. Machine learning and statistical analysis for biomass torrefaction: a review. Bioresour. Technol. 369, 128504.
[95] Mari Selvam, S., Balasubramanian, P., 2022. Influence of biomass composition and microwave pyrolysis conditions on biochar yield and its properties: a machine learning approach. Bioenergy Res.
[96] Meena, M., Shubham, S., Paritosh, K., Pareek, N., Vivekanand, V., 2021. Production of biofuels from biomass: predicting the energy employing artificial intelligence modelling. Bioresour. Technol. 340, 125642.
[97] Mohamed, B.A., Bilal, M., Salama, E.S., Periyasamy, S., Fattah, I.R., Ruan, R., Awasthi, M.K., Leng, L., 2022. Phenolic-rich bio-oil production by microwave catalytic pyrolysis of switchgrass: experimental study , life cycle assessment , and economic analysis. J. Clean. Prod. 366, 132668.
[98] Molino, A., Chianese, S., Musmarra, D., 2016. Biomass gasification technology: the state of the art overview. J. Energy Chem. 25(1), 10-25.
[99] 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.
[100] Mutlu, A.Y., Yucel, O., 2018. An artificial intelligence based approach to predicting syngas composition for downdraft biomass gasification. Energy. 165, 895-901.
[101] Ögren, Y., Tóth, P., Garami, A., Sepman, A., Wiinikka, H., 2018. Development of a vision-based soft sensor for estimating equivalence ratio and major species concentration in entrained flow biomass gasification reactors. Appl. Energy. 226, 450-460.
[102] Olatunji, o.o., akinlabi, s., madushele, n., adedeji, p.a., 2019. estimation of the elemental composition of biomass using hybrid adaptive neuro-fuzzy inference system. Bioenergy Res. 12, 642-652.
[103] Onsree, T., Tippayawong, N., 2021. Machine learning application to predict yields of solid products from biomass torrefaction. Renewable Energy. 167, 425-432.
[104] Ozbas, E.E., Aksu, D., Ongen, A., Aydin, M.A., Ozcan, H.K., 2019. Hydrogen production via biomass gasification, and modeling by supervised machine learning algorithms. Int. J. Hydrogen Energy. 44(32), 17260-17268.
[105] 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. Technol. 56(7), 4187-4198.
[106] Pathy, A., Meher, S., Balasubramanian, P., 2020. Predicting algal biochar yield using eXtreme Gradient Boosting (XGB) algorithm of machine learning methods. Algal Res. 50, 102006.
[107] Pavlov, Y.L., 2001. Random forests. De Gruyter. Berlin, Boston.
[108] Peng, C., Zhai, Y., Zhu, Y., Wang, Tengfei, Xu, B., Wang, Tao, Li, C., Zeng, G., 2017. Investigation of the structure and reaction pathway of char obtained from sewage sludge with biomass wastes, using hydrothermal treatment. J. Clean. Prod. 166, 114-123.
[109] Perera, S.M., Wickramasinghe, C., Samarasiri, B.K.T., Narayana, M., 2021. Modeling of thermochemical conversion of waste biomass-a comprehensive review. Biofuel Res. J. 8(4), 1481-1528.
[110] 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. Energy Environ. Sci. 1(1), 32-56.
[111] Qasem, M.A.A., Al-Mutairi, E.M., Jameel, A.G.A., 2023. Smoke point prediction of oxygenated fuels using neural networks. Fuel. 332, 126026.
[112] Quinlan, J.R., 1986. Induction of decision trees. Mach. Learn. 1, 81-106.
[113] 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. 344(6185), 1246843.

Please cite this article as: Li H., Chen J., Zhang W., Zhan H., He C., Yang Z., Peng H., Leng L. Machine-learning-aided thermochemical treatment of biomass: a review. Biofuel Research Journal 37 (2023) 1786-1809. DOI: 10.18331/BRJ2023.10.1.4

[114] Sawangkeaw, R., Ngamprasertsith, S., 2013. A review of lipid-based biomasses as feedstocks for biofuels production. Renew. Sust. Energy Rev. 25, 97-108.
[115] Sezer, S., Kartal, F., Özveren, U., 2022. Prediction of combustion reactivity for lignocellulosic fuels by means of machine learning. J. Therm. Anal. Calorim. 147(17), 9793-9809.
[116] Sezer, S., Kartal, F., Özveren, U., 2021. Prediction of chemical exergy of syngas from downdraft gasifier by means of machine learning. Therm. Sci. Eng. Prog. 26, 101031.
[117] Sezer, S., Özveren, U., 2021. Investigation of syngas exergy value and hydrogen concentration in syngas from biomass gasification in a bubbling fluidized bed gasifier by using machine learning. Int. J. Hydrogen Energy. 46(39), 20377-20396.
[118] 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.
[119] Shahbeik, H., Rafiee, S., Shafizadeh, A., Jeddi, D., Jafary, T., Lam, S.S., Pan, J., Tabatabaei, M., Aghbashlo, M., 2022. Characterizing sludge pyrolysis by machine learning: towards sustainable bioenergy production from wastes. Renewable Energy. 199, 1078-1092.
[120] 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(24), 12739-12757.
[121] Sigmund, G., Gharasoo, 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. Technol. 54(7), 4583-4591.
[122] Su, Y., Zhu, W., Gong, M., Zhou, H., Fan, Y., Amuzu-Sefordzi, B., 2015. Interaction between sewage sludge components lignin (phenol) and proteins (alanine) in supercritical water gasification. Int. J. Hydrogen Energy. 40(30), 9125-9136.
[123] Sun, Y., Zhang, Yuyao, Lu, L., Wu, Y., Zhang, Y., Kamran, M.A., Chen, B., 2022. The application of machine learning methods for prediction of metal immobilization remediation by biochar amendment in soil. Sci. Total Environ. 829, 154668.
[124] 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.
[125] Tang, Q., Chen, Y., Yang, H., Liu, M., Xiao, H., Wu, Z., Chen, H., Naqvi, S.R., 2020. Prediction of bio-oil yield and hydrogen contents based on machine learning method: effect of biomass compositions and pyrolysis conditions. Energy Fuels. 34(9), 11050-11060.
[126] Tao, J., Liang, R., Li, J., Yan, B., Chen, G., Cheng, Z., Li, W., Lin, F., Hou, L., 2020. Fast characterization of biomass and waste by infrared spectra and machine learning models. J. Hazard. Mater. 387, 121723.
[127] Hoang Pham, L.K.,Tran, T.T.V., Kongparakul, S., Reubroycharoen, P., Ding, M., Guan, G., Vo, D.V.N., Jaiyong, P., Youngvises, N., Samart, C., 2021. Data-driven prediction of biomass pyrolysis pathways toward phenolic and aromatic products. J. Environ. Chem. Eng. 9(2), 104836.
[128] Tuck, C.O., Pérez, E., Horváth, I.T., Sheldon, R.A., Poliakoff, M., 2012. Valorization of biomass: deriving more value from waste. Science. 337(6095), 695-699.
[129] Ullah, Z., khan, M., Naqvi, S.R., 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.
[130] Ullah, Z., Khan, M., Naqvi, S.R., Khan, M.N.A., Farooq, W., Anjum, M.W., Yaqub, M.W., AlMohamadi, H., Almomani, F., 2022. An integrated framework of data-driven, metaheuristic, and mechanistic modeling approach for biomass pyrolysis. Process Saf. Environ. Prot. 162, 337-345.
[131] Umenweke, G.C., Afolabi, I.C., Epelle, E.I., Okolie, J.A., 2022. Machine learning methods for modeling conventional and hydrothermal gasification of waste biomass: a review. Bioresour. Technol. Rep. 17, 100976.
[132] 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.
[133] Vassilev, S. V., Baxter, D., Andersen, L.K., Vassileva, C.G., 2010. An overview of the chemical composition of biomass. Fuel. 89(5), 913-933.
[134] Vassilev, S.V., Baxter, D., Andersen, L.K., Vassileva, C.G., Morgan, T.J., 2012. An overview of the organic and inorganic phase composition of biomass. Fuel. 94, 1-33.
[135] Wang, S., Dai, G., Yang, H., Luo, Z., 2017. Lignocellulosic biomass pyrolysis mechanism: a state-of-the-art review. Prog. Energy Combust. Sci. 62, 33-86.
[136] 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.
[137] Wang, F., Harindintwali, J.D., Yuan, Zhizhang, Wang, M., Wang, Faming, Li, S., Yin, Z., Huang, L., Fu, Y., Li, L., Chang, S.X., Zhang, L., Rinklebe, J., Yuan, Zuoqiang, Zhu, Q., Xiang, L., Tsang, D.C.W., Xu, L., Jiang, X., Liu, J., Wei, N., Kästner, M., Zou, Y., Ok, Y.S., Shen, J., Peng, D., Zhang, W., Barceló, D., Zhou, Y., Bai, Z., Li, B., Zhang, B., Wei, K., Cao, H., Tan, Z., Zhao, L. bin, He, X., Zheng, J., Bolan, N., Liu, X., Huang, C., Dietmann, S., Luo, M., Sun, N., Gong, J., Gong, Y., Brahushi, F., Zhang, T., Xiao, C., Li, X., Chen, W., Jiao, N., Lehmann, J., Zhu, Y.G., Jin, H., Schäffer, A., Tiedje, J.M., Chen, J.M., 2021a. Technologies and perspectives for achieving carbon neutrality. Innovation. 2(4), 100180.
[138] Wang, Y., Liao, Z., Mathieu, S., Bin, F., Tu, X., 2021b. Prediction and evaluation of plasma arc reforming of naphthalene using a hybrid machine learning model. J. Hazard. Mater. 404, 123965.
[139] Wang, S., Shi, Z., Jin, Y., Zaini, I.N., Li, Y., Tang, C., Mu, W., Wen, Y., Jiang, J., Jönsson, P.G., Yang, W., 2022. A machine learning model to predict the pyrolytic kinetics of different types of feedstocks. Energy Convers. Manage. 260, 115613.
[140] Xia, C., Cai, L., Zhang, H., Zuo, L., Shi, S.Q., Lam, S.S., 2021. A review on the modeling and validation of biomass pyrolysis with a focus on product yield and composition. Biofuel Res. J. 8(1), 1296-1315.
[141] Xing, J., Luo, K., Wang, H., Fan, J., 2019. Estimating biomass major chemical constituents from ultimate analysis using a random forest model. Bioresour. Technol. 288, 121541.
[142] Yan, Y., Mattisson, T., Moldenhauer, P., Anthony, E.J., Clough, P.T., 2020. Applying machine learning algorithms in estimating the performance of heterogeneous, multi-component materials as oxygen carriers for chemical-looping processes. Chem. Eng. J. 387, 124072.
[143] Xu, S., Chen, J., Peng, H., Leng, S., Li, Hui, Qu, W., Hu, Y., Li, Hailong, Jiang, S., Zhou, W., Leng, L., 2021a. Effect of biomass type and pyrolysis temperature on nitrogen in biochar, and the comparison with hydrochar. Fuel. 291, 120128.
[144] Xu, Y., Liu, Xin, Cao, X., Huang, C., Liu, E., Qian, S., Liu, Xingchen, Wu, Y., Dong, F., Qiu, C.W., Qiu, J., Hua, K., Su, W., Wu, J., Xu, H., Han, Y., Fu, C., Yin, Z., Liu, M., Roepman, R., Dietmann, S., Virta, M., Kengara, F., Zhang, Z., Zhang, Lifu, Zhao, T., Dai, J., Yang, J., Lan, L., Luo, M., Liu, Z., An, T., Zhang, B., He, X., Cong, S., Liu, Xiaohong, Zhang, W., Lewis, J.P., Tiedje, J.M., Wang, Q., An, Z., Wang, Fei, Zhang, Libo, Huang, T., Lu, C., Cai, Z., Wang, Fang, Zhang, J., 2021b. Artificial intelligence: a powerful paradigm for scientific research. Innovation. 2(4), 100179.
[145] 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. Pyrolysis. 167, 105678.
[146] Yang, K., Wu, K., Zhang, H., 2022a. Machine learning prediction of the yield and oxygen content of bio-oil via biomass characteristics and pyrolysis conditions. Energy. 254, 124320.
[147] Yang, L., Wang, G., Liu, T., Wan, Y., Peng, H., Leng, L., Zhong, Q., 2022b. Experimental and ReaxFF molecular dynamic study on pyrolysis of phenylalanine. Fuel. 324, 124690.
[148] Yang, X., Yuan, C., He, S., Jiang, D., Cao, B., Wang, S., 2023. Machine learning prediction of specific capacitance in biomass

Please cite this article as: Li H., Chen J., Zhang W., Zhan H., He C., Yang Z., Peng H., Leng L. Machine-learning-aided thermochemical treatment of biomass: a review. Biofuel Research Journal 37 (2023) 1786-1809. DOI: 10.18331/BRJ2023.10.1.4

derived carbon materials : effects of activation and biochar characteristics. Fuel. 331, 125718.
[149] Yapıcı, E., Akgün, H., Özkan, K.E.M.A.L., Günkaya, Z., Özkan, A., Banar, M., 2022. Prediction of gas product yield from packaging waste pyrolysis: support vector and Gaussian process regression models. Int. J. Environ. Sci. Technol. 20, 461-476.
[150] Yuan, X., Leng, L., Huang, H., Chen, X., Wang, H., Xiao, Z., Zhai, Y., Chen, H., Zeng, G., 2015. Speciation and environmental risk assessment of heavy metal in bio-oil from liquefaction/pyrolysis of sewage sludge. Chemosphere. 120, 645-652.
[151] Yuan, X., Suvarna, M., Low, S., Dissanayake, P.D., Lee, K.B., Li, J., Wang, X., Ok, Y.S., 2021. Applied machine learning for prediction of CO_2 adsorption on biomass waste-derived porous carbons. Environ. Sci. Technol. 55(17), 11925-11936.
[152] 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.
[153] Zhang, J., Liu, J., Evrendilek, F., Zhang, X., Buyukada, M., 2019. TG-FTIR and Py-GC/MS analyses of pyrolysis behaviors and products of cattle manure in CO_2 and N_2 atmospheres: kinetic, thermodynamic, and machine-learning models. Energy Convers. Manage. 195, 346-359.
[154] Zhang, K., Zhong, S., Zhang, H., 2020. Predicting aqueous adsorption of organic compounds onto biochars, carbon nanotubes, granular activated carbons, and resins with machine learning. Environ. Sci. Technol. 54(11), 7008-7018.
[155] Zhang, T., Cao, D., Feng, X., Zhu, J., Lu, X., Mu, L., Qian, H., 2022. Machine learning prediction of bio-oil characteristics quantitatively relating to biomass compositions and pyrolysis conditions. Fuel. 312, 122812.
[156] Zhang, W., Chen, Q., Chen, J., Xu, D., Zhan, H., Peng, H., Pan, J., Vlaskin, M., Leng, L., Li, H., 2023. Machine learning for hydrothermal treatment of biomass : a review. Bioresour. Technol. 370, 128547.
[157] 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.
[158] 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.
[159] Zhu, X., Li, Y., Wang, X., 2019a. Machine learning prediction of biochar yield and carbon contents in biochar based on biomass characteristics and pyrolysis conditions. Bioresour. Technol. 288, 121527.
[160] Zhu, X., Wang, X., Ok, Y.S., 2019b. The application of machine learning methods for prediction of metal sorption onto biochars. J. Hazard. Mater. 378, 120727.
[161] Zhu, X., Tsang, D.C.W., Wang, L., Su, Z., Hou, D., Li, L., Shang, J., 2020. Machine learning exploration of the critical factors for CO_2 adsorption capacity on porous carbon materials at different pressures. J. Clean. Prod. 273, 122915.
[162] Zhou, X., Zhao, J., Chen, M., Wu, S., Zhao, G., Xu, S., 2022a. Effects of hydration parameters on chemical properties of biocrudes based on machine learning and experiments. Bioresour. Technol. 350, 126923.
[163] Zhou, X., Zhao, J., Chen, M., Zhao, G., Wu, S., 2022b. Influence of catalyst and solvent on the hydrothermal liquefaction of woody biomass. Bioresour. Technol. 346, 126354.
[164] Zhu, X., He, M., Sun, Y., Xu, Z., Wan, Z., Hou, D., Alessi, D.S., Tsang, D.C.W., 2022a. Insights into the adsorption of pharmaceuticals and personal care products (PPCPs) on biochar and activated carbon with the aid of machine learning. J. Hazard. Mater. 423, 127060.
[165] Zhu, L.T., Chen, X.Z., Ouyang, B., Yan, W.C., Lei, H., Chen, Z., Luo, Z.H., 2022b. Review of Machine Learning for Hydrodynamics, Transport, and Reactions in Multiphase Flows and Reactors. Ind. Eng. Chem. Res. 61(28), 9901-9949.

Prof. Lijian Leng obtained his doctoral degree at Hunan University, China. He is currently working as a Professor at the School of Energy Science and Engineering at Central South University, China. He has authored over 100 peer-review journal papers. He was awarded "Highly Cited Researchers" in 2021 and 2022 by Clarivate. His research focuses on biomass/biowastes conversion and valorization by thermochemical technologies: (i) production of bio-oil & chemical precursors and bio-char & material precursors by hydrothermal liquefaction or pyrolysis and their upgradings; (ii) engineering of bio-oil & biochar with target properties or compositions and engineering of biochar for various applications aided by machine learning (ML); (iii) mechanistic study of biomass/organic waste conversion by density functional theory and ReaxFF molecular dynamic. His research profile on Google Scholar can be found at the following link: https://scholar.google.com/citations?hl=en&user=dGcVcqwAAAAJ.

Prof. Hailong Li obtained his PhD (2011) from Huazhong University of Science and Technology. After a short postdoctoral training at Central South University, he worked as an associate professor at Central South University from 2012 to 2015. Between 2015 and 2017, he worked as a post-doctoral fellow at the University of Hong Kong. Since 2015, he has held a full professor position at the department of thermal engineering, Central South University. His research interests are air pollution control, CCUS, bioenergy, waste management, etc. He has published >100 SCI-indexed journal papers and filed >20 patents. He was awarded the highly cited researcher 2021 by Clarivate, Hong Kong Scholar, and the Excellent Young Scholar of Hunan province, China. His research profile is available at https://orcid.org/0000-0003-0652-6655.

Please cite this article as: Li H., Chen J., Zhang W., Zhan H., He C., Yang Z., Peng H., Leng L. Machine-learning-aided thermochemical treatment of biomass: a review. Biofuel Research Journal 37 (2023) 1786-1809. DOI: 10.18331/BRJ2023.10.1.4

Jiefeng Chen is a doctoral student from the School of Energy Science and Engineering at Central South University, China. He has a Master's degree in Environmental Engineering from Nanchang University, China. His research interests include (1) the application of biochar/hydrochar; (2) heavy metal pollution control; (3) machine learning. His research profile can be found at the following link: https://www.researchgate.net/profile/Jiefeng-Chen-3.

Weijin Zhang is a master's student at the School of Energy Science and Engineering at Central South University, China. His major research interests are the applications of machine learning in the hydrothermal treatment of biomass. His research profile can be found at the following link: https://www.researchgate.net/profile/Weijin-Zhang-3.

Dr. Zequn Yang is an associate professor in the School of Energy Science and Engineering at Central South University, China. He has a Ph.D. degree from The University of Hong Kong. His research interests include (1) Carbon Dioxide Capture, Storage, and Conversion and (2) Heavy Metal Remediation and Recycling from Industrial Sources. His research profile can be found at the following link: https://www.researchgate.net/profile/Zequn-Yang-2

Dr. Haoyi Peng is an associate professor in the School of Energy Science and Engineering at Central South University, China. His major research interests are biomass energy efficient conversion and utilization technology, efficient, clean combustion technology, industrial energy saving technology, etc. His research profile can be found at the following link: https://www.scopus.com/authid/detail.uri?authorId=8631471100.

Dr. Chao He is an Associate Professor in Bio and Circular Economy at Tampere University, Finland. Meanwhile, Dr. He is holding Academy Research Fellowship from the Academy of Finland. Currently, Dr. He also acts as a board member of the Finnish Thermal Energy Research Association, a university member of the International Water Association, an Associate Editor in the Journal of Environmental Chemical Engineering (Elsevier), a Social Media Editor in Separation and Purification Technology (Elsevier), and editorial board member in Waste (MDPI). His research group focuses on circular economy innovation by developing low-carbon clean technologies for waste management and carbon neutrality based on multidisciplinary sciences in environmental and energy engineering, thermochemical and physicochemical processes, chemical reaction engineering, materials design, and applied catalysis. Dr. He has been exploring energy-efficient processes, technologies, and catalysts for energy conversion and nutrients recycling of sewage sludge, sustainable disposal of municipal solid wastes, bio-refinery of biomass wastes, and technology-critical elements recovery from hyperaccumulators in terms of fundamental sciences, key reaction processes, industrial applications, and commercialization. His research profile is available at https://www.researchgate.net/profile/Chao-He-17.

Dr. Hao Zhan is a lecturer at the School of Energy Science and Engineering at Central South University, People's Republic of China. He received his Ph.D. from the University of Chinese Academy of Sciences. His research interests include the value-added thermochemical utilization of biowastes, the application of bio-based fuel charcoal and carbon material, biomass gasification, and pyrolytic polygeneration. He has previously published over 40 peer-reviewed publications, and his research profile on ResearchGate can be found at the following link: https://www.researchgate.net/profile/Hao-Zhan.

Please cite this article as: Li H., Chen J., Zhang W., Zhan H., He C., Yang Z., Peng H., Leng L. Machine-learning-aided thermochemical treatment of biomass: a review. Biofuel Research Journal 37 (2023) 1786-1809. DOI: 10.18331/BRJ2023.10.1.4

Supplementary Information

Table S1.
Machine learning aided prediction of the yield of thermochemical treatment products.

Biomass	Inputs [a]	Data number	Outputs [b]	Predictive performance, test data			Reference
				Algorithms [c]	R^2	RMSE [d]	
Wet thermochemical treatment (hydrothermal treatment)							
Biomass	EC, AR, OC	108	Oil	RF	0.85	5.83	Leng et al. (2022b)
	EC, AR, BC, OC	91			0.92	3.31	
Biomass	EC, OC, solvent	448	Oil	DRT	0.82	8.07	Li et al. (2021b)
				RF	0.89	6.17	
				GBR	0.90	6.14	
	EC, BC, OC, solvent	382		DRT	0.84	7.04	
				RF	0.87	6.41	
				GBR	0.87	6.39	
Algae	EC, OC	310	Oil	GBR	0.86	5.52	Zhang et al. (2021)
				RF	0.85	6.16	
	AR, OC	310		GBR	0.88	5.07	
				RF	0.85	5.08	
	BC, OC	310		GBR	0.90	4.75	
				RF	0.86	6.06	
	EC, AR, BC, OC	310		GBR	0.90	4.69	
				RF	0.87	5.72	
Biomass	BC, OC	325	Oil	XGB	0.81	6.56	Katongtung et al. (2022)
	EC, AR, OC				0.82	6.28	
	EC, AR, BC, OC				0.90	4.77	
Biomass	EC, OC		Oil	GPR	0.95	NA [e]	Shafizadeh et al. (2022)
Biomass	BC, OC	525	Oil	RF	0.78	8.07	Cheng et al. (2022)
				XGB	0.77	8.26	
Biomass	EC, solvent, OC, catalyst	488	Oil	RF	0.90	6.03	Castro Garcia et al. (2022)
Biomass	EC, OC	650	Oil	GPR	0.95	NA	Shafizadeh et al. (2022)
				NNR	0.84		
				GAM	0.98		
				SVR	0.88		
Biomass	PA, EC, OC	248	Char	SVR	0.88	7.83	Li et al. (2020)
Biomass	PA, EC, BC, OC	649	Char	RF	0.95	4.06	Li et al. (2018)
Biomass	PA, EC, AR, OC	248	Char	DNN	0.90	7.05	Li et al. (2021c)
Biomass	EC, solvent, OC, catalyst	488	Char	RF	0.94	3.84	Castro Garcia et al. (2022)
Biomass	EC, OC	340	Char	GPR	0.94	NA	Shafizadeh et al. (2022)
				NNR	0.80		
				GAM	0.85		
				SVR	0.89		
Biomass	PA, EC, OC	296	Char	NN	0.86	NA	Mu et al. (2022)
				PSO-NN	0.86		

Please cite this article as: Li H., Chen J., Zhang W., Zhan H., He C., Yang Z., Peng H., Leng L. Machine-learning-aided thermochemical treatment of biomass: a review. Biofuel Research Journal 37 (2023) 1786-1809. DOI: 10.18331/BRJ2023.10.1.4

Table S1.
Continued.

Biomass	Inputs [a]	Data number	Outputs [b]	Predictive performance, test data			Reference
				Algorithms [c]	R^2	RMSE [d]	
Biomass	EC, OC	216	Gas	GPR	0.86	NA	Shafizadeh et al. (2022)
				NNR	0.90		
				GAM	0.98		
		197	Aqueous	SVR	0.93		
				GPR	0.87		
				NN	0.80		
				GAM	0.98		
				SVR	0.76		
Dry thermochemical treatment							
Biomass	BC, OC	245	Char	RF	0.85	3.50	Zhu et al. (2019)
	EC, OC				0.80	3.96	
	EC, BC, OC				0.85	3.40	
Biomass	PA, OC	115	Char	ANN	0.98	0.06	Abdulsalam et al. (2020)
Biomass	PA, EC, OC	165	Char	SVR	0.88	5.86	Li et al. (2020)
Algae	PA, EC, AR, OC	91	Char	XGB	0.74	NA	Pathy et al. (2020)
	EC, OC				0.73		
Biomass	PA, EC, OC	800	Char	KRR	0.86	0.08	Onsree and Tippayawong (2021)
				GTB	0.87	0.07	
Biomass	PA, EC, BC, OC	122	Char	RF	0.94	2.72	Leng et al. (2021)
	EC, BC, OC				0.94	2.75	
	PA, EC, OC				0.93	2.92	
	PA, BC, OC				0.94	2.57	
	PA, OC				0.94	2.80	
	BC, OC				0.95	2.56	
	EC, OC				0.90	3.59	
Biomass	PA, EC, metal contents, OC		Char	ANN	0.98	1.61	Gu et al. (2021)
Biomass-plastic mixture	PA, EC, OC	94	Char	DNN	0.93	2.96	Alabdrabalnabi (2021)
				XGB	0.91	3.18	
Biomass	PA, EC, OC, gasifier	312	Char	ANN	0.71	0.30	Ascher et al. (2022)
Biomass	PA, EC, BC, OC	226	Char	MLP-NN	0.96	3.4	Li et al. (2022)
				ANFIS	0.88	4.9	
Biomass	PA, BC, OC	402	Char	ANN-GWO	0.85	2.82	Khan et al. (2022)
				ANN-RA 1	0.92	1.89	
				ANN-RA 2	0.93	1.74	
				ANN-SCA	0.84	2.86	
				ANN-GA	0.92	1.87	
				ANN-PSO	0.92	1.82	
				ANN-GWO	0.87	2.52	
				ANN-RA 1	0.70	3.55	

Please cite this article as: Li H., Chen J., Zhang W., Zhan H., He C., Yang Z., Peng H., Leng L. Machine-learning-aided thermochemical treatment of biomass: a review. Biofuel Research Journal 37 (2023) 1786-1809. DOI: 10.18331/BRJ2023.10.1.4

Table S1.
Continued.

Biomass	Inputs [a]	Data number	Outputs [b]	Predictive performance, test data			Reference
				Algorithms [c]	R^2	RMSE [d]	
				ANN-RA 2	0.88	2.23	
				ANN-SCA	0.78	3.43	
				ANN-GA	0.88	2.25	
				ANN-PSO	0.88	2.18	
Biomass	PA, EC, OC, gasifier	312	Oil	ANN	0.92	0.07	Ascher et al. (2022)
Biomass	PA, OC	264	Oil	RF	0.92	2.13	Tang et al. (2020)
	EC, OC				0.87	3.05	
Biomass	PA, EC, BC, OC	122	Oil	RF	0.94	3.40	Leng et al. (2021)
	EC, BC, OC				0.92	3.89	
	PA, EC, OC				0.94	3.29	
	PA, BC, OC				0.94	3.43	
	PA, OC				0.94	3.69	
	BC, OC				0.95	3.28	
	EC, OC				0.92	4.03	
Biomass	PA, EC, metal contents, OC	NA	Oil	ANN	0.80	2.28	Gu et al. (2021)
Biomass-plastic mixture	PA, EC, OC	96	Oil	DNN	0.88	4.44	Alabdrabalnabi (2021)
				XGB	0.91	3.56	
Biomass	PA, OC	263	Oil	RF	0.98	1.06	Ullah et al. (2021)
				SVR	0.96	1.62	
				DRT	0.96	1.69	
				RF	0.99	0.56	
	EC, OC			SVM	0.96	1.53	
				DRT	0.91	2.18	
Biomass	EC, AR, BC, OC	282	Oil	RF	0.89	2.89	Zhang et al. (2022)
	EC, OC				0.93	2.33	
	BC, OC				0.88	3.11	
	PA, OC				0.84	3.98	
Biomass	PA, OC	292	Oil	RF	0.93		Yang et al. (2022)
	EC, OC				0.83	NA	
	BC, OC				0.85		
Biomass	PA, EC, metal contents, OC	NA	Gas	ANN	0.92	2.75	Gu et al. (2021)
Biomass	PA, EC, OC	194	Gas	RF	0.85	3.80	Tang et al. (2021)
				SVM	0.81	4.37	
Biomass	PA, EC, BC, OC	122	Gas	RF	0.91	3.42	Leng et al. (2021)
	EC, BC, OC				0.91	3.46	
	PA, EC, OC				0.92	3.21	
	PA, BC, OC				0.92	3.27	
	PA, OC				0.91	3.52	
	BC, OC				0.93	3.10	

Please cite this article as: Li H., Chen J., Zhang W., Zhan H., He C., Yang Z., Peng H., Leng L. Machine-learning-aided thermochemical treatment of biomass: a review. Biofuel Research Journal 37 (2023) 1786-1809. DOI: 10.18331/BRJ2023.10.1.4

Table S1.
Continued.

Biomass	Inputs [a]	Data number	Outputs [b]	Predictive performance, test data			Reference
				Algorithms [c]	R^2	RMSE [d]	
	EC, OC				0.90	3.70	
Biomass	PA, EC, OC, gasifier	312	Gas	ANN	0.96	0.08	Ascher et al. (2022)

[a] EC: elemental compositions of biomass; AR: atomic ratios of biomass; BC: biological compositions of biomass; PA: proximate analysis of biomass; OC: operating conditions of the thermochemical process.
[b] Yields of oil, char, gas, or aqueous phases.
[c] RF: random forest; NN: neural network; DRT: decision regression tree; GBR: gradient boosting regression; GBDT: gradient boosting decision tree; GAM: generalized additive model; GPR: gaussian process regression; GTB: gradient tree boosting; XGB: extreme gradient boosting; KRR: kernel ridge regression; SVR: support vector regression; SVM: support vector machine; MLP: multi-layer perceptron; GA: genetic algorithm; GWO: grey wolf optimization; RA: Rao algorithms; PSO: particle swarm optimization; SCA: sine cosine algorithm; ANFIS: artificial neuro-fuzzy inference system; DNN: deep neural network; ANN: artificial neural network; MARS: multivariate adaptive regression splines.
[d] Root mean square error.
[e] NA: not available or data available were calculated based on a unified dataset.

Table S2.
Machine learning aided prediction of the elemental compositions, proximate analysis, and gas compositions of thermochemical treatment products.

Biomass	Inputs [a]	Data number	Outputs [b]	Predictive performance, test data			Reference
				Algorithms [c]	R^2	RMSE [d]	
Wet thermochemical treatment (hydrothermal treatment)							
Biomass	EC, AR, OC	86	N_oil	RF	0.94	0.48	Leng et al. (2022b)
	EC, AR, BC, OC	69			0.95	0.47	
Algae	EC, OC	310	N_oil	GBR	0.89	1.71	Zhang et al. (2021)
				RF	0.87	1.85	
			O_oil	GBR	0.88	0.77	
				RF	0.87	0.78	
	AR, OC	310	N_oil	GBR	0.87	1.88	
				RF	0.86	1.97	
			O_oil	GBR	0.87	0.80	
				RF	0.86	0.82	
	BC, OC	310	N_oil	GBR	0.89	1.71	
				RF	0.87	1.85	
			O_oil	GBR	0.88	0.77	
				RF	0.87	0.78	
	EC, AR, BC, OC	310	N_oil	GBR	0.90	1.68	
				RF	0.87	1.91	
			O_oil	GBR	0.90	0.68	
				RF	0.88	0.74	
Biomass	EC, OC	268	C_oil	GPR	0.91	NA [e]	Shafizadeh et al. (2022)
			H_oil		0.98		
			O_oil		0.92		
			N_oil		0.99		
		263	S_oil		0.99		

Please cite this article as: Li H., Chen J., Zhang W., Zhan H., He C., Yang Z., Peng H., Leng L. Machine-learning-aided thermochemical treatment of biomass: a review. Biofuel Research Journal 37 (2023) 1786-1809. DOI: 10.18331/BRJ2023.10.1.4

Table S2.
Continued.

Biomass	Inputs [a]	Data number	Outputs [b]	Predictive performance, test data			Reference
				Algorithms [c]	R^2	RMSE [d]	
		268	C_oil	NNR	0.68	NA [e]	Shafizadeh et al. (2022)
			H_oil		0.84		
			O_oil		0.79		
			N_oil		0.95		
		263	S_oil		0.93		
		268	C_oil	GAM	0.89		
			H_oil		0.90		
			O_oil		0.99		
			N_oil		0.76		
		263	S_oil		0.98		
		268	C_oil	SVR	0.63		
			H_oil		0.83		
			O_oil		0.70		
			N_oil		0.87		
		263	S_oil		0.91		
Biomass	PA, EC, BC, OC	622	C_char	RF	0.95	2.40	Li et al. (2018)
Biomass	PA, EC, AR, OC	248	C_char	DNN	0.95	2.91	Li et al. (2021c)
			H/C_char		0.89	0.08	
			O/C_char		0.91	0.06	
			N/C_char		0.89	0.01	
Biomass	PA, EC, OC	296	ash_char	NN	0.97	NA	Mu et al. (2022)
				PSO-NN	0.98		
			N/C_char	NN	0.96		
				PSO-NN	0.96		
Sewage sludge	PA, EC, OC	138	N_char	ANN	0.97	-	Djandja et al. (2021)
Biomass	EC, OC	295	CO_gas	NN	0.95	0.31	Li et al. (2021a)
			CO_2_gas		0.97	1.06	
			CH_4_gas		0.94	0.45	
			H_2_gas		0.97	1.27	
	EC, OC, alkali catalyst	117	CO_gas		0.86	0.40	
			CO_2_gas		0.82	2.22	
			CH_4_gas		0.80	0.74	
			H_2_gas		0.91	2.08	
	EC, OC, transition-metal catalyst	73	CO_gas		0.87	0.51	
			CO_2_gas		0.98	0.96	
			CH_4_gas		0.92	1.58	
			H_2_gas		0.93	1.60	

Please cite this article as: Li H., Chen J., Zhang W., Zhan H., He C., Yang Z., Peng H., Leng L. Machine-learning-aided thermochemical treatment of biomass: a review. Biofuel Research Journal 37 (2023) 1786-1809. DOI: 10.18331/BRJ2023.10.1.4

Table S2.
Continued.

Biomass	Inputs [a]	Data number	Outputs [b]	Predictive performance, test data			Reference
				Algorithms [c]	R^2	RMSE [d]	
	EC, OC, with or without catalyst	233	CO_gas		0.80	0.26	
			CO_2_gas		0.91	2.38	
			CH_4_gas		0.86	1.22	
			H_2_gas		0.95	2.77	
Biomass	EC, OC	95	H_2_gas	GPR	0.95	2.93	Zhao et al. (2021)
				ANN	0.92	1.85	
				SVM	0.98	0.87	
				RF	0.98	1.02	
Food wastes	OC	40	CO_gas	NN	0.99	-	Shenbagaraj et al. (2021)
			CO_2_gas		0.98	-	
			CH_4_gas		0.99	-	
			H_2_gas		0.99	-	
Biomass	PA, EC, OC	125	H_2_gas	GPR	0.96	0.20	Haq et al. (2022)
				DT	0.92	1.23	
				SVM	0.69	2.97	
				ANN	0.86	1.99	
Dry thermochemical treatment							
Biomass	BC, OC	245	C_char	RF	0.76	6.94	Zhu et al. (2019)
	EC, OC				0.83	6.22	
	EC, BC, OC				0.84	5.81	
Biomass	PA, EC, BC, OC	226	FC_char	MLP-NN	0.90	5.1	Li et al. (2022)
			ash_char		0.94	2.0	
			VM_char		0.90	6.1	
			C_char		0.92	3.1	
			H_char		0.86	0.5	
			O_char		0.89	3.3	
			N_char		0.89	0.4	
			FC_char	ANFIS	0.79	7.4	
			ash_char		0.92	2.6	
			VM_char		0.81	7.7	
			C_char		0.85	4.1	
			H_char		0.84	0.6	
			O_char		0.86	3.6	
			N_char		0.87	0.5	
Biomass	EC, OC	236	N_char	RF	0.82	0.93	Leng et al. (2022a)
				GBR	0.92	0.64	
	PA, OC	170		RF	0.84	0.80	
				GBR	0.86	0.75	
	BC, OC	75		RF	0.88	0.44	
				GBR	0.61	0.80	

Please cite this article as: Li H., Chen J., Zhang W., Zhan H., He C., Yang Z., Peng H., Leng L. Machine-learning-aided thermochemical treatment of biomass: a review. Biofuel Research Journal 37 (2023) 1786-1809. DOI: 10.18331/BRJ2023.10.1.4

Table S2.
Continued.

Biomass	Inputs [a]	Data number	Outputs [b]	Predictive performance, test data			Reference
				Algorithms [c]	R^2	RMSE [d]	
	EC, EA, OC	161		RF	0.90	0.63	
				GBR	0.92	0.56	
	EC, EA, BC, OC	67		RF	0.77	0.64	
				GBR	0.65	0.79	
Biomass	PA, OC EC, OC BC, OC	171	O_oil	RF	0.80 0.90 0.82	NA	Yang et al. (2022)
Biomass	PA, OC EC, OC	264	H_Oil	RF	0.79 0.84	0.54 0.56	Tang et al. (2020)
Biomass	EC, AR, BC, OC EC, OC BC, OC PA, OC	214	H/C_oil	RF	0.75 0.74 0.78 0.68	0.63 1.12 0.67 1.24	Zhang et al. (2022)
Biomass	EC, AR, BC, OC EC, OC BC, OC PA, OC	232	O/C_oil	RF	0.78 0.92 0.89 0.87	0.14 0.17 0.13 0.18	Zhang et al. (2022)
One biomass	EC, OC	5237	CO_gas	RF	0.71	5.14	Mutlu and Yucel (2018)
				LS-SVM	0.90	1.52	
			CO_2_gas	RF	0.81	2.80	
				LS-SVM	0.93	0.64	
			CH_4_gas	RF	0.88	1.80	
				LS-SVM	0.94	0.34	
			H_2_gas	RF	0.81	5.04	
				LS-SVM	0.96	0.59	
Biomass	PA, EC, OC	120	CO_gas	RF	0.87	4.30	Tang et al. (2021)
				SVM	0.55	8.02	
			CO_2_gas	RF	0.87	5.71	
				SVM	0.76	7.51	
			CH_4_gas	RF	0.86	4.24	
				SVM	0.89	3.71	
			H_2_gas	RF	0.86	3.39	
				SVM	0.89	3.36	
Biomass	PA, EC, OC, gasifier	312	N_2_gas	ANN	0.98	0.10	Ascher et al. (2022)
			H_2_gas		0.95	0.11	
			CO_gas		0.93	0.11	
			CO_2_gas		0.90	0.11	
			CH_4_gas		0.80	0.18	
			C_2H_n_gas		0.52	0.19	

[a] Abbreviations see Table S1. [b] Contents of C, H, O, N, and S in oil/char; compositions of H_2, CO, CO_2, CH_4, C_2H_n, N_2 in gas; or yields of ash, volatile matter (VM), fixed carbon (FC) of char. [c] LS: least-squares. For other abbreviations for machine learning algorithms, see Table S1. [d] Root mean square error. [e] NA: not available or data available were calculated based on a unified dataset.

Please cite this article as: Li H., Chen J., Zhang W., Zhan H., He C., Yang Z., Peng H., Leng L. Machine-learning-aided thermochemical treatment of biomass: a review. Biofuel Research Journal 37 (2023) 1786-1809. DOI: 10.18331/BRJ2023.10.1.4

Table S3.
Machine learning aided prediction of the caloric value of thermochemical treatment products.

Biomass	Inputs [a]	Data number	Outputs [b]	Predictive performance, test data			Reference
				Algorithms [c]	R^2	RMSE [d]	
Wet thermochemical treatment (hydrothermal treatment)							
Biomass	EC, OC	292	HHV_oil	GPR	0.97	-[e]	Shafizadeh et al. (2022)
				NNR	0.80	-	
				GAM	0.92	-	
				SVR	0.79	-	
Biomass	BC, OC	325	HHV_oil	XGB	0.84	1.59	Katongtung et al. (2022)
	EC, AR, OC				0.84	1.60	
	EC, AR, BC, OC				0.86	1.47	
Biomass	PA, EC, OC	248	HHV_char	SVR	0.96	1.39	Li et al. (2020)
Biomass	PA, EC, BC, OC	475	HHV_char	RF	0.97	3.06	Li et al. (2018)
Biomass	PA, EC, AR, OC	248	HHV_char	DNN	0.95	1.53	Li et al. (2021c)
Biomass	PA, EC, OC	296	HHV_char	NN	0.90	-	Mu et al. (2022)
				PSO-NN	0.90	-	
Dry thermochemical treatment							
Biomass	EC, AR, BC, OC	198	HHV_oil	RF	0.68	1.68	Zhang et al. (2022)
	EC, OC	282			0.74	2.20	
	BC, OC	282			0.81	1.78	
	PA, OC	282			0.51	2.37	
Biomass	PA, EC, BC, OC	92	HHV_oil	RF	0.89	1.74	Leng et al. (2021)
	EC, BC, OC				0.84	2.09	
	PA, EC, OC				0.89	1.69	
	PA, BC, OC				0.85	2.03	
	PA, OC				0.90	1.60	
	BC, OC				0.87	1.88	
	EC, OC				0.93	1.38	
Biomass	PA, OC	115	HHV_char	ANN	0.997	0.03	Abdulsalam et al. (2020)
Biomass	PA, EC, OC	165	HHV_char	SVR	0.95	1.37	Li et al. (2020)
Biomass	PA, EC, AR, VM/FC, OC	149	HHV_char	GBR	0.93	1.74	Chen et al. (2022)
				RF	0.95	1.45	
				SVM	0.92	1.85	
Biomass	PA, EC, OC	312	LHV_gas	ANN	0.96	0.10	Ascher et al. (2022)
Biomass	EC, OC	5237	HHV_gas	RF	0.88	2.32	Mutlu and Yucel (2018)
				LS-SVM	0.96	0.38	

[a] VM/FC: ratio of volatile matter and fixed carbon. For other abbreviations, see **Table S1**.
[b] Higher heating value (HHV) or Lower heating value (LHV) of oil/char/gas phases.
[c] LS: least-squares. For other abbreviations for machine learning algorithms, see **Table S1**.
[d] Root mean square error.
[e] Not available or data available were calculated based on a unified dataset.

Please cite this article as: Li H., Chen J., Zhang W., Zhan H., He C., Yang Z., Peng H., Leng L. Machine-learning-aided thermochemical treatment of biomass: a review. Biofuel Research Journal 37 (2023) 1786-1809. DOI: 10.18331/BRJ2023.10.1.4

References

[1] Abdulsalam, J., Lawal, A.I., Setsepu, R.L., Onifade, M., Bada, S., 2020. Application of gene expression programming, artificial neural network and multilinear regression in predicting hydrochar physicochemical properties. Bioresour. Bioprocess. 7(1), 1-22.

[2] Alabdrabalnabi, A., 2021. Utilization of machine learning to predict bio-oil and biochar yields from copyrolysis of biomass with waste polymers. Fuel. 328, 125303.

[3] Ascher, S., Sloan, W., Watson, I., You, S., 2022. A comprehensive artificial neural network model for gasification process prediction. Appl. Energy. 320, 119289.

[4] Castro Garcia, A., Shuo, C., Cross, J.S., 2022. Machine learning based analysis of reaction phenomena in catalytic lignin depolymerization. Bioresour. Technol. 345, 126503.

[5] 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.

[6] 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.

[7] 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.

[8] Gu, C., Wang, X., Song, Q., Li, H., Qiao, Y., 2021. Prediction of gas-liquid-solid product distribution after solid waste pyrolysis process based on artificial neural network model. Int. J. Energy Res. 45(9), 13786-13800.

[9] 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.

[10] Katongtung, 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.

[11] Khan, M., Ullah, Z., Mašek, O., Naqvi, S.R., Khan, M.N.A., 2022. Artificial neural networks for the prediction of biochar yield: a comparative study of metaheuristic algorithms. Bioresour. Technol. 355, 127215.

[12] Leng, E., He, B., Chen, J., Liao, G., Ma, Y., Zhang, F., Liu, S., Jiaqiang, E., 2021. Prediction of three-phase product distribution and bio-oil heating value of biomass fast pyrolysis based on machine learning. Energy. 236, 121401.

[13] Leng, L., Yang, L., Lei, X., Zhang, W., Ai, Z., Yang, Z., Zhan, H., Yang, J., Yuan, X., Peng, H., Li, H., 2022a. Machine learning predicting and engineering the yield, N content, and specific surface area of biochar derived from pyrolysis of biomass. Biochar. 4(1), 63.

[14] Leng, L., Zhang, W., Chen, Q., Zhou, J., Peng, H., Zhan, H., Li, H., 2022b. Machine learning prediction of nitrogen heterocycles in bio-oil produced from hydrothermal liquefaction of biomass. Bioresour. Technol. 362, 127791.

[15] 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.

[16] Li, J., Pan, L., Suvarna, M., Wang, X., 2021a. Machine learning aided supercritical water gasification for H_2-rich syngas production with process optimization and catalyst screening. Chem. Eng. J. 426, 131285.

[17] Li, J., Zhang, W., Liu, T., Yang, L., Li, H., Peng, H., Jiang, S., Wang, X., Leng, L., 2021b. 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.

[18] Li, J., Zhu, X., Li, Y., Tong, Y.W., Ok, Y.S., Wang, X., 2021c. 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.

[19] 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.

[20] Li, Y., Gupta, R., You, S., 2022. Machine learning assisted prediction of biochar yield and composition via pyrolysis of biomass. Bioresour. Technol. 359, 127511.

[21] 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.

[22] Mutlu, A.Y., Yucel, O., 2018. An artificial intelligence based approach to predicting syngas composition for downdraft biomass gasification. Energy. 165, 895-901.

[23] Onsree, T., Tippayawong, N., 2021. Machine learning application to predict yields of solid products from biomass torrefaction. Renewable Energy. 167, 425-432.

[24] Pathy, A., Meher, S., Balasubramanian, P., 2020. Predicting algal biochar yield using eXtreme Gradient Boosting (XGB) algorithm of machine learning methods. Algal Res. 50, 102006.

[25] 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.

[26] 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(24), 12739-12757.

[27] Tang, Q., Chen, Y., Yang, H., Liu, M., Xiao, H., Wang, S., Chen, H., Naqvi, S.R., 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.

[28] Tang, Q., Chen, Y., Yang, H., Liu, M., Xiao, H., Wu, Z., Chen, H., Naqvi, S.R., 2020. Prediction of bio-oil yield and hydrogen contents based on machine learning method: effect of biomass compositions and pyrolysis conditions. Energy Fuels. 34(9), 11050-11060.

[29] Ullah, Z., khan, M., Naqvi, S.R., 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.

[30] Yang, K., Wu, K., Zhang, H., 2022. Machine learning prediction of the yield and oxygen content of bio-oil via biomass characteristics and pyrolysis conditions. Energy. 254, 124320.

[31] Zhang, T., Cao, D., Feng, X., Zhu, J., Lu, X., Mu, L., Qian, H., 2022. Machine learning prediction of bio-oil characteristics quantitatively relating to biomass compositions and pyrolysis conditions. Fuel. 312, 122812.

[32] 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.

[33] 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.

[34] Zhu, X., Li, Y., Wang, X., 2019. Machine learning prediction of biochar yield and carbon contents in biochar based on biomass characteristics and pyrolysis conditions. Bioresour. Technol. 288, 121527.

Please cite this article as: Li H., Chen J., Zhang W., Zhan H., He C., Yang Z., Peng H., Leng L. Machine-learning-aided thermochemical treatment of biomass: a review. Biofuel Research Journal 37 (2023) 1786-1809. DOI: 10.18331/BRJ2023.10.1.4