Korean Journal of Chemical Engineering (2024) 41:1923–1953
https://doi.org/10.1007/s11814-024-00181-7

REVIEW ARTICLE

Progresses and Challenges of Machine Learning Approaches in Thermochemical Processes for Bioenergy: A Review

Nafiu Olanrewaju Ogunsola[1] · Seung Seok Oh[2] · Pil Rip Jeon[3] · Jester Lih Jie Ling[2] · Hyun Jun Park[2] · Han Saem Park[2] · Ha Eun Lee[2] · Jung Min Sohn[1] · See Hoon Lee[1,2]

Received: 29 January 2024 / Revised: 23 April 2024 / Accepted: 1 May 2024 / Published online: 10 May 2024

Abstract

Thermochemical conversions of nonedible biomass into energy are promising alternatives for ensuring a sustainable energy society. However, determining the optimum design and operating conditions of the processes remains a major challenge due to the laborious and costly experimental methods. Machine learning techniques are cost-effective and non-time consuming and have been widely utilized in thermochemical conversion process modelling with robust and accurate results and solutions. Nonetheless, no standard method has been proposed for applying ML models to biomass thermochemical processes. Consequently, the general development procedure for ML models with high accuracy and robustness remains unclear. This review provides a comprehensive review of machine learning techniques for predicting biofuel yield and composition. It is recommended that quality datasets be ensured to enable the development of more robust machine learning-aided models for practical engineering applications. Finally, solutions to the identified challenges and prospective future research directions on machine learning-based biomass thermochemical conversion processes are recommended to accelerate the optimization and large-scale deployment of these processes.

Keywords Bioenergy · Thermochemical conversion · Machine learning · Artificial neural networks · Sustainable biomass utilization

Abbreviations

ABC	Artificial bee colony
AC	Ash content
BC	Biomass content in blending
C-char	Carbon contents of biochar
CART	Classification and regression tree
Ce	Cellulose content
CFNN	Cascade forward neural network
DSS	Decision support system
Ea	Activation energy
Err	Relative error
ERT	Extremely randomized trees
F	Flow rate
FC	Fixed carbon
GAM	Generalized additive model
GBR	Gradient tree boosting
GPR	Gaussian process regression
H/C	H-to-C atomic ratio
HHV	Higher heating value
Hm	Hemicellulose content
HR	Optimum heating rate
HT	Holding time
HTT	Highest treatment temperature
K_0	Pre-exponential factor
Lg	Lignin content
MC	Moisture content
MCS	Monte Carlo simulation
MLnR	Multi-nonlinear regression
MLR	Multilinear regression
N/C	N-to-C atomic ratio

Co-first authors: Nafiu Olanrewaju Ogunsola, Seung Seok Oh and Pil Rip Jeon.

✉ See Hoon Lee
donald@jbnu.ac.kr

[1] Department of Mineral Resources and Energy Engineering, Jeonbuk National University, 567 Baekje-daero, Deokjin-gu, Jeonju-si, Jeollabuk-do, Republic of Korea

[2] Department of Environment and Energy, Jeonbuk National University, 567 Baekje-daero, Deokjin-gu, Jeonju-si, Jeollabuk-do, Republic of Korea

[3] Department of Chemical Engineering, Kongju National University, Cheonan-daero 1223-24, Seobuk-gu, Cheonan-si, Chungcheongnam-do 31080, Republic of Korea

n	Reaction order
O/C	O-to-C atomic ratio
P	Production capacity
PC	Pyrolysis conditions
PS	Particle size
RBF	Radial basis function
RMSE	Root mean square error
RT	Residence time
SGB	Stochastic gradient boosting regression trees
SHAP	Shapley additive explanation
SML	Sample mass loading
SS	Sample size
T	Operating temperature
TDSS	Tunable decision support system
TRS	Tunable recommendation system
V	Space velocity
VM	Volatile matter

Introduction

Currently, the modern world faces a series of worldwide energy security issues, starting with a growing population and an accompanying rise in energy consumption that results in environmental pollution and climate change [1, 2]. Among them, the 'shortage of energy' is a foremost issue because it is closely linked to the serious pollution induced by excessive use of conventional fossil fuels. Fossil fuel sources such as coal, natural gas, and petroleum are continuously being utilized to fulfill a substantial portion of the global energy demands [2, 3]. The escalating use of fossil fuels has resulted in a persistent rise in the levels of greenhouse gases (GHGs) including carbon dioxide (CO_2) and methane (CH_4), and atmospheric pollutants like sulfur oxides (SOx), nitrogen oxides (NOx) and so on. This has consequently given rise to significant environmental challenges [4, 5]. Climate change arises from global warming due to the extensive dependence on fossil fuels. This dependency contributes to elevated surface temperatures, sea level rise because of melting glaciers, and various other environmental concerns [6, 7]. The increasing awareness of environmental damage and GHG emissions, in conjunction with fluctuations in the price of petroleum products and gasoline, has opened opportunities to explore alternative and renewable energy sources [8]. Therefore, it is imperative to develop suitable, sustainable, and renewable alternatives to fossil fuels to mitigate GHG emissions and adverse environmental consequences that come with them.

In contrast to fossil fuels, biomass energy conversion processes are sustainable and renewable. Solid biomass, such as lignocellulosic and municipal waste, has been receiving growing interest as a reliable alternative to fossil fuels for the past several decades. Biomass energy conversion methods, including biochemical and thermochemical processes, have been recently growing owing to their carbon neutrality and eco-friendly advantages [9]. Biomass consists of around 38–50% of cellulose, 23–32% of hemicellulose, 15–25% of lignin, and 5–13% of other elements, including inorganic substances and extractives like sodium, potassium, and calcium [10–12]. These components can be converted into heat, electricity, fuels, and other value-added products through either biochemical or thermo-chemical processes.

Biochemical conversion processes such as anaerobic and syngas fermentation utilize bacteria, enzymes, and microorganisms to decompose biomass into fuels and other products such as biogas, biodiesel, bioethanol, and other specific constituents [13, 14]. However, the main drawback of these biochemical processes is their low reaction rate, which requires hours, days, or even weeks to complete the necessary reactions [7]. Thermo-chemical conversion of biomass is more commonly used than biochemical processes due to its several advantages and higher conversion efficiency. Unlike biochemical processes, thermochemical conversion does not rely on microorganisms. Instead, they utilize thermal energy and chemical reactions to decompose the biomass into various products [15–17]. They are relatively faster than biochemical processes, with reactions taking only a few minutes or seconds. Additionally, these processes can generate high yields and are adaptable to different types of feedstocks [18].

Thermochemical conversion processes are categorized into four classes: pyrolysis or torrefaction, gasification, hydrothermal treatment, and combustion or incineration [2, 15, 17]. Pyrolysis is a crucial technology used for transforming various types of biomasses, including lignocellulosic and waste biomass, into useful fuel or materials such as bio-oil, biochar, and gases such as H_2, CH_4, and so on. Pyrolysis is commonly classified into slow, fast, and flash depending on heating rate and solid residence time in pyrolizers [19, 20]. The efficiency and characteristics of pyrolysis products are significantly influenced by operation factors including reaction temperatures, solid residence time, heating rate, and water content of biomass [21]. Gasification is the process of transforming solid biomass into syngas. This can be classified into conventional and hydrothermal types based on factors such as reaction conditions, gasifying agents, and the syngas quality. [2, 15, 22]. Compared to other technologies that use dry biomass, hydrothermal treatment (HTT) technology transforms wet biomass into end products such as gaseous and solid fuels that are entirely different from those produced by dry conversion processes by utilizing hot pressurized water as a reactant, catalyst, and solvent [7]. HTT can be categorized as either hydrothermal carbonation or hydrothermal liquefaction, depending on factors such as temperature, pressure, and proportion of the intermediate products [23]. Combustion converts solid biomass into heat

and electricity for industrial and household uses [17, 24]. The combustion process can be classified as single combustion or co-combustion, depending on the fuel type. Figure 1 shows the various types of thermochemical conversion processes. However, the modeling of these products, technologies, and systems presents various technical bottlenecks, most of which are due to their multiparametric and complex characteristics.

The optimization of typical thermochemical conversion processes at diverse research levels often requires numerous rigorous experiments to identify the operating conditions and the optimal design. To enhance the efficiency of optimization and lower operational costs, innovative modeling approaches like computational fluid dynamics (CFD), kinetic and thermodynamic models have been employed [25–27]. However, implementing these types of modeling using conventional techniques can be difficult and sometimes unrealistic owing to their assumptions, complexities, and limitations [28]. In addition, these models are difficult to incorporate with different variations in thermochemical conversion process yields and kinetics under operating conditions [27, 29]. For instance, CFD modeling is time-consuming and computationally costly due to its significant amount of calculation and predefined parameters, whereas the kinetic model requires complex reaction rate estimations since the reaction mechanism is either unidentified or not completely comprehended. Thermodynamic modeling frequently assumes the equilibrium state, despite the fact that the majority of experimental reactions rarely reach equilibrium [2]. Therefore, it is essential to devise a precise, time-efficient, economical, and resilient modeling method to establish correlations for the intricacies of thermochemical conversion processes.

Machine learning (ML) techniques, in response, have been increasingly utilized as a reliable alternative to conventional modeling techniques for comprehending complex biomass thermochemical conversion processes. Unlike conventional models, ML models can effectively describe intricate relationships between input and output variables, offering a more accurate and dependable approach to modeling complex problems. Several studies have employed ML in various aspects of biomass thermochemical conversion, including bioenergy and biofuel conversion processes, emissions prediction of coal-fired boilers, biomass pyrolysis in fluidized bed reactors, and steam methane reforming (SMR) [30–35]. Most pioneering review studies on thermochemical conversion processes have focused on the application of conventional modeling techniques to product yields and kinetics [36–38]. However, recent studies have demonstrated the potential of ML models for thermochemical conversion processes, particularly pyrolysis and gasification. Nevertheless, there is a shortage of comprehensive reviews on the application of ML models in HTT and combustion processes [2, 31].

In a recent review of the application of ML in thermochemical conversion processes, Ascher et al. [31] highlighted the recent advances in pyrolysis and gasification product yields, distributions, and kinetics optimization and prediction. However, other thermochemical conversion processes, including HTT and combustion, were not considered. Umenweke et al. [2] investigated and summarized biomass gasification and its recent advances in the application of ML

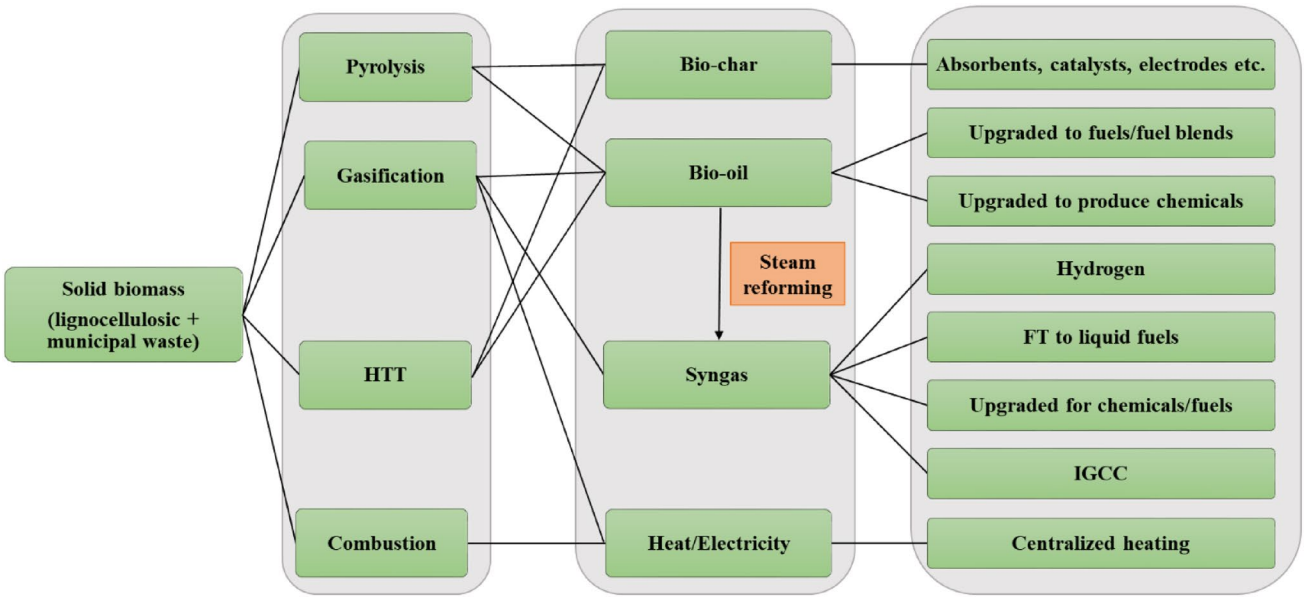

Fig. 1 Classification of thermochemical conversion processes utilizing biomass [7]

to product yields and distribution during conventional and hydro-thermal gasification. However, this review also did not consider other thermochemical conversion processes, such as pyrolysis, HTT, and combustion. Recently, Jeon et al. [39] reviewed the detailed procedures of ML development for specific biomass conversion processes, including pyrolysis, gasification, and hydrothermal treatment. They compared the relative importance of input variables in predicting output variables and discussed the application of ML in techno-economic analysis. However, their focus primarily centered on the detailed ML development procedure to enhance model performance, overlooking crucial insights into improving the efficiency of biomass conversion processes, such as variations in product yields under different operation conditions.

Information gathering through a comprehensive review of all thermochemical conversion processes could help determine the most important parameters and optimal operating conditions that affect product yields and kinetics, which can aid in the optimization and large-scale deployment of the processes. Therefore, this study aims to comprehensively cover all areas in which ML has been employed in this field, highlight the challenges encountered in existing ML techniques, and suggest potential solutions and future research directions. The application of ML in biomass thermochemical conversion processes, encompassing the combustion process, is reviewed to determine optimal operational conditions for a given biomass. Given that the successful commercialization of biomass thermochemical conversion relies on reactor and process designs [40], this review also investigates the types of reactors employed in the literature. Ultimately, this study elicits recent advances in the application of ML to solid biomass thermochemical conversion processes and proposes strategies to enhance the efficiency of biomass thermochemical conversion process under specific conditions.

General Machine Learning Approaches

Machine learning (ML) methods have extensively been applied in addressing societal challenges across diverse fields, including biomass thermochemical conversion processes. These approaches offer the benefit of being efficient and yielding accurate results that closely align with experimental findings. ML algorithms can be categorized as supervised, unsupervised, and semi-supervised.

Supervised ML involves feeding labelled input and corresponding output data to the model, which then learns to map input data to its targeted output [41]. This type of learning can be divided into regression and classification techniques. Both approaches employ similarly labelled datasets for predictions distinctively. While regression methods aim to find the best match between dependent and independent variables in continuous data, classification methods aim to find that in discrete data or identify behavioral patterns across datasets [42]. Unsupervised ML, also known as self-organization, differs from supervised learning in that it works well with unlabeled data and excels at identifying hidden patterns in the data. When input data is fed into an unsupervised ML model, the model searches and sorts the similarities and differences among the data. Unsupervised ML methods encompass principal component analysis (PCA), dimensionality reduction, and clustering.

In contrast, semi-supervised ML integrates the benefits of both supervised and unsupervised learning. It can handle systems with both labeled and unlabeled input and helps address challenges in supervised learning [41]. Labeling data for supervised learning can be difficult and time-consuming, but semi-supervised learning can overcome this limitation by integrating a portion of the unlabeled input data into the supervised learning process [41, 43]. An overview of the classes of ML techniques applicable in the fields of science and engineering, including biomass applications is illustrated in Fig. 2. A wide variety of ML models have been employed for various aspects of biomass thermochemical conversion processes, with the artificial neural network (ANN) being mostly explored model in this research area. A brief discussion of the ANN and other models that have been applied in diverse aspects of this field are discussed below.

Artificial Neural Network

Artificial neural network (ANN) is an ML algorithm that imitates the human brain's ability to extract, process, and interpret information. It follows a mathematical model based on the functioning of biological neurons in solving complex problems [44]. It offers the benefit of straightforward implementation and does not necessitate a predetermined or recognized connection between the inputs (features) and outputs (labels). The ANN is comprised of interconnected nodes, and its capacity to process information is encoded in the weights that connect these nodes. The arrangements of connections among the nodes in the ANN are referred to as the ANN network architecture. Figure 3 illustrates a standard depiction of the ANN network architecture.

The ANN network architecture comprises three essential layers: an input layer containing input features, a hidden layer comprising hidden neurons, and an output layer that provides the predicted target. Also, ANN can be divided into two main network architectures: feed-forward and feedback or recurrent network [45]. The feed-forward networks lack loops, as seen in a multilayer perceptron where input and output neurons are layered with one-way direction between them. In general, the feed-forward networks generate a set of outputs rather than a sequence in response to a given input.

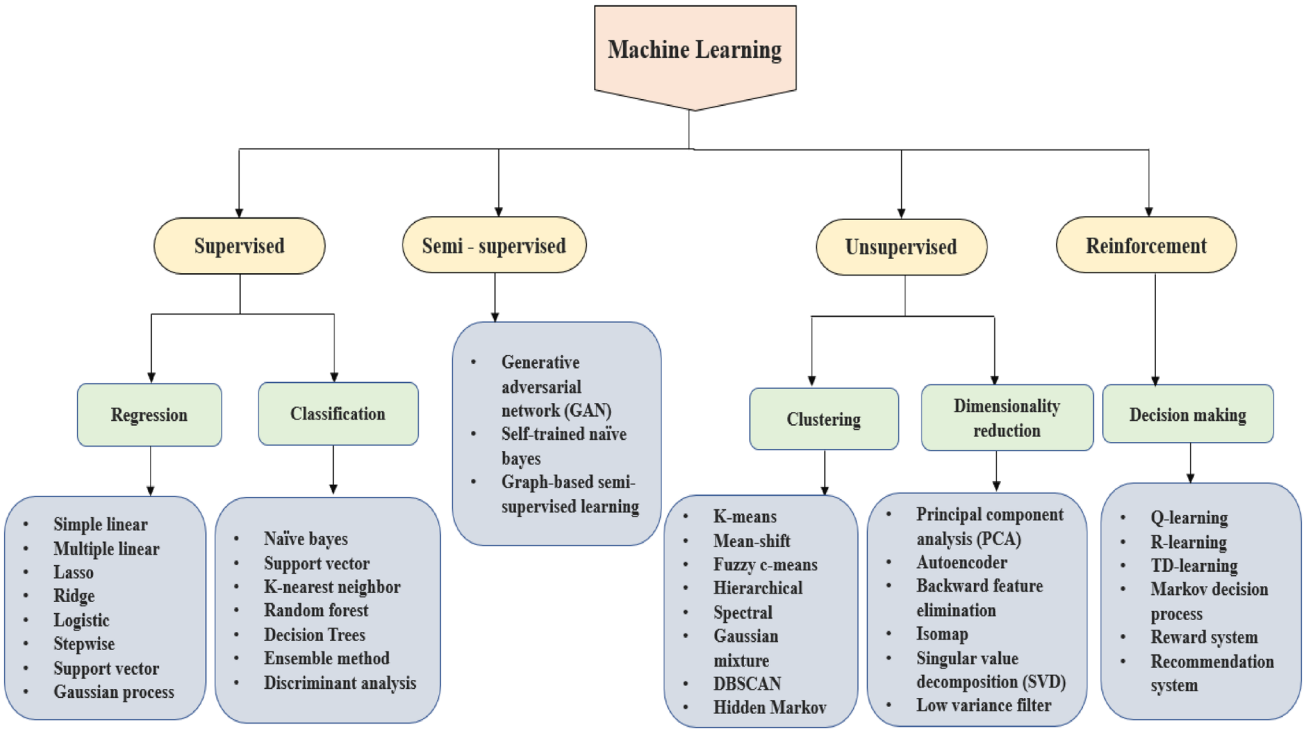

Fig. 2 Classification of general ML algorithms

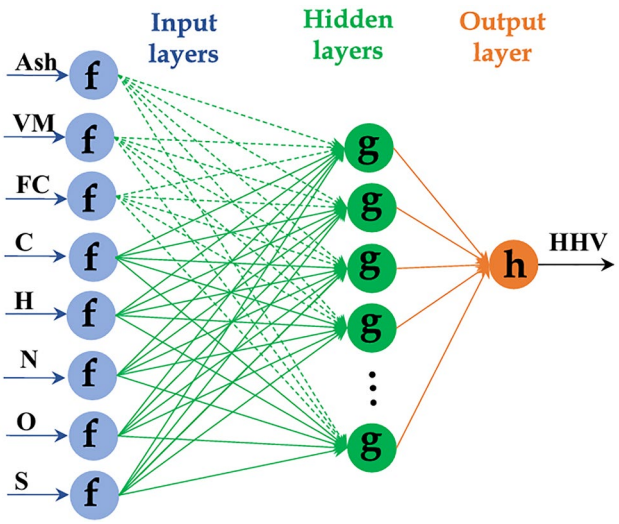

Fig. 3 A typical network architecture of ANN [47]

In contrast, feedback or recurrent networks have variability [46]. Regardless of the mentioned network architectures, the crucial aspect is the learning process.

Decision Tree and Random Forest

The decision tree (DT) model possesses a distinctive ability to address both linear and nonlinear problems, rendering it well-suited for intricate non-linear models. The tree models are more pertinent than other black box ML models due to their straightforward comprehension and interpretability [48, 49]. The primary drawback linked to DT models is the risk of overfitting with the increased number of datasets as the model tends to develop complex tree structures [50]. Another constraint of conventional DT models is their incapacity to account for noise in the dataset and the absence of smoothness in their corresponding functions [31]. The DT models, however, have benefits in terms of simplicity and ease of interpretation coupled with the low computational cost [51]. Therefore, other models such as random forest (RF) have been concurrently employed to improve the veracity of each DT models.

RF is a collection of DT designed to address the overfitting limitation. To mitigate overfitting, RF employs an averaging approach across smaller decision trees rather than relying on a single massive DT [52, 53]. RF stands out as a user-friendly ML model suitable for both supervised and unsupervised learning techniques because it has fewer hyper-parameters. Even with sub-optimal hyperparameters, it exhibits commendable performance and accuracy. The two most commonly used parameters among the various options in RF are the total number of individual tree parameters (Nvar) and decision trees (n). The performance of model is significantly influenced by the number of decision trees running in parallel, depending on the desired output [54].

Figure 4 illustrated a standard graphical representation of both DT and RF techniques.

Support Vector Machine

Support vector machine (SVM) employs a kernel approach, integrating both classifications, known as support vector classifier (SVC), and regression, known as support vector regression (SVR). This is applicable to linear, non-linear, and multi-dimensional problems, guided by the Vapnik–Chervonenkis (VC) theory [56, 57]. A kernel method is an analysis technique merely based on the dot-products of available datasets. The SVM encompasses the combination of four fundamental concepts: the hyperplane of separation, the hyperplane with maximum margin, flexible margin, and kernel function. The operation of the SVMs depends on identifying the hyperplane which minimizes the distance between examples [58]. A typical structure of the SVM is depicted in Fig. 5.

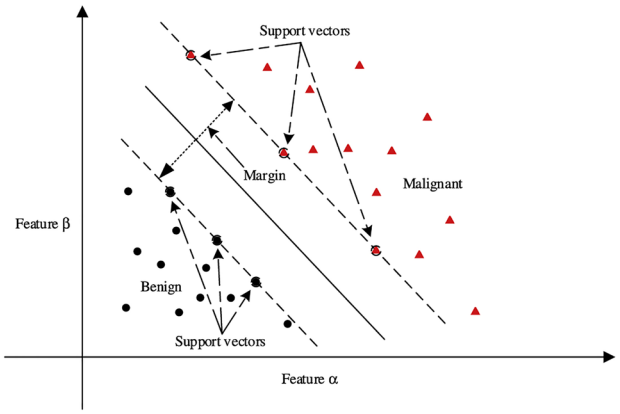

Fig. 5 A typical structure of support vector machine in two-dimension (modified from [59])

Deep Learning

Deep learning (DL) is characterized as the distinct type of neural network architecture featuring multiple layers. A graphical representation of the distinction between shallow learning and deep learning is shown in Fig. 6. Different from shallow learning that only has one hidden layer, DL is characterized by multiple hidden layers and activation functions. DL employs a linear regression framework built upon numerous neural nodes or networks. The capacity of DL models to formulate intricate hypotheses renders them potent for comprehending complex, nonlinear, and multi-dimensional relationships. Therefore, the DL can solve problems that shallow learning models could hardly solve [60]. Examples of DL algorithms include recurrent neural network (RNN), long short-term memory (LSTM), convolutional neural network (CNN) and deep neural network (DNN). RNN is a kind of network combinations in a loop through which information persistence occurs [61]. The LSTM are a form of RNN that is explicitly designed to avoid the problem of RNN networks' long-term dependency. The standard LSTM module has four neural network layers known as the repeated module interacting. The DNN is one of feed-forward neural networks containing multiple layers of hidden units situated between its input and output layers.

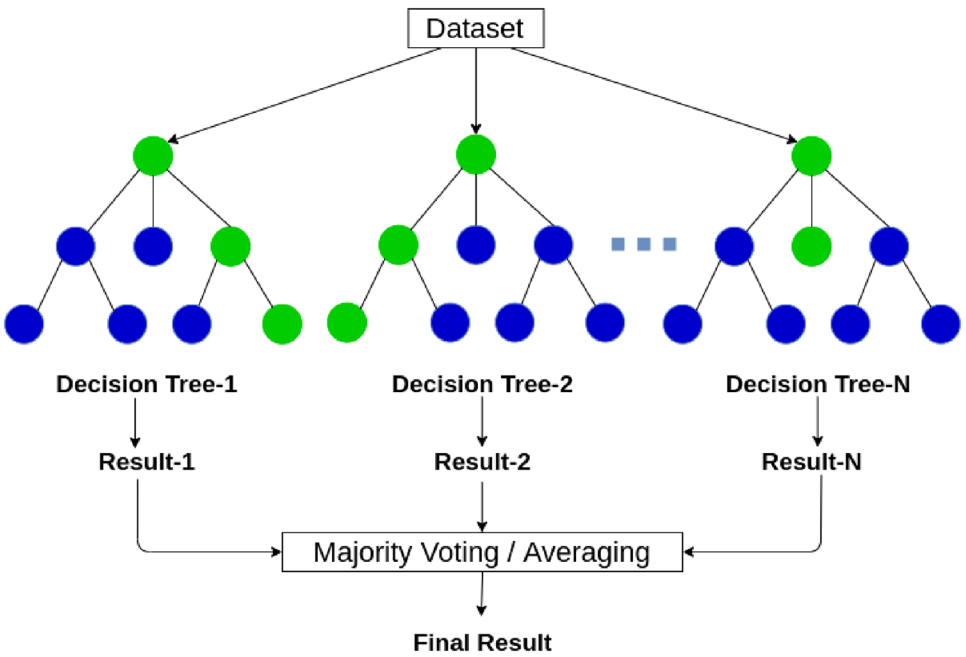

Fig. 4 A schematic diagram of decision tree and random forest pathway (modified from [55])

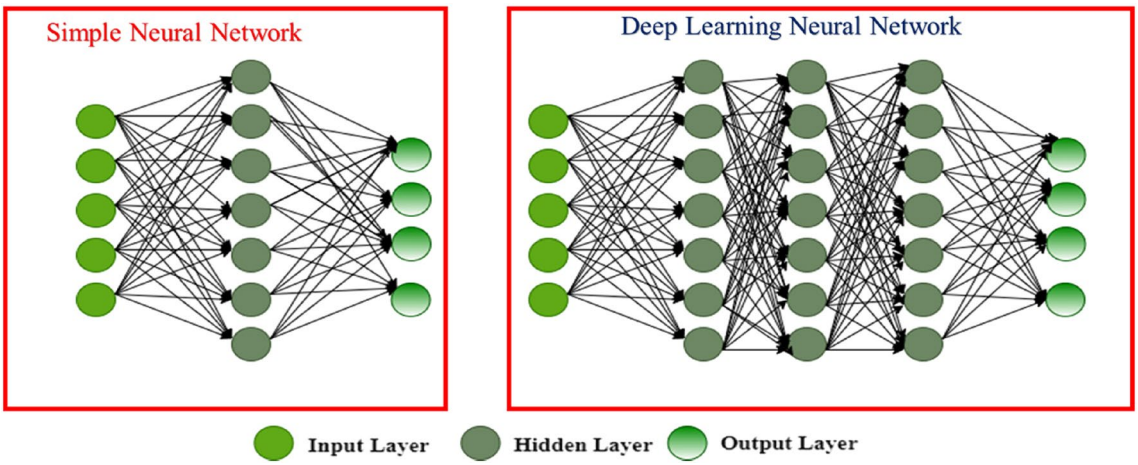

Fig. 6 Distinction between simple and deep neural network architecture (modified from [2])

Machine Learning Applications in Biomass Thermochemical Conversion

The implementation of ML techniques has been expanded in the thermo-chemical conversion processes. Such implementation allows prediction of yield and kinetics of pyrolysis process, syngas or producer gas yield prediction for a gasification process, process modeling, and quantifying char and tar formation among others. The following section and sub-sections are aimed at summarizing existing literature in this field and highlighting the application of ML techniques for operation optimization and research advancement.

Machine Learning Applications in Pyrolysis Processes

Pyrolysis is usually implemented in an oxygen-free condition at elevated temperatures of about 300–700 ℃ [62]. ML techniques have been applied to biomass pyrolysis thermochemical conversion processes including the prediction of pyrolysis yield and kinetics. Some of the commonly employed input variables include particle characteristics and operating conditions. The particle scale parameters comprise the particle size and chemical compositions. The operating conditions comprise temperature, gas velocity, heating rate, production capability, and sample weight [63]. The application of ML techniques in biomass pyrolysis prediction is discussed in the sub-sections that follow.

Yield Prediction of Biomass Pyrolysis

The prediction of biomass pyrolysis yields has been previously reported in the literature using the ANN and some other ML algorithms, in terms of biochar, bio-oil, and biogas and their constituents inclusive of fixed carbon, volatile matter, ash, and water content. The quality of end products of pyrolysis has been known to depend strongly on various operation conditions and particle characteristics [21]. These variables are the features usually used to develop the pyrolysis yield prediction ML models.

Earlier studies employed the ANN model to estimate the quality and quantity of liquid and gaseous pyrolysis products generated from three biomasses waste [64]. Catalyst type, amount, temperature, and biomass diversity were input variables for the pyrolysis process while the output of the model is the hydrogen-rich gas (H-rG) ratio. A total of 168 datasets consisting of 102 training, 33 testing, and 33 validation datasets were used to progress the ANN model. The forecasting performance and accuracy of the ANN model were superior with the regression coefficient (R) of 0.975, 0.955, and 0.902 and mean square error (MSE) of 3.25, 6.97, and 9.20 for training, testing, and validation datasets, respectively. The developed model can be applied to similar experimental programs provided the range of model parameters are within the range used in this model.

In a recent study [65], ANN network architecture was used to calculate the basic yields of liquid, gas, and solid product from a pyrolizer. Eighteen datasets with eight input parameters consisting of particle scale parameters and operating conditions, particularly temperature for four biomasses waste, were used to develop the ANN model. The model developed showed high predictive accuracy with an R value of 0.9999 and 0.9941 and MSE of 0.0176 and 5.1714 for both the training and testing dataset, respectively. The major yields or products of pyrolysis (char, gas, and tar) using seven different ML algorithms including Linear regression (LR), ANN, K-nearest neighbor (KNN), SVR, DT, RF, and DNN were predicted in a spouted bed [66]. The labeled dataset for the training of the ML models was generated by the computational particle fluid dynamics (CPFD) simulation.

The input parameters used for the simulation include temperature and gas residence time. The developed ML models showed a better agreement with the product yields of the CPFD than those of the lumped process models (Fig. 7). Finally, the results of the ML applications were evaluated with those of the CPFD, and process analysis (PA) lumped

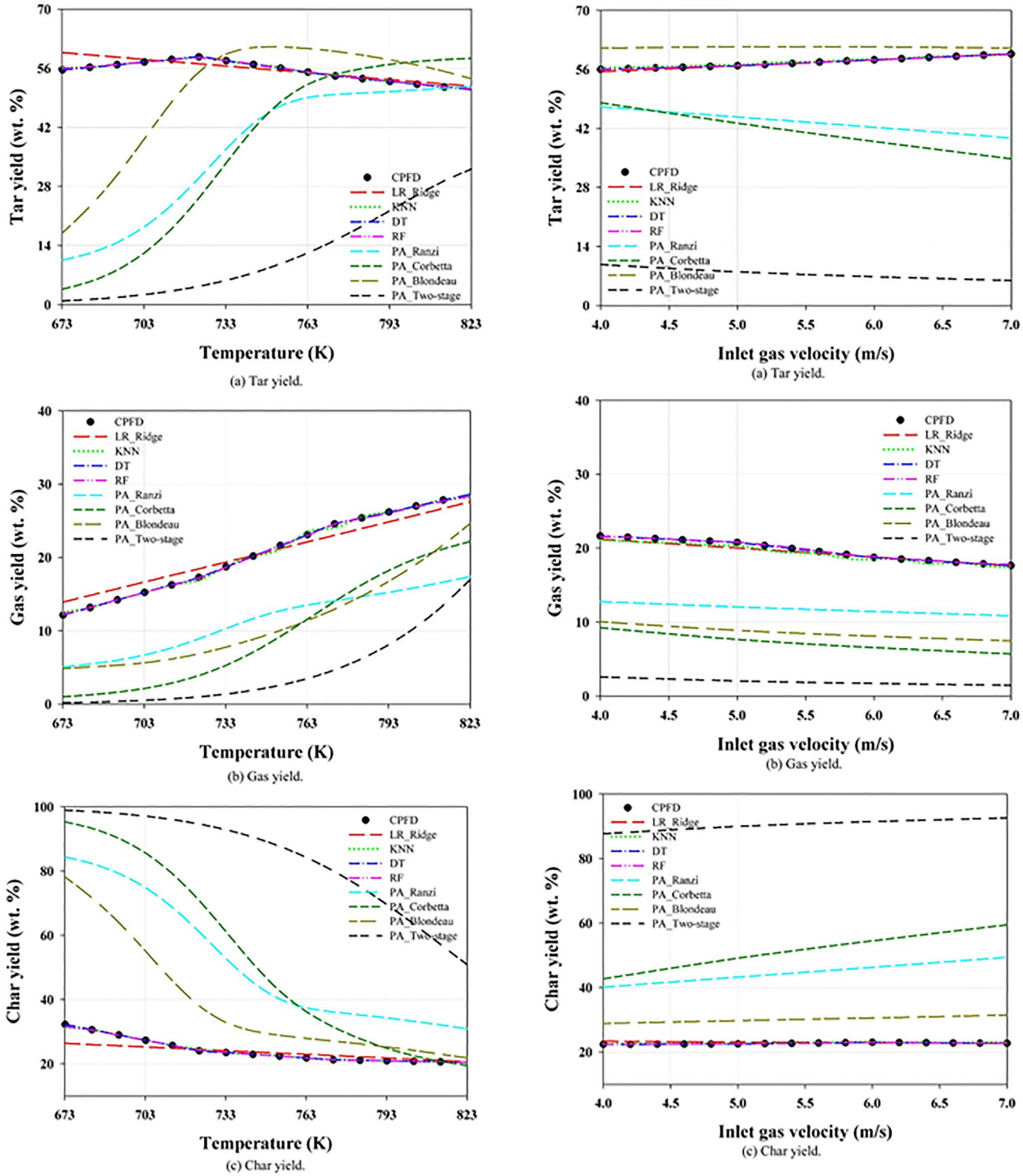

Fig. 7 Comparisons of the yields of fast pyrolysis products for various prediction models and CPFD with respect to the reaction temperature and inlet gas velocity (modified from [66])

models for all products. The comparison study indicates that the developed ML models produce highly accurate predictions compared to lumped process models. A summary of other recently developed ML models for pyrolysis yield prediction is highlighted in Table 1.

The review of constructed models on pyrolysis product yields showed that the biomass pyrolysis condition is the most sensitive feature for pyrolysis yield prediction. Tang et al. [77] found that the pyrolysis conditions predominantly have a larger effect on the process outputs than the biomass composition. However, in the other two studies by Cheng et al. [80] and Zhu et al. [72], pyrolysis temperatures were observed to have a powerful effect on the production of biochar from a range of biomass feedstocks.

Machine Learning Applications in Biomass Pyrolysis Kinetic Parameters Prediction

The ANN has been widely applied to predict the kinetic parameters or constituent distributions of a biomass pyrolysis process. Zhong et al. [81] developed an ANN model to build CFD-based reduced order models (ROMs) using pyrolysis temperature and coordinates in the reactor as input variables to predict the distributions or kinetics of gas, liquid, and solid fractions in a bubbling fluidized bed pyrolyzer. The developed ANNs have high predictive accuracy with an average R^2 of 0.9940 and were able to show good results for tar, gas, and biomass distributions and kinetics simultaneously.

Sasithorn et al. [82] employed a total of 150 datasets from the thermogravimetric experiments of various biomass resources to develop an ANN model to correlate biomass constituents with the kinetic parameters of the pyrolysis process, in terms of pre-exponential constant (k_0), activation energy (Ea), and reaction order (n). Although non-linear relationships were depicted between the biomass components and the output variables, the developed ANN model shows a predictive accuracy, R^2 of more than 0.9. The results obtained from the developed ANN models showed minor deviation compared to the experimental results from thermogravimetric analyses (TGA).

In another study, the RF method was successfully applied to develop a prediction model for the pyrolysis activation energy of 5 different biomass resources [83]. For this study, 281 datasets comprising 10 features (types, ultimate analysis results (C, H, O, N, S), H/C ratio, O/C ratio, ash, and model-free method) were used to develop the model. The developed model showed a high predictive performance with the coefficient of determination reaching a value as high as 0.9964. Based on the findings, the reliability of the RF model for the estimation of pyrolysis kinetic has been verified. A summary of other recently developed ML models for biomass pyrolysis kinetic parameters prediction is highlighted in Table 2.

Machine Learning Applications in Gasification Processes

Gasification is an important thermo-chemical conversion process that has been vastly researched to produce hydrogen-rich syngas or producer gas from biomass. The gasification process usually involves four distinct processes including drying, pyrolysis, partial combustion and tar cracking, and reduction. The application of novel ML techniques in gasification studies has been tailored towards the prediction of syngas yield and solid residue (char and tar) yield, and catalysis screening and selection. The application of ML techniques in the gasification process has helped in designing a cost-effective process with higher carbon conversion efficiency, as well as reducing the processing time and expenses of complex and time-consuming practices [84]. The various applications of ML techniques in the prediction of biomass gasification processes are elaborated in the subsequent "Machine learning applications in conventional gasification process" and "Machine learning applications in hydrothermal gasification process".

Machine Learning Applications in Conventional Gasification Process

Conventional gasification has a lot of merits, such as low capital cost, low operating cost and less risk of explosion with decreased operating pressure. However, the process requires drying of waste biomass with moisture content above 35wt% which is different from the hydrothermal gasification. Pioneer studies with the implementation of ML algorithms in the conventional gasification process have demonstrated some promising results. Two separate ANN network architectures were employed to predict the syngas yields including CO, H_2, CO_2, and CH_4, during the gasification of biomass in fluidized bed gasifiers [93]. The datasets used for the ANN model development were obtained from 18 and 36 experimental runs. Biomass composition and equivalent ratios were used as input conditions. The developed ANN models displayed robust and accurate predictive performance with R^2 of > 0.97. However, the model is constrained to only one type of feedstock and the specific range of the experimental conditions used in model development. The sensitivity analysis was also carried out to figure out the relation between inputs and predicted outputs and showed that two inputs have greater effects on the syngas yields [93]. The results and limitations depicted in the previous study [93] coincided with the results obtained by Mikulandrić et al. [94]. They used gasification temperature, gas flow, and residence time to estimate the yield and compositions of syngas in a fixed-bed gasifier.

In their attempt to solve the lack of model generalization in the pioneer studies, Baruah et al. [95] developed an ANN

Table 1 Summary of pioneering investigation on machine learning application for biomass pyrolysis

Biomass	Reactor type	Method	Model inputs	Model outputs	Model performance	Major findings	References
Pine sawdust	Quartz tubular (L: 1 m; D: 50 mm)	ANN	T, PS, V	CO, CO_2, CH_4, H_2	$R^2 > 0.99$	According to the ANN model, the primary gas products were predominantly influenced by temperature. The total selectivity of produced gas increased from 2.91% (300 ℃) to 34.31% (900 ℃)	[67]
Algal	Diverse	XGBoost, SHAP	C, H, O, N, H/C, O/C, N/C, AC, FC, VM, T, HR, and RT	Biochar yield	R^2: 0.84	Pyrolysis temperature, carbon, and ash content showed greater effects on the algal biochar yield	[68]
lingo-cellulosic biomass, herbaceous plants, food waste, and algae	Fixed bed	RF, MLR, SVM, DT	C, H, O, N, AC, FC, VM PC	Bio-oil yield	R^2: 0.98	The RF model showed better prediction performance, and partial dependence plot analysis showed that the optimal features selected using a genetic algorithm-based approach have a great influence on biooil yield	[69]
Ligno-cellulose	Batch-type setups	ANN	H, RT, PS, SS, T, MC, AC, Lg, Hm, Ce, HT	Biochar, tar, and biogas yields	R^2: 0.66, 0.70, 0.84	The ANN models could consistently show reliable relations between yields and temperatures	[70]
Ligno-cellulose	Batch and continuous setups	ANN, MLnR	T, RT, PS, AC, FC, P	Biochar, FC, VM, ash, HHV	R^2: 0.79	Compared to the MnLR model, the accuracy of ANN models was higher The ANN models can estimate and optimize the production and quality of biochar from the slow pyrolysis	[71]

Table 1 (continued)

Biomass	Reactor type	Method	Model inputs	Model outputs	Model performance	Major findings	References
Ligno-cellulose	Fixed bed	RF	AC, RT, H, T, PS, L–C–H, C–H–O–N	Biochar and C-char yields	RMSE:3.4	Pyrolysis temperature had a significant impact on the variation in the two outputs. Regarding feedstock characteristics, the structural information proved to be more crucial than element compositions for predicting yield, although the opposite was observed for C-char	[72]
Rice straw, pinewood	Fixed bed (bench-scale)	ANN, ANN-PSO	T, C, H, O, AC, MC, FC, VM	Water, tar, biochar, biogas (CO and CO_2)	Err: 7.7%	PSO-ANN increased the prediction accuracy, particularly for data points located around the database boundary was largely improved. The relative error decreased from 23.07% for ANN model to 7.72% for the PSO-ANN model, which proved to be suitable for modeling biomass oxidative pyrolysis	[73]
Ligno-cellulose	Fixed bed	ANN, GBR, RF, SVM, SHAP	T, RT, N_2, O_2, CO_2, SS, SML, F, MC, VM, FC, AC, C, H, N, O, S, H/C	Solid yields, HHV	RMSE: 0.064	The operating conditions were identified as having the most significant impact on yields, with features from torrefaction reactor properties being equally crucial as those derived from biomass characteristics	[74]

Table 1 (continued)

Biomass	Reactor type	Method	Model inputs	Model outputs	Model performance	Major findings	References
Ligno-cellulose	Fixed bed	CFNN, ABC, CFN-ABC	C, H, N, O, FC, VM, MC, AC, HTT, PS, and HR	Bio-oil yield	R: 0.95	The average difference in the results of CFNN and experiments was below 5%. The analysis results of biomass can be used to calculate the yield of bio-oil	[75]
Ligno-cellulose	Fixed bed	RBF, SVM, RF	C, H, O, MC, AC, VM, FC, and BC	Tar yield, water yield, and char yield	R^2: 0.99	The fusion model (FU) made better predicted accuracy than the three base models. Moreover, FU model significantly varied the feature importance of base models during fusing model process	[76]
Agriculture waste, forest waste, algae	Fixed bed	RF, SVM	C, N, O, HTT, HR, PS, FR	Biogas yield, CO_2, H_2, CO, CH_4	R^2: 0.85	The influence of pyrolysis conditions on yield (55%), CO_2 (73%), and H_2 (81%) was greater than that of biomass characteristics, which was inverse for CO (12%) and CH_4 (38%)	[77]
Ligno-cellulose	TGA	ANN	C, H, O, N, S, VM, FC and Ash	Kinetic Parameters (n, E_a, A)	R^2: 0.99	The biochemical analysis was found to be the highest ~ 38%, elemental and proximate analysis with ~ 29% and ~ 22% in estimating the kinetics	[78]
Hazelnut shell, Almond Shell, Husk, Wheat Straw, Olive Pomace	TGA	ANN	C, H, N, S, FC, VM and Ash	Hemi-cellulose, Cellulose and Lignin	$R^2 > 0.99$	The ANN model predicted the mass loss and DTG values of various biomass resources	[79]

Table 2 Summary of machine learning applications for biomass pyrolysis kinetic parameters prediction

Biomass	Reactor type	Method	Model inputs	Model outputs	Model performance	Major findings	References
Ligno-cellulosic forest residue, olive oil residue	TGA	ANN	T, HR	Ea	R^2: 0.99	The experimental thermal data demonstrated a robust correspondence with the data predicted by the ANN ($R_2 > 0.99$), but a more intricate network topology was necessary for the combined network because of the complexity of the dataset	[85]
Sawdust of different trees, crop straw, crop shell and manures	TGA	EC, ANN, RF	HR, Lg, Hm, Ce	log (K_0), log (Ea), log(n)	R^2: 0.92	Empirical correlations struggle to precisely capture intricate nonlinear relationships, and methods such as ANN and RF are better suited for representing these complex correlations, with the RF model being particularly effective	[86]
Ligno-cellulose	Fixed bed	ANFIS	RT, Hm, Ce, Lg	Ea	R^2: 0.97	The findings indicate the utility of the ANFIS model in forecasting the pyrolysis reaction of lignocellulosic biomass	[87]
Wood, grass, and crops	TGA	CRNN	Ce, O_2, S_2, S_3	Ea	MAE: 0.049	The CRNN-aided model can accurately predict the mass fractions of products and analyze reaction kinetics	[88]
Diverse (Pine sawdust, municipal solid waste etc.)	TGA	RF, CART, ERT	C, O_2, N_2, H_2, FC, S_2, VM, AC	Ea	$R^2 > 0.99$	The input variable importance measurement results indicate that temperature plays the most important role in co-pyrolysis, followed by blending ratio, then ultimate and proximate analysis of samples, and finally heating rate	[89]
Ligno-cellulose	TGA	SGB	C, O, H, N, S, HR	Ea	R^2: 0.99	Sensitivity analysis showed that the heating rate strongly affected results with a relevancy factor of 0.43	[90]
Ligno-cellulose	Batch reactor	ANN, DT	T, HR, C, H	Ea	R^2: 0.99	The developed models exhibit remarkable agreement with the result of simulated reactions in a batch reactor	[91]
Algal	TGA	ANN	Tinst, Ttarget, HR	Mass loss rate	R^2: 0.97	The ANN model projected the results of pyrolysis with various heating rates. The DTG curve generated by the neural network model closely mirrors the DTG curve derived from experimental data obtained through the TGA	[92]

model by utilizing experimental datasets from different types of woody biomass to predict syngas yield and composition. The inputs included the zone temperature, ultimate, and proximate analysis of different woody biomass utilized for the ANN model. The model performance was high, and the predicted output showed adequate agreement with experimental results with an R^2 value between 0.98 and 0.99 [95]. A similar study by George et al. [96], that utilized datasets from five different waste biomass for ANN model training, also showed a high predictive accuracy with an R^2 value of 0.9. These studies suggested that expanding the used dataset is desirable to improve the generalization of the model. Similar studies that utilized ANN for model development and prediction were reported by Ascher et al. [31].

Recent research efforts have aimed to explore alternative ML algorithms for predicting syngas or producer gas yields and composition, driven by the desire to overcome the black-box nature of ANN and to assess the predictive capabilities of different ML approaches. Elmaz et al. [97] conducted a comparative study involving four ML algorithms: decision trees (DT), support vector regression (SVR), polynomial regression (PR), and multilayer perceptron ANN (MLP-ANN). These models were utilized to estimate the yield and compositions of syngas in a downdraft gasifier. The developed models exhibited high predictive performance, with both MLP-ANN and DT models outperforming the other two in predicting the levels of CO, CO_2, CH_4, H_2, and HHV. Similarly, Ozbas et al. [98] also predicted the syngas composition and its HHV using four different ML models, including the KNN, Linear Regression (LR), SVR, and DT. Their predictive performances are also compared. The developed models demonstrated high accuracy, with R-squared (R_2) values exceeding 0.99. Similar studies were also conducted by Fang et al. [99] and Li et al. [100].

An attempt has been made recently by Serrano and Castelló [101] to include the reactor's bed material as one of the input variables for the prediction of syngas compositions and gas yield in a bubbling fluidized bed gasifier. Nine input variables were chosen for the model training in this study. Apart from the bed materials, the remaining input variables include C, H, O, N, S, MC, ash, equivalence ratio (ER), temperature, and steam/biomass ratio. It was noted that the developed models successfully forecasted the compositions and yield of syngas with good accuracy (R^2 of > 0.94 and MSE of $< 1.7 \times 10^{-3}$) [101].

Prediction of outputs other than syngas yield and composition such as solid residues (char, tar, and ash) yield during gasification are believed to be equally important. Solid residue generation can be one of the limiting factors hindering commercial scale utilization. Therefore, the accurate prediction of these solid residue yields could help in establishing an effective solid residue reduction strategy. To this end, Serrano and Castelló [101] developed an ANN model to predict char deposition in a bubbling fluidized bed gasifier. Datasets used for the model training were obtained from experimental tar sampling, collection, analysis methods, and literature. The developed model showed good predictive ability with R^2 of > 0.97 for both testing and validation datasets. In addition, the calculated relative errors were less than 20% for most of the tested samples. Parametric studies conducted with the developed ANN model reveal that the tar yield profile with both temperature and equivalent ratio (ER), exhibits an excellent fit to experimental values reported in the literature, outperforming previous models. A similar study employed various inputs including C, H, N, S, O, moisture, ash content, temperature, and ER to predict LHV of syngas and tars [102]. In the investigation, two ANN models were employed: one with multiple-input and single-output (MISO) and another with multiple-input and multiple-output (MIMO). These were trained using the Levenberg–Marquardt backpropagation algorithm. They showed a high estimation performance with R^2 of > 0.99 for tar and syngas LHV value prediction [102]. A summary of other recently developed ML models for syngas yield and composition, and residue yield during conventional gasification are highlighted in Table 3 (Fig. 8).

Machine Learning Applications in Hydrothermal Gasification Process

Hydro-thermal or supercritical water gasification is preferable to conventional gasification because of its capability to handle feedstock of high moisture content without requiring drying [108]. In addition, the produced syngas from this method are often free from nitrogen and sulfur-containing compounds [109]. Different ML algorithms have been applied to hydrothermal gasification processes, particularly for H_2 yield prediction and optimization, and catalyst selection and screening. ANN, GPR, SVM, and RF models were developed to estimate H_2 yield for supercritical water gasification (SCWG) of agricultural waste and municipal solid waste biomass.

The developed models utilized eight input variables that includes both biomass characteristics: C, H, O, ash; and the SCWG conditions: biomass concentration (BC), gasification temperature (T), pressure (P) and residence time (RT). All the developed ML models showed promising predictive capability with R^2 of > 0.98. They also assessed the effect of biomass properties (C, H, O contents) on H_2 yield and the exergy efficiency. The proposed ML models showed good prospects to be used for future design and optimization of hydro-thermal gasification process [113].

In another study, Shenbagaraj et al. [114] developed a multi-layered feed-forward back-propagation algorithm-aided artificial neural network (FFBPNN) model to evaluate the effect of different factors affecting syngas composition

Table 3 Summary of machine learning applications in conventional gasifiers for syngas and residue yield

Biomass	Reactor type	Method	Model inputs	Model outputs	Model performance	Major findings	References
Municipal solid waste	Fluidized bed	DT, XGB, RF, SVR, MLP-ANN	C, H, N, O, S, MC, AC, ER, T	Syngas yield, composition, LHV, and LHV of other products	R^2: 0.99	Based on the input importance analysis, all input variables are found to affect the modelling of the gasification but with different levels of influence. It turns out that temperature is the most influential variable for predicting LHV, LHVp, and gas yield with relative importance score of 40.8%, 39%, and 35%, respectively	[110]
Woody and plastic	TGA	ANN	PS1 (rubber seed shell), PS2 (HDPE), T, plastic mass	H_2 yield	R^2: 0.99	All the input variables were found to significantly influence the model outputs as indicated by the relative importance analysis	[111]
Wood	Fixed bed	MCS, ANN	T, particle shape, emissivity, thermal conductivity, size, water content, and porosity	Syngas yield	$R^2 > 0.99$	The sensitivity analysis of parameters showed that the main factors for syngas yield are water content, size, and gasification temperature. Other particle characteristics including shape, emissivity, conductivity, and porosity may be unnecessary	[99]
Food waste, sewage sludge, and manure	Fluidized and fixed beds	GBR, ANN, SVR, RF	C, H, O, N, S, AC, ER, T, S/B	Three-phase products (char, tar, and syngas) and syngas yields	R^2: 0.96	The elucidation of the GBR model showed that fuel and operation parameters, such as C, H, O, ash, and temperature were the main factors to affect the yield of syngas and by-products	[100]

Table 3 (continued)

Biomass	Reactor type	Method	Model inputs	Model outputs	Model performance	Major findings	References
Various (from literature)	Fixed, fluidized, others	ANN	Feedstock, operation, gasifying agent, reactor, bed material, catalyst, system scale, CHNSO, AC, MC, VM, FC, PS, LHV, Lg, Hm, Ce, T, RT, S/B, ER	Syngas, char, and tar yields	R^2: 0.93	ANN model for biomass gasifiers can estimate gasification outputs by using various biomass, reactor, and gasification agents	[31]
Wood	Bubbling fluidized bed (lab-scale)	ANN	C, H, O, MC, AC, ER, T, SM	Syngas, tar and LHV	R^2: 0.97	Results from the ANN model agreed with previous results regarding the formation and behavior of tar under various gasification temperature and ER. Especially, the characteristics of tar from the model were compared with experimental results	[101]
Wood	Fixed bed downdraft	ANNs, NARXNN	1. ER, F(air), T (distribution, T0-T5) 2. C, H, O, N, MC,VM, FC, AC, ER, F(air), T0	CO, CO_2, H_2, CH_4 yields, and syngas HHV	$R^2 > 0.99$	Using temperature distribution as features can reduce the number of features (inputs) required for a fixed bed gasifier while enhancing accuracy. The second advantage of the reduced number of features is an increase of the speed in matching	[112]

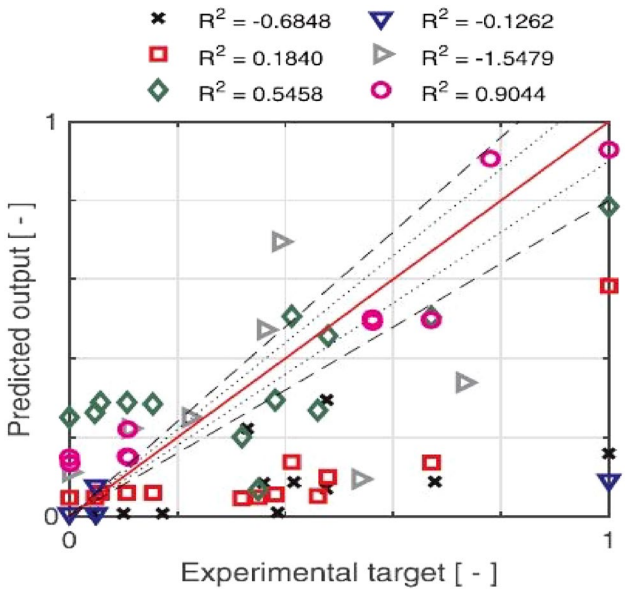

Fig. 8 Comparisons of experimental results with simulation results in the literature [(times symbol) Stark et al. [103]; (white square), (diamond suit) Sridhar [104]; (black down pointing small triangle) Wojnicka et al. [105]; (black right-pointing pointer) Hejazi et al. [106]; (white circle) Rameshkumar and Mayilsamy [107])

during the SCWG of food waste. The model displayed a high prediction accuracy greater than 98% and an MSE value less than 0.3. The effect of feed concentration, reaction temperature and time on syngas composition was also studied [114]. A similar study was reported by Zhang et al. [115] on SCWG of waste biomass using linear regression with the single- and two-hidden layer neural networks. Predictive performance of the single-hidden layer neural network has shown superior performance compared to a two-hidden layer neural network. Li et al. [116] proposed the use of Gradient Boosting Regression (GBR) and hybrid GBR-Particle Swarm Optimization (PSO) algorithms to predict the hydrogen (H_2) yield in a hydro-thermal gasification process. The aim was to address the black-box nature of ML algorithms and ensure the interpretability of the developed models. The GBR model was constructed based on the composition of the feedstock (C, H, O, N, ash) and operation parameters (temperature, pressure, and time), achieving an R^2 value exceeding 0.90. Subsequently, the model was enhanced by integrating Particle Swarm Optimization (PSO). In order to validate the ML models, a comparison was made with mechanistic modeling using Aspen Plus simulation. This comparative analysis between ML algorithms and mechanistic models enhances user confidence in the GBR-PSO models.

An attempt was made by Li et al. [128] for screening and selection of catalysts during the hydro-thermal gasification of waste biomass using ML method of principal component analysis (PCA). Applied datasets were divided into three subcategories of non-catalyst, alkali-metal catalyst, and transition-metal catalyst with the aid of the PCA. The developed model displayed high prediction accuracy in identifying and screening materials for increasing H_2 and decreasing CO_2 during SCWG of waste biomass. Similar studies on catalyst screening and selection for H_2 yield during SCWG of biomass using ML algorithms are reported by Gopirajan et al. [117], Guan et al. [41], and Fózer et al. [118]. A summary of other recently developed ML models for syngas yield and composition, and residue yield during hydro-thermal gasification are highlighted in Table 4.

Machine Learning Applications in Hydrothermal Processes

In HTT, wet biomass is converted into gaseous or solid fuels and other constituents under elevated pressure and temperature [7]. As mentioned above, HTT is classified as hydrothermal carbonation (HTC) and hydrothermal liquefaction (HTL) depending on temperature, pressure, and the proportions of the converted intermediate products [23]. The HTC is usually conducted at relatively low temperatures of between 150 and 260 ℃ and pressure of 1.5–5 MPa to obtain a solid product (hydro-char), while HTL is mostly undertaken at moderate temperature of between 250 and 350 ℃ and pressure 5 ~ 20 MPa in subcritical water to yield liquid fuel (bio-oil) [120]. ML techniques have been applied to various aspects of wet biomass hydro-thermal treatment prediction, particularly for the prediction of biofuel and bio-char.

Machine Learning Applications in Hydrothermal Carbonation

ML algorithms have demonstrated their effectiveness in the field of hydrothermal carbonization (HTC) and the pretreatment of biomass that can transform wet biomass into uniform, carbon-rich hydro-char with a higher heating value (HHV). The ML algorithms have been successfully utilized to predict diverse aspects of the HTC process, including the composition and properties of the final product (hydro-char), process kinetics, nitrogen, and phosphorus content in hydro-char, process optimization, and synergistic catalysis.

Vardiambasis et al. [121] employed four ANN models, developed using 144 datasets, to predict hydro-char yield and HHV during hydro-thermal carbonization of sewage sludge and food waste biomass. The developed models utilized elemental content (C, H, O), HTC temperature, and time as the input variables. The developed ANN models showed promising predictive capability with R^2 of > 0.917. The sensitivity analysis results confirmed that carbon content (C) had the greatest influence on hydro-char yield and HHV. The proposed ML models confirmed good perspectives to be used for future design and optimization of the

Table 4 Summary of machine learning applications in SCWG for syngas and residue yield prediction

Biomass	Reactor type	Method	Model inputs	Model outputs	Model performance	Major findings	References
Waste	Stainless steel reactor (lab-scale)	TDSS, TRS	C, H, O, N, S, MC, VM, AC, T, P, SV/B, RT, mass of catalyst	H_2 yield, CO_2, and CO	Accuracy: 94%	Results showed that more than 94% accuracy was achieved with all sets of data and non-suitable input variables were directing the system to a specially devised Tunable Recommendation System	[117]
Food waste	Batch reactor setup	ANN	T, RT, BC	CO, CO_2, H_2, CH_4 yields, catalyst selection	$R^2 > 0.98$	The syngas composition and yield predicted by the FFBPNN models correspond with experimental results. Its mean squared error was below 0.300 and R_2 was above 98%	[114]
Microalgae and waste	Plug-flow tubular reactor	ANN	C, H, O, N, S, T, P, B/W, RT, Cat/Susp	Bio-methanol	$R^2 > 0.97$	The ANN model was used to evaluate techno-economic analysis. Depending on the supply processes of hydrogen, bio-ethanol costs varied in the range of 316–339 US $ $(tMeOH)^{-})^{-1}$	[118]
Wood (lignite)	Quartz tube reactor	ANN	H/C, O/C, N/C, AC, protein, lipid, carbohydrate, T, Solid content	Syngas yield and composition	R^2: 0.99	For the variable contribution analysis, the ranking of the sensitivity of each gas is almost $H_2 > CO_2 > CH_4 > CO$. What's more, the MIV of temperature, residence time, and concentration are 0.7493, 0.2188, − 0.1051, respectively. The temperature has the largest contribution to gas production, and the next is residence time and concentration	[115]

Table 4 (continued)

Biomass	Reactor type	Method	Model inputs	Model outputs	Model performance	Major findings	References
Wet waste	STOICH reactor and batch setup	GBR, GBR-PSO	C, H, N, O, AC, SD (solid content), T, P	Syngas yield	R^2: 0.99	The optimal conditions recommended by the ML-based optimization for H_2-rich syngas production were validated by Aspen simulation and the error was found to be around 10% and 20% for H_2 and CO_2 yields	[116]
Horti-cultural waste and sewage sludge	Fluidize and fixed beds	LR, ANN, SVM, GPR	T, S/B, O, AC, F, mass of biomass, adsorbent to biomass, PS, C	H^2-rich syngas	$R^2 > 0.99$	The sensitivity analysis showed that all factors influenced gasification results. However, gasification temperature is the most dominant parameters for gasification results	[119]

hydro-thermal gasification process [121]. Zhu et al. [72] developed an RF model to predict the yield, HHV, and carbon in char of municipal waste during HTC operation. The predictive results showed a high accuracy with R^2 of 0.80, 0.91, and 0.95 for hydro-char yield, HHV, and C_ char, respectively. Similar studies that utilized ML models for HTC process prediction can be found in Li et al. [71] , Ismail et al. [122], Kapetanakis et al. [123], and Mu et al. [124].

The prediction of process kinetics during HTC of cellulose, poplar, and wheat straw biomass was reported by Aghaaminiha et al. [125]. Four different ML models of KNN, ANN, SVR, and DF employed 132 datasets with input variables of experimental type (isothermal, dynamic), temperature, time, nitrogen content, sulfur content, and hydrogen content. The developed models showed high predictive accuracy in all the scenarios investigated. In addition, the performance of the developed ML methods in interpolating kinetics results was evaluated, considering situations where experimental data is limited to only a few time-points. An 'extrapolation model' was additionally developed using kinetics data from the first three time-points as input. The kinetic data for subsequent time steps was predicted to evaluate its capability in extrapolating kinetics when data is available from only a small number of initial time-points. Djandja et al. [126] developed an RF model for HTC processes to access the phosphorous concentration of char based on 109 data points. Particle properties such as FC, VM, ash, moisture, C, H, O, N, S, and phosphorous concentration, and operation conditions including temperature, residence time, pH of feedwater, and dry matter content (DM), were considered as features for designing the model. The model predictive performance was high with R^2 of > 0.92–0.95. In addition, operation conditions exhibit a positive effect on the phosphorous concentration in char [126]. A similar study by Djandja et al. [127] was developed to predict the nitrogen content in hydro-char. A summary of recently developed ML models for HTC is highlighted in Table 5 (Fig. 9).

Machine Learning Applications in Hydro-thermal Liquefaction

Hydro-thermal liquefaction (HTL) is considered as a viable option for producing bio-crude oil from wet biomass with varying moisture content, which can be upgraded to be used as a transportation fuel and for chemicals. HTL also produces solid biochar and a gaseous fraction. It has some advantages over the classical process, as it has a lower processing time and a higher yield [130]. The process is mainly dependent on feed quality, reaction time, temperature, catalyst, and mixing ratio. The application of ML in HTL is mainly for estimating the yield and composition of bio-oil.

Cheng et al. [131] developed ML models by utilizing 570 experimental datasets for HTL of different biomass

Table 5 Summary of the existing research on machine learning applications for hydrothermal carbonization of wet biomass

Biomass	Reactor type	Method	Model inputs	Model outputs	Model performance	Major findings	References
Poultry litter	8.0 L Parr stirred pressure reactor (lab-scale)	ANN	RT, T	Hydro-char yield	$R^2 > 0.94$	Reaction temperature was the most important factor for HTC processes, as the rate of change in inorganic phosphorus (IP) and carbon recovery exhibits a much greater variation with temperature than with changes in time	[122]
Sewage sludge, food waste, and manure	Fluidized, fixed, others	DNN	C, H, N, O, FC, AC, VM, MC, RT, T,	CO, CO_2, H_2, CH_4 yields, catalyst selection	$R^2 > 0.91$	Based on the importance analysis, particle properties and operation conditions were critical to the estimation of process performance	[128]
Sewage sludge	Fluidized, fixed, others	ANN	RT, T, C, H, O	Hydro-char yield and HHV	$R^2 > 0.97$	The ANN models were successfully applied to access HHV and contents of C and H	[123]
Wood biomass, herbaceous biomass, food waste	Fluidized, fixed, others	ANN, ANN-PSO	C, H, O, N, S FC, AC, VM, T, RT, B/W	Fuel properties of hydro-char	$R^2 > 0.85$	The results showed that C content and temperature were important to the prediction of the mass yield of hydro-char and the O content had positively affected the char yield	[124]
Sewage sludge	Fluidized, fixed, others	ANN	C, H, O, N, VM, FC, AC, T, RT, Mass of water	Nitrogen content of hydro-char	$R^2 > 0.98$	The significant factors found to predict the nitrogen in sewage sludge-derived hydro-char were found to be temperature, C, N, VM, and FC of materials	[126]
Wheat straw, corn straw, and sorghum straw	TGA	LR, ANN, SVR, RF	HR, T, BC	Iodine number, Methylene blue number	$R^2 > 0.99$	The evaluation indicators were all good. Also, straws were effectively used as precursors to produce activated carbon which can be utilized for the removal of dyes from wastewater	[129]

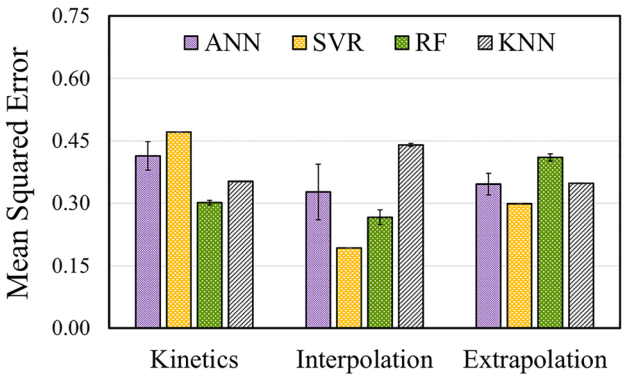

Fig. 9 Comparison of the predictive performance of the machine learning models [125]

feedstocks in literature (algae, lignocellulosic biomass, food waste, manure, sludge, bioethanol residue, municipal solid waste, and seed plants) to predict bio-crude yield. Eight ML models including MLR, Ridge regression (RR), Lasso regression (LR), SVR, DT, MLP-ANN, RF, and XG Boost were developed for this purpose, and their performance was compared. The performance of the above models was relatively high, and results showed good agreement with experiments. This RF model was adjudged the optimal model with an RMSE of 8.07. The accuracy of the optimal RF model and a probabilistic economic analysis are found to be enough to arrange the order of resources delayed on the basis of the estimated minimum fuel price [131]. Katongtung et al. [132] predicted bio-crude yields and HHV of biomass resources in HTL processes by using 17 inputs in 325 datasets. Novel ML algorithms, including SVR, Kernel ridge regression (KRR), RF, and extreme GB (XGB) based on tenfold cross-validation were adopted for the prediction purpose. The developed models showed a high predictive accuracy with an R^2 value of > 0.9. Similar studies that utilized ML algorithms for bio-crude and composition prediction from HTL of wet biomass can be found in the literature [133, 134].

In another study, two ML models of DT and RF with 257 datasets were developed to predict pH, TOC, TP, and TN in the aqueous phase (AP) produced in HTL processes by using algae, food waste, sludge, and manure [135]. Both developed models presented high predictive performance, with the DT showing better performance than RF for both single and multi-target predictions. In addition, the ML-based feature importance and partial dependence analysis showed that temperature, solid, and nitrogen content were mostly important factors for pH, TN, and TP. A similar study used both GBR and RF to predict bio-oil yield, content of oxygen (O_ oil), and nitrogen (N_ oil) from HTL of wet biomass [115]. The results indicated that the GBR, with an averaged R^2 of > 0.90 for the test set, displayed better predictive capability than the RF model for both single and multi-target task prediction. A summary of other recently developed ML models for bio-crude and its composition prediction from HTL of biomass are highlighted in Table 6.

Machine Learning Application in Combustion Processes

In the combustion process, organic matters are thermally converted into CO_2 and H_2O with the help of an oxidant, generally oxygen. The combustion process generates heat and electricity for home and industrial use. The process can be classified into single combustion or co-combustion of two fuels depending on the fuel type [7]. In the combustion process, ML can be employed for modeling boiler wall at different operating conditions, fault diagnosis, automation in the generation process, and the prediction of gaseous pollutants, thermal properties of biomass, and output energy [136]. The thermal combustion characteristics, which include combustion reactivity [137], heat capacity [138, 139], oxidation kinetic parameters [140–144], and co-combustion of biomass are also included [145–147].

The presence of moisture content in biomass significantly affects its properties and combustion characteristics. Taking this into account, Rico-Contreras et al. [148] devised a prediction model based on FL and ANN to estimate the thermal properties of poultry litter. They utilized various input variables including density, temperature, duration of storage, and feedstock handling, which affect the moisture content of the feedstock. A total of 108 samples were tested with this artificial framework and an average moisture content of 30.16% was reported. In addition, the developed FL-ANN system showed an accuracy of 92.88% when modelled with 20 different farm study results. Importantly, this proposed model holds applicability across a wide range of bioenergy generation systems, encompassing not only the poultry sector but also other combustion or alternative processes. Krzywanski et al. [33] developed an FL model to predict emission contents from both wood chips biomass and bituminous coal in a lab-scale 5 KW dual fluidized chemical looping combustor. This model estimated various emissions including CO_2, CO, NOx, and SOx. These pollutants were made across conventional and novel combustion processes including air-firing, oxy-fuel combustion, chemical looping combustion, and so on. The developed FL model displayed a high predictive capability with a maximum relative error between target and prediction lower than 8%. Prediction by the FL models for gaseous pollutant emissions was in good agreement with experimental results. A comparable study was conducted by Li et al. [149] and Li et al. [150] to investigate NOx emitted from biomass combustors using the DNN. In both studies, the predicted NOx emissions exhibit good agreement with the measurement results. Further studies on GHG emissions

Table 6 Summary of pioneering investigation on machine learning application for hydrothermal liquefaction of wet biomass

Biomass	Reactor type	Method	Model Inputs	Model outputs	Model performance	Major findings	References
Ligno-cellulose, algae, sewage sludge, manure, and municipal waste	Fluidized, fixed, others	ANN, GAM, SVR, GPR	C, H, O, N, S, AC, T, P, RT, B/W	Bio-crude oil yield, bio-crude oil composition, aqueous phase yield, hydro-char yield, syngas yield	$R^2 > 0.93$	Gaussian process regression (GPR) model was adjudged the optimized. The mean absolute error of GPR model was lower than 0.031	[134]
Pine wood, castor residue etc	Hydro-thermal reactor (4598, Parr reactor)	DSS	C, H, N, O, MC, AC, T, P, RT, SV/B, catalyst, solvent	Bio-oil yield and HHV	$R^2 > 0.99$	More than 94% accuracy was obtained for the predicted results and about 95% accuracy was achieved for targeted optimization	[133]
Microalgae, crops/forest residues, energy crops, bio-degradable organic wastes	Fluidized, fixed, others	MLR, RF, DT	O/C, H/C, AC, T, RT, Mass of biomass	Bio-crude yield, hydro-char yield, ACP yield, gas yield, energy, and carbon contents of bio-crude and hydro-char	$R^2 > 0.90$	Results indicated that the RF models outperformed other models in modeling HTT of feedstocks and predicting the mass yields of various products (biocrude, hydro-char, as well as the energy and carbon contents of hydro-char and biocrude	[80]

prediction from biomass-fired combustion process can be found in [151–153].

To become a sustainable society, carbon-neutral fuels like biomass, hydrogen, and ammonia should be utilized for heat and electricity and their usage consequentially increases in most countries. Therefore, HHV is an important factor in assessing the economic feasibility of biomass utilization. Noushabadi et al. [154] developed and proposed an ML methods framework to evaluate the HHV of biomass by using C, H, N, O, and S weight fractions. A total of 535 datasets were utilized to construct five ML models. Different types of biomasses, such as fruits, agricultural wastes, grasses/leaves/fibrous materials, wood chips/tree species, and various organic wastes including municipal solid wastes, were utilized in combustion processes. The developed models demonstrated excellent predictive performance, as evidenced by the low average absolute relative deviation (AARD) values.

Specifically, the MNR and GA-RBF algorithms achieved an AARD of 3.5% and 3.4%, respectively, indicating their suitability for estimating the HHV of biomass. You et al. [155] estimated the low heating value of biomass (municipal solid waste, MSW) using MLP-ANN, ANFIS, SVM, and RF models. The result indicated that a high-accuracy ML model could improve the CFB operation and contribute stable energy supply [155]. Further studies on the prediction of HHV from biomass-fired combustion processes can be found in [47, 86, 156–159]. A summary of other recently developed ML models for thermo-chemical combustion of biomass is highlighted in Table 7.

Futuristic Prospects

As discussed, ML is widely applied to biomass thermochemical conversion processes. However, there needs to be a more detailed description of the procedure of ML development to improve further. For example, in many instances, the 'number of epochs' is used as the termination criterion when the desired correlation (R^2) and minimum RMSE are not achieved [160]. In this chapter, some challenges, and prospects in the application of ML to biomass thermochemical conversion are introduced with respect to the datasets and training and testing of the ML model. Furthermore, relying on a single statistical index is inadequate as it only provides a limited perspective on model errors, focusing on specific aspects of the error characteristics. To comprehensively assess model performance, a combination of statistical indices, including variance accounted for (VAF), mean absolute percentage error (MAPE), and mean squared deviation (MSD), should be employed. By considering multiple indices, a more thorough evaluation of the model's accuracy and performance can be achieved.

Previous research has revealed a scarcity of data collected directly from biomass thermochemical experiments because of the costly and time-consuming nature of the data acquisition process. To overcome this challenge, researchers often depend on importing experimental data from various sources. Therefore, employing these sets of experimental data often results in models with good predictive accuracy but low extrapolability since these models are only compatible with a specific dataset. The developed models are only applicable to experimental data within the range of the data they were trained on. Integrating ML models with theoretical modeling approaches, such as kinetics, thermodynamics, and CFD studies, to simulate the experimental conditions should provide a high-quality dataset. These datasets provide extensive information regarding fundamental reaction mechanisms and pathways, facilitating the development of optimal models during the training process and improving the predictive capabilities of the models.

Most ML algorithms are called "black box" because the relationships between the input and output variables are, most of the time, complex and are not always depicted in the form of tractable mathematical equations that can be easily understood. On the other hand, the advantages of developing a reactor model with detailed reaction kinetics and hydrodynamics for biomass thermochemical conversion processes lie in its better extrapolability and interpretability compared to ML models. This becomes particularly important when dealing with the inherent variability in biomass feedstock. While a physically and chemically consistent feedstock is required for the robust operation of the process, the inconsistency within a single resource due to varying growth and harvesting conditions [161], especially during climate crises, can present challenges in black-box modeling. Thus, elucidating the reaction pathways of biomass decomposition under various conditions is necessary for the robust operation of biomass thermochemical conversion processes. To enhance the extrapolability of ML models, it has been suggested to develop ML models guided by theoretical models. The methodology for such a combination was described elsewhere [39]. Future work should be directed towards unboxing the black-box nature to ease the implementation and increase the reproducibility and interpretability of these ML models. These models should be converted into tractable mathematical equations, allowing quantitative interpretation and accurate prediction. Additionally, researchers have proposed the use of feature permutations to examine the behavior and relevance of various input variables. This approach enabled a comprehensive evaluation of the significance of each input variable in the ML model [31], increasing the interpretability of the model. Moreover, it is advisable to incorporate sensitivity and techno-economic analyses to assess the influence of each predictor or feature

Table 7 Summary of machine learning applications for biomass combustion

Biomass type	Reactor type	Method	Model inputs	Model outputs	Model performance	Major findings	References
Oily sludge	Fixed, fluidized, others	MLR	C, H, O, N, S, AC	HHV of oily sludge	$R^2 > 0.95$	It was found that the accuracy of the correlation is improved for the samples with HHV higher than 20 MJ/kg, which are more favorable to be utilized as fuel resources	[156]
Biomass solid fuel, waste solid fuel	Fixed, drop-tube reactor, etc	ANN	C, H, O, N, S	HHV of biomass solid fuel, waste solid fuel, and their combination	$R^2 > 0.91$, 0.90, and 0.87	The developed optimal ANN network architecture model was 5–25-1. This model effectively estimated the HHV of solid fuels from their ultimate analysis information with high R values	[158]
Commercial fuels, wastes, forest wastes, energy crops and cereals	TGA	ANN	C, H, O, N, S, VM, AC, FC	HHV	$R^2 > 0.96$	The ANN model results about HHV by using C, H, O, N, S, ash, and moisture contents were highly reliable with 0.962 of average correlation coefficient	[47]
Hazelnut shell, almond shell, barley straw, wheat straw, olive pomace, rice husk and straw	TGA	ANN	C, H, O, HR, T	Combustion reactivity (instantaneous combustion index)	$R^2 > 0.99$	At 40 °C/min heating rate, reactivity results from ANN model were reliable with 0.98 of correlation coefficient	[137]
Cellulosic biomass	Batch set up	ANN, LSSVR, ANFIS	S, AC, T, Crystallinity Index	Heat capacity of cellulosic biomass	$R^2 > 0.99$	Ranking analysis using seven uncertainty indices over 819 intelligent models from three different classes (i.e., ANN, SVM, and ANFIS) revealed the least- squares vector machine are most useful approach	[138]

Table 7 (continued)

Biomass type	Reactor type	Method	Model inputs	Model outputs	Model performance	Major findings	References
Soya husk, rice husk, bagasse, and peanut shell, eucalyptus wood and coconut shell	TGA	MLP-ANN	C, H, O, AC, VM, T, HR	Activation energy of biomass wastes	RMSE: 0.23	The regression coefficient (R^2) was found to be 1 and predicted the same findings as in the experiment, with very few errors in the calculations during the testing and evaluation of the input parameters through the neural network perceptron	[143]
Sewage sludge, coffee industry residues	TGA	ANN, PCA	T, HR, Mixing ratio	Combustion behavior	$R^2 > 0.99$	1st component variance and 2nd component variance were 54.43% and 45.06%. Cumulative variance was 99.49%	[146]
Peanut hull	TGA (TG–DTA, Setaram Labsys, France)	ANN, MSO, MCS	Blending ratio, HR, T	Co-combustion of peanut hull and coal	$R^2 > 0.99$	The strong consistency of experiments and estimation was caused by the high accuracy of ANN model to estimate the mass loss in co-combustion processes	[145]

on the predicted thermochemical conversion process and the economic importance, respectively.

Several ML algorithms employed to study biomass thermochemical conversion processes have shown good relevance. Nevertheless, even though ML algorithms offer several advantages over conventional modeling techniques, such as CFD, it is crucial to remark that no single ML algorithm can effectively address all engineering problems related to conversion processes, as stated by the No-Free-Lunch (NFL) theorem [162]. Advancements in conversion technologies, such as carbon capture and sequestration processes and liquid fuel transportation, which require additional inputs and increase the complexity of the models. At present, accurate predictive models for these conversion technologies have been extensively developed. Future studies should focus on applying other existing ML algorithms that are yet to be used to biomass thermochemical conversion processes while also developing new algorithms for these advanced technologies. Thus, researchers will be able to develop ML models with higher predictive accuracy and interpretability.

Assessing the predictive performance of ML models is as crucial as training them. During the performance evaluation, the predicted outputs of the models were compared with relevant observations or experimental data using various statistical measures or error metrics, commonly known as statistical indices. These statistical metrics illustrate the disparities between the anticipated and observed values of the outputs, specifically indicating the extent of the residuals' dispersion. The statistical indices used to appraise the performance of ML models have several limitations. The two primary error metrics, namely the mean square error (MSE) and root mean square error (RMSE) are regarded as less reliable owing to their sensitivity to different data fractions and their susceptibility to outliers [163]. In particular, the RMSE is affected by extreme scores, which can hinder the convergence of errors within specified tolerance limits.

Conclusions

This study provides a comprehensive review of published articles that focus on the application of ML techniques in biomass thermochemical conversion processes. Compared to conventional modeling techniques such as computational fluid dynamics (CFD), thermodynamic and kinetic models, process models, and ML techniques offer several advantages, including accuracy, efficiency, simplicity, and robustness in modeling the complex nature of these processes. Various ML algorithms have been successfully used to predict yields and kinetics and optimize pyrolysis, gasification, HTT, and combustion processes, demonstrating high predictive performance and accuracy. Among these techniques, ANNs have been widely utilized by researchers owing to their capacity to manage intricate associations between input and output variables without prior knowledge of their interconnections. However, a limitation of ANN and other ML models is their lack of interpretability because the relationships between inputs and outputs can be intricate. To improve the understanding and interpretability of these processes, it is crucial to develop models that are easy to interpret. Ensuring the interpretability of the developed ML models and improving their predictive accuracy is imperative for large-scale and industrial deployment of these processes. Fine-tuning the hyperparameters of the developed process models using metaheuristic algorithms could help ensure higher predictive accuracy of the processes. Establishing broad input and output variable datasets could aid in improving the model performance because it has been established that the model performance increases with the increasing number of datasets.

Consequently, there is a notable need for the development of additional ML algorithms and the inclusion of a wider range of operating conditions and parameters in the datasets used for model development. This approach can help to create universal models with high predictive accuracy. By using novel algorithms and incorporating comprehensive data, researchers can enhance the capabilities of ML models and ensure their applicability to different biomass thermochemical conversion processes. This will ultimately contribute to the advancement of the field and facilitate more accurate predictions in various operational scenarios.

Acknowledgements This work was supported by a Grant from the National Research Foundation of Korea, which was funded by the Ministry of Education, Science and Technology (2023R1A2C2004497 and RS-2023-00281706) and the research grant of the Kongju National University in 2022.

Data availability The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

1. J.L.J. Ling, H.W. Kim, E.S. Go, S.S. Oh, H.J. Park, C.S. Jeong, S.H. Lee, Analysis of operational characteristics of biomass oxygen fuel circulating fluidized bed combustor with indirect supercritical carbon dioxide cycle. Energy Convers. Manag **259**, 115569 (2022)
2. G.C. Umenweke, I.C. Afolabi, E.I. Epelle, J.A. Okolie, Machine learning methods for modeling conventional and hydrothermal gasification of waste biomass: a review **17**, 100976 (2022)
3. M. Mandegari, S. Farzad, J.F. Görgens, A new insight into sugarcane biorefineries with fossil fuel co-combustion: techno-economic analysis and life cycle assessment. Energy Convers. Manag. **165**, 76 (2018)
4. W.S. Chai, Y.L. Bao, P.F. Jin, G. Tang, L. Zhou, A review on ammonia, ammonia-hydrogen and ammonia-methane fuels. Renew. Sustain. Energy Rev. **147**, 111254 (2021)
5. S. Nanda, S.N. Reddy, S.K. Mitra, J.A. Kozinski, The progressive routes for carbon capture and sequestration **4**, 99 (2016)

6. R. Lee, Y.R. Gwak, J.M. Sohn, S.H. Lee, The prediction of co2 emissions in domestic power generation sector between 2020 and 2030 for Korea. **32**, 855 (2021)
7. M.W. Seo, S.H. Lee, H. Nam, D. Lee, D. Tokmurzin, S. Wang, Y.K. Park, Recent advances of thermochemical conversion processes for biorefinery. Bioresour. Technol. **343**, 126109 (2022)
8. P. Debiagi, G. Gentile, A. Cuoci, A. Frassoldati, E. Ranzi, T. Faravelli, A predictive model of biochar formation and characterization. J. Anal. Appl. Pyrol. **134**, 326 (2018)
9. D. Lee, H. Nam, M.W. Seo, S.H. Lee, D. Tokmurzin, S. Wang, Y.K. Park, Recent progress in the catalytic thermochemical conversion process of biomass for biofuels. Chem. Eng. J. **447**, 137501 (2022)
10. F.X. Collard, J. Blin, A review on pyrolysis of biomass constituents: mechanisms and composition of the products obtained from the conversion of cellulose, hemicelluloses and lignin. Renew. Sustain. Energy Rev. **38**, 594 (2014)
11. S. Hameed, A. Sharma, V. Pareek, H.W. Wu, Y. Yu, A review on biomass pyrolysis models: kinetic, network and mechanistic models. Biomass Bioenergy **123**, 104 (2019)
12. M. Madadi, A. Abbas, Lignin degradation by fungal pretreatment: a review **8**, 1 (2017)
13. S. Brethauer, M.H. Studer, Biochemical conversion processes of lignocellulosic biomass to fuels and chemicals—a review. Chimia (Aarau) **69**, 572 (2015)
14. C.Y. Lin, W.S. Chai, C.H. Lay, C.C. Chen, C.Y. Lee, P.L. Show, Optimization of hydrolysis-acidogenesis phase of swine manure for biogas production using two-stage anaerobic fermentation. Processes **9**, 1324 (2021)
15. K. Kim, Y. Kim, C. Yang, J. Moon, B. Kim, J. Lee, U. Lee, S. Lee, J. Kim, W. Eom, S. Lee, M. Kang, Y. Lee, Long-term operation of biomass-to-liquid systems coupled to gasification and fischer-tropsch processes for biofuel production. Bioresour. Technol. **127**, 391 (2013)
16. S.H. Lee, M.S. Eom, K.S. Yoo, N.C. Kim, J.K. Jeon, Y.K. Park, B.H. Song, S.H. Lee, The yields and composition of bio-oil produced from quercus acutissima in a bubbling fluidized bed pyrolyzer. J. Anal. Appl. Pyrol. **83**, 110 (2008)
17. S.H. Lee, T.H. Lee, S.M. Jeong, J.M. Lee, Economic analysis of a 600 mwe ultra supercritical circulating fluidized bed power plant based on coal tax and biomass co-combustion plans. Renew. Energy **138**, 121 (2019)
18. J.A. Okolie, S. Nanda, A.K. Dalai, F. Berruti, J.A. Kozinski, A review on subcritical and supercritical water gasification of biogenic, polymeric and petroleum wastes to hydrogen-rich synthesis gas. Renew. Sustain. Energy Rev. **119**, 109546 (2020)
19. A.V. Bridgwater, Review of fast pyrolysis of biomass and product upgrading. Biomass Bioenergy **38**, 68 (2012)
20. K. Yogalakshmi, P. Sivashanmugam, S. Kavitha, Y. Kannah, S. Varjani, S. AdishKumar, G. Kumar, Lignocellulosic biomass-based pyrolysis: a comprehensive review **286**, 131824 (2022)
21. G.B. Chen, J.W. Li, H.T. Lin, F.H. Wu, Y.C. Chao, A study of the production and combustion characteristics of pyrolytic oil from sewage sludge using the taguchi method. Energies **11**, 2260 (2018)
22. S.B. Seo, E.S. Go, J.L.J. Ling, S.H. Lee, Techno-economic assessment of a solar-assisted biomass gasification process. Renew. Energy **193**, 23 (2022)
23. S.R. Wang, G.X. Dai, H.P. Yang, Z.Y. Luo, Lignocellulosic biomass pyrolysis mechanism: a state-of-the-art review. Prog. Energy Combust. **62**, 33 (2017)
24. S.B. Seo, H. Ahn, E.S. Go, L.J.J. Ling, N.J. Siambun, Y.K. Park, S.H. Lee, Evaluation of the solar thermal storage of fluidized bed materials for hybrid solar thermo-chemical processes. Biomass Convers. Bioresour. **1** (2022)
25. E.S. Go, H.W. Kim, S.Y. Kang, S.I. Keel, J.L.J. Ling, S.H. Lee, Analysis of combustion characteristics using cpfd in 0.1 mw oxy-fuel cfb. Int. J. Nanotechnol. **19**, 63 (2022)
26. Y.R. Gwak, J.H. Yun, S.I. Keel, S.H. Lee, Numerical study of oxy-fuel combustion behaviors in a 2mwe cfb boiler. Korean J. Chem. Eng. **37**, 1878 (2020)
27. J.W. Kook, I.S. Gwak, Y.R. Gwak, M.W. Seo, S.H. Lee, A reaction kinetic study of CO2 gasification of petroleum coke, coals and mixture. Korean J. Chem. Eng. **34**, 3092 (2017)
28. K.I. Wong, P.K. Wong, C.S. Cheung, C.M. Vong, Modelling of diesel engine performance using advanced machine learning methods under scarce and exponential data set. Appl. Soft Comput. **13**, 4428 (2013)
29. T. Chen, X.K. Ku, J.Z. Lin, H. Ström, Cfd-dem simulation of biomass pyrolysis in fluidized-bed reactor with a multistep kinetic scheme. Energies **13**, 5358 (2020)
30. M. Aghbashlo, W.X. Peng, M. Tabatabaei, S.A. Kalogirou, S. Soltanian, H. Hosseinzadeh-Bandbafha, O. Mahian, S.S. Lam, Machine learning technology in biodiesel research: a review. Prog. Energy Combust. **85**, 100904 (2021)
31. S. Ascher, W. Sloan, I. Watson, S.M. You, A comprehensive artificial neural network model for gasification process prediction. Appl. Energy **320**, 119289 (2022)
32. A.D. Ferreira, S.D. Ferreira, S.R.D. Neto, Study of the influence of operational parameters on biomass conversion in a pyrolysis reactor via cfd. Korean J. Chem. Eng. **40**, 2787 (2023)
33. J. Krzywanski, T. Czakiert, W. Nowak, T. Shimizu, A. Zylka, K. Idziak, M. Sosnowski, K. Grabowska, Gaseous emissions from advanced clc and oxyfuel fluidized bed combustion of coal and biomass in a complex geometry facility: a comprehensive model. Energy **251**, 123896 (2022)
34. Z. Lian, Y. Wang, X. Zhang, A. Yusuf, L. Famiyeh, D. Murindababisha, H. Jin, Y. Liu, J. He, Y. Wang, Hydrogen production by fluidized bed reactors: a quantitative perspective using the supervised machine learning approach **4**, 266 (2021)
35. Z. Wang, X. Peng, A. Xia, A.A. Shah, Y. Huang, X. Zhu, X. Zhu, Q. Liao, The role of machine learning to boost the bioenergy and biofuels conversion. Bioresour. Technol. **343**, 126099 (2022)
36. D. Baruah, D.C. Baruah, Modeling of biomass gasification: a review. Renew. Sustain. Energy Rev. **39**, 806 (2014)
37. T.K. Patra, P.N. Sheth, Biomass gasification models for downdraft gasifier: a state-of-the-art review. Renew. Sustain. Energy Rev. **50**, 583 (2015)
38. A. Ramos, E. Monteiro, A. Rouboa, Numerical approaches and comprehensive models for gasification process: a review. Renew. Sustain. Energy Rev. **110**, 188 (2019)
39. P.R. Jeon, J.H. Moon, N.O. Ogunsola, S.H. Lee, J.L.J. Ling, S.M. You, Y.K. Park, Recent advances and future prospects of thermochemical biofuel conversion processes with machine learning. Chem. Eng. J. **471**, 144503 (2023)
40. S.M.H.D. Perera, C. Wickramasinghe, B.K.T. Samarasiri, M. Narayana, Modeling of thermochemical conversion of waste biomass—a comprehensive review. Biofuel Res. J. **8**, 1481 (2021)
41. Y. Guan, D. Chaffart, G. Liu, Z. Tan, D. Zhang, Y. Wang, J. Li, L. Ricardez-Sandoval, Machine learning in solid heterogeneous catalysis: recent developments, challenges and perspectives **248**, 117224 (2022)
42. M.W.H. Wang, J.M. Goodman, T.E.H. Allen, Machine learning in predictive toxicology: recent applications and future directions for classification models. Chem. Res. Toxicol. **34**, 217 (2021)
43. P. Sahoo, I. Roy, Z. Wang, F. Mi, L. Yu, P. Balasubramani, L. Khan, J.F. Stoddart, Multicon: a semi-supervised approach for predicting drug function from chemical structure analysis. J. Chem. Inf. Model. **60**, 5995 (2020)
44. A.I. Lawal, G.O. Oniyide, S. Kwon, M. Onifade, E. Köken, N.O. Ogunsola, Prediction of mechanical properties of coal

from non-destructive properties: a comparative application of mars, ann, and ga. Nat. Resour. Res. **30**, 4547 (2021)
45. A.I. Lawal, S. Kwon, Application of artificial intelligence to rock mechanics: an overview. J. Rock Mech. Geotech. **13**, 248 (2021)
46. A.K. Jain, J.C. Mao, K.M. Mohiuddin, Artificial neural networks: a tutorial. Computer **29**, 31 (1996)
47. F. Güleç, D. Pekaslan, O. Williams, E. Lester, Predictability of higher heating value of biomass feedstocks via proximate and ultimate analyses—a comprehensive study of artificial neural network applications **320**, 123944 (2022)
48. B. Doerr, J. Lengler, T. Kötzing, C. Winzen, Black-box complexities of combinatorial problems. In *Proceedings of the 13th Annual Conference on Genetic and Evolutionary Computation*, pp. 981–988 (2011)
49. T. Hastie, R. Tibshirani, J.H. Friedman, J.H. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (Springer, Berlin, 2009)
50. M. Kannangara, R. Dua, L. Ahmadi, F. Bensebaa, Modeling and prediction of regional municipal solid waste generation and diversion in Canada using machine learning approaches. Waste Manag. **74**, 3 (2018)
51. S.K. Shreyas, A. Dey, Application of soft computing techniques in tunnelling and underground excavations: state of the art and future prospects. Innov. Infrastruct. Solut. **4**, 1 (2019)
52. A. Liaw, M. Wiener, Classification and regression by randomforest. **2**, 18 (2002)
53. E. Scornet, G. Biau, J.P. Vert, Consistency of random forests. Ann. Stat. **43**, 1716 (2015)
54. X.C. Nguyen, Q.V. Ly, W. Peng, V.H. Nguyen, D.D. Nguyen, Q.B. Tran, T.T. Huyen Nguyen, C. Sonne, S.S. Lam, H.H. Ngo, P. Goethals, Q.V. Le, Vertical flow constructed wetlands using expanded clay and biochar for wastewater remediation: a comparative study and prediction of effluents using machine learning. J. Hazard. Mater. **413**, 125426 (2021)
55. A. Sharma, Decision tree vs. random forest–which algorithm should you use? **12** (2020)
56. R.M. Hlihor, M. Diaconu, F. Leon, S. Curteanu, T. Tavares, M. Gavrilescu, Experimental analysis and mathematical prediction of cd(ii) removal by biosorption using support vector machines and genetic algorithms. New Biotechnol. **32**, 358 (2015)
57. V. Vapnik, S. Golowich, A. Smola, Support vector method for function approximation, regression estimation and signal processing. Adv. Neural Inf. Process. Syst. (1996)
58. W.S. Noble, What is a support vector machine? **24**, 1565 (2006)
59. H.F. Wang, B.C. Zheng, S.W. Yoon, H.S. Ko, A support vector machine-based ensemble algorithm for breast cancer diagnosis. Eur. J. Oper. Res. **267**, 687 (2018)
60. J.J. Verbeek, N. Vlassis, B. Krose, Efficient greedy learning of gaussian mixture models. Neural Comput. **15**, 469 (2003)
61. A. Mahmoodzadeh, M. Mohammadi, A. Daraei, H. Farid Hama Ali, A. Ismail Abdullah, N. Kameran Al-Salihi, Forecasting tunnel geology, construction time and costs using machine learning methods **33**, 321 (2021)
62. H.K.S. Panahi, M. Dehhaghi, Y.S. Ok, A.S. Nizami, B. Khoshnevisan, S.I. Mussatto, M. Aghbashlo, M. Tabatabaei, S.S. Lam, A comprehensive review of engineered biochar: production, characteristics, and environmental applications. J. Clean. Prod. **270**, 122462 (2020)
63. L. Lu, X. Gao, J.-F. Dietiker, M. Shahnam, W.A. Rogers, Mfix based multi-scale cfd simulations of biomass fast pyrolysis: a review **248**, 117131 (2022)
64. A. Karaci, A. Caglar, B. Aydinli, S. Pekol, The pyrolysis process verification of hydrogen rich gas (h-rg) production by artificial neural network (ann). Int. J. Hydrog. EnergyHydrog. Energy **41**, 4570 (2016)
65. B. Aydinli, A. Caglar, S. Pekol, A. Karaci, The prediction of potential energy and matter production from biomass pyrolysis with artificial neural network. Energy Explor. Exploit.Explor. Exploit. **35**, 698 (2017)
66. T.H. Kim, M.K. Choi, H.S. Choi, Biomass fast pyrolysis prediction model through data-based prediction models coupling with cpfd simulation. J. Anal. Appl. Pyrol. **162**, 105448 (2022)
67. Y.F. Sun, L.N. Liu, Q. Wang, X.Y. Yang, X. Tu, Pyrolysis products from industrial waste biomass based on a neural network model. J. Anal. Appl. Pyrol. **120**, 94 (2016)
68. A. Pathy, S. Meher, P. Balasubramanian, Predicting algal biochar yield using extreme gradient boosting (xgb) algorithm of machine learning methods. Algal Res. **50**, 102006 (2020)
69. Z. Ullah, M. Khan, S.R.N Aqvi, W. Farooq, H. Yang, S. Wang, D.V.N Vo, A comparative study of machine learning methods for bio-oil yield prediction – A genetic algorithm-based features selection. Biores. Technol. **335**, 125292 (2021). https://doi.org/10.1016/j.biortech.2021.125292
70. C. Tsekos, S. Tandurella, W. de Jong, Estimation of lignocellulosic biomass pyrolysis product yields using artificial neural networks. J. Anal. Appl. Pyrol. **157**, 105180 (2021)
71. T.Y. Li, H. Xiang, Y. Yang, J. Wang, G. Yildiz, Prediction of char production from slow pyrolysis of lignocellulosic biomass using multiple nonlinear regression and artificial neural network. J Anal. Appl. Pyrol. **159**,105286 (2021). https://doi.org/10.1016/j.jaap.2021.105286
72. X. Zhu, Y. Li, X. Wang, Machine learning prediction of biochar yield and carbon contents in biochar based on biomass characteristics and pyrolysis conditions. Bioresour. Technol. **288**, 121527 (2019)
73. S. Zhao, W. Xu, L. Chen, The modeling and products prediction for biomass oxidative pyrolysis based on pso-ann method: an artificial intelligence algorithm approach **312**, 122966 (2022)
74. T. Onsree, N. Tippayawong, S. Phithakkitnukoon, J. Lauterbach, Interpretable machine-learning model with a collaborative game approach to predict yields and higher heating value of torrefied biomass. Energy **249**, 123676 (2022)
75. Z. Ullah, M. Khan, S.R. Naqvi, M.N.A. Khan, W. Farooq, M.W. Anjum, M.W. Yaqub, H. AlMohamadi, F. Almomani, An integrated framework of data-driven, metaheuristic, and mechanistic modeling approach for biomass pyrolysis. Process. Saf. Environ. **162**, 337 (2022)
76. J. Song, C. Tang, S. Yu, X. Yang, L. Yang, Prediction of product yields using fusion model from co-pyrolysis of biomass and coal. Bioresour. Technol. **353**, 127132 (2022)
77. Q. Tang, Y. Chen, H. Yang, M. Liu, H. Xiao, S. Wang, H. Chen, S. Raza Naqvi, Machine learning prediction of pyrolytic gas yield and compositions with feature reduction methods: effects of pyrolysis conditions and biomass characteristics. Bioresour. Technol. **339**, 125581 (2021)
78. H.K. Balsora, S. Kartik, V. Dua, J.B. Joshi, G. Kataria, A. Sharma, A.G. Chakinala, Machine learning approach for the prediction of biomass pyrolysis kinetics from preliminary analysis. J. Environ. Chem. Eng. **10**, 108025 (2022)
79. F. Kartal, Y. Dalbudak, U. Özveren, Prediction of thermal degradation of biopolymers in biomass under pyrolysis atmosphere by means of machine learning. Renew. Energy **204**, 774 (2023)
80. F. Cheng, M.D. Porter, L.M. Colosi, Is hydrothermal treatment coupled with carbon capture and storage an energy-producing negative emissions technology? **203**, 112252 (2020)
81. H.B. Zhong, Q.G. Xiong, L.N. Yin, J.T. Zhang, Y.Q. Zhu, S.R. Liang, B. Niu, X.Y. Zhang, Cfd-based reduced-order modeling of fluidized-bed biomass fast pyrolysis using artificial neural network. Renew. Energy **152**, 613 (2020)

82. S. Sasithorn, B. Chalermsinsuwan, P. Piumsomboon, Artificial neural network model for the prediction of kinetic parameters of biomass pyrolysis from its constituents. Fuel **193**, 142 (2017)
83. S.L. Wang, Z.Y. Shi, Y.H. Jin, I.N. Zaini, Y. Li, C.C. Tang, W.Z. Mu, Y.M. Wen, J.C. Jiang, P.G. Jonsson, A machine learning model to predict the pyrolytic kinetics of different types of feedstocks. Energy Convers. Manag. **260**, 115613 (2022)
84. A.Y. Mutlu, O. Yucel, An artificial intelligence based approach to predicting syngas composition for downdraft biomass gasification. Energy **165**, 895 (2018)
85. Ö. Çepelioğullar, İ. Mutlu, S. Yaman, H. Haykiri-Acma, Activation energy prediction of biomass wastes based on different neural network topologies **220**, 535 (2018)
86. J.K. Xing, K. Luo, H.O. Wang, Z.W. Gao, J.R. Fan, A comprehensive study on estimating higher heating value of biomass from proximate and ultimate analysis with machine learning approaches. Energy **188**, 116077 (2019)
87. N. Lerkkasemsan, Fuzzy logic-based predictive model for biomass pyrolysis. Appl. Energy **185**, 1019 (2017)
88. W.Q. Ji, F. Richter, M.J. Gollner, S.L. Deng, Autonomous kinetic modeling of biomass pyrolysis using chemical reaction neural networks. Combust. Flame **240**, 111992 (2022)
89. H. Wei, K. Luo, J. Xing, J. Fan, Predicting co-pyrolysis of coal and biomass using machine learning approaches. **310**, 122248 (2022)
90. L. Dong, R.R. Wang, P.D. Liu, S. Sarvazizi, Prediction of pyrolysis kinetics of biomass: new insights from artificial intelligence-based modeling. Int. J. Chem. Eng. **2022** (2022)
91. B.R. Hough, D.A.C. Beck, D.T. Schwartz, J. Pfaendtner, Application of machine learning to pyrolysis reaction networks: reducing model solution time to enable process optimization. Comput. Chem. Eng. **104**, 56 (2017)
92. A.P. Mayol, J.M.Z. Maningo, A.G.A.Y. Chua-Unsu, C.B. Felix, P.I. Rico, G.S. Chua, E.V. Manalili, D.D. Fernandez, J.L. Cuello, A.A. Bandala, Application of artificial neural networks in prediction of pyrolysis behavior for algal mat (lablab) biomass. In *2018 IEEE 10th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management (HNICEM), IEEE*, pp. 1–5 (2018)
93. M. Puig-Arnavat, E.A. Tora, J.C. Bruno, A. Coronas, State of the art on reactor designs for solar gasification of carbonaceous feedstock. Sol. Energy **97**, 67 (2013)
94. R. Mikulandrić, D. Lončar, D. Böhning, R. Böhme, M. Beckmann, Artificial neural network modelling approach for a biomass gasification process in fixed bed gasifiers **87**, 1210 (2014)
95. D. Baruah, D.C. Baruah, M.K. Hazarika, Artificial neural network based modeling of biomass gasification in fixed bed downdraft gasifiers. Biomass Bioenergy **98**, 264 (2017)
96. J. George, P. Arun, C. Muraleedharan, Assessment of producer gas composition in air gasification of biomass using artificial neural network model. Int. J. Hydrog. EnergyHydrog. Energy **43**, 9558 (2018)
97. F. Elmaz, O. Yücel, A.Y. Mutlu, Predictive modeling of biomass gasification with machine learning-based regression methods. Energy **191**, 116541 (2020)
98. E.E. Ozbas, D. Aksu, A. Ongen, M.A. Aydin, H.K. Ozcan, Hydrogen production via biomass gasification, and modeling by supervised machine learning algorithms. Int. J. Hydrog. EnergyHydrog. Energy **44**, 17260 (2019)
99. Y. Fang, L. Ma, Z.Y. Yao, W.L. Li, S.M. You, Process optimization of biomass gasification with a monte carlo approach and random forest algorithm. Energy Convers. Manag. **264**, 115734 (2022)
100. J. Li, L.Y. Li, Y.W. Tong, X.A. Wang, Understanding and optimizing the gasification of biomass waste with machine learning. Green Chem. Eng. **4**, 123 (2023)
101. D. Serrano, D. Castelló, Tar prediction in bubbling fluidized bed gasification through artificial neural networks. Chem. Eng. J. **402**, 126229 (2020)
102. D. Pandey, V.C. Pandey, Sacred plants from ancient to modern era: traditional worshipping towards plants conservation **3**, 136 (2016)
103. A.K. Stark, R.B. Bates, Z.L. Zhao, A.F. Ghoniem, Prediction and validation of major gas and tar species from a reactor network model of air-blown fluidized bed biomass gasification. Energy Fuel **29**, 2437 (2015)
104. R. Sridhar, Chemistry models for major gas species estimation and tar prediction in fluidized bed biomass gasification. Massachusetts Institute of Technology (2016)
105. B. Wojnicka, M. Ściążko, J.C. Schmid, Modelling of biomass gasification with steam **11**, 1787 (2021)
106. B. Hejazi, J.R. Grace, X.T. Bi, A. Mahecha-Botero, Kinetic model of steam gasification of biomass in a bubbling fluidized bed reactor. Energy Fuel **31**, 1702 (2017)
107. R. Rameshkumar, K. Mayilsamy, Prediction of tar and particulate in biomass gasification using adaptive neuro fuzzy inference system. J. Intell. Fuzzy Syst. **27**, 361 (2014)
108. K. Kang, R. Azargohar, A.K. Dalai, H. Wang, Hydrogen generation via supercritical water gasification of lignin using ni-co/mg-al catalysts. Int. J. Energy Res. **41**, 1835 (2017)
109. J.A. Okolie, E.I. Epelle, S. Nanda, D. Castello, A.K. Dalai, J.A. Kozinski, Modeling and process optimization of hydrothermal gasification for hydrogen production: a comprehensive review. J. Supercrit. Fluid **173**, 105199 (2021)
110. N. Kardani, A.N. Zhou, M. Nazem, X.S. Lin, Modelling of municipal solid waste gasification using an optimised ensemble soft computing model. Fuel **289**, 119903 (2021)
111. B.V. Ayodele, S.I. Mustapa, R. Kanthasamy, M. Zwawi, C.K. Cheng, Modeling the prediction of hydrogen production by co-gasification of plastic and rubber wastes using machine learning algorithms. Int. J. Energy Res. **45**, 9580 (2021)
112. O. Yucel, E.S. Aydin, H. Sadikoglu, Comparison of the different artificial neural networks in prediction of biomass gasification products. Int. J. Energy Res. **43**, 5992 (2019)
113. S. Zhao, J. Li, C. Chen, B.B. Yan, J.Y. Tao, G.Y. Chen, Interpretable machine learning for predicting and evaluating hydrogen production via supercritical water gasification of biomass. J. Clean. Prod. **316**, 128244 (2021)
114. S. Shenbagaraj, P.K. Sharma, A.K. Sharma, G. Raghav, K.B. Kota, V. Ashokkumar, Gasification of food waste in supercritical water: an innovative synthesis gas composition prediction model based on artificial neural networks. Int. J. Hydrog. EnergyHydrog. Energy **46**, 12739 (2021)
115. W. Zhang, J. Li, T. Liu, S. Leng, L. Yang, H. Peng, S. Jiang, W. Zhou, L. Leng, H. Li, Machine learning prediction and optimization of bio-oil production from hydrothermal liquefaction of algae. Bioresour. Technol. **342**, 126011 (2021)
116. J. Li, M. Suvarna, L.J. Pan, Y.R. Zhao, X.N. Wang, A hybrid data-driven and mechanistic modelling approach for hydrothermal gasification. Appl. Energy **304**, 117674 (2021)
117. P.V. Gopirajan, K.P. Gopinath, G. Sivaranjani, J. Arun, Optimization of hydrothermal gasification process through machine learning approach: experimental conditions, product yield and pollution. J. Clean. Prod. **306**, 127302 (2021)
118. D. Fózer, A.J. Tóth, P.S. Varbanov, J.J. Klemeš, P. Mizsey, Sustainability assessment of biomethanol production via hydrothermal gasification supported by artificial neural network **318**, 128606 (2021)
119. A. Bahadar, R. Kanthasamy, H.H. Sait, M. Zwawi, M. Algarni, B.V. Ayodele, C.K. Cheng, L.J. Wei, Elucidating the effect of process parameters on the production of hydrogen-rich syngas

by biomass and coal co-gasification techniques: a multi-criteria modeling approach. Chemosphere **287**, 132052 (2022)
120. J. Li, L. Li, M. Suvarna, L. Pan, M. Tabatabaei, Y.S. Ok, X. Wang, Wet wastes to bioenergy and biochar: a critical review with future perspectives. Sci. Total. Environ. **817**, 152921 (2022)
121. I.O. Vardiambasis, T.N. Kapetanakis, C.D. Nikolopoulos, T.K. Trang, T. Tsubota, R. Keyikoglu, A. Khataee, D. Kalderis, Hydrochars as emerging biofuels: recent advances and application of artificial neural networks for the prediction of heating values. Energies **13**, 4572 (2020)
122. H.Y. Ismail, S. Shirazian, I. Skoretska, O. Mynko, B. Ghanim, J.J. Leahy, G.M. Walker, W. Kwapinski, Ann-kriging hybrid model for predicting carbon and inorganic phosphorus recovery in hydrothermal carbonization. Waste Manag. **85**, 242 (2019)
123. T.N. Kapetanakis, I.O. Vardiambasis, C.D. Nikolopoulos, A.I. Konstantaras, T.K. Trang, D.A. Khuong, T. Tsubota, R. Keyikoglu, A. Khataee, D. Kalderis, Towards engineered hydrochars: application of artificial neural networks in the hydrothermal carbonization of sewage sludge. Energies **14**, 3000 (2021)
124. L. Mu, Z. Wang, D. Wu, L. Zhao, H.C. Yin, Prediction and evaluation of fuel properties of hydrochar from waste solid biomass: machine learning algorithm based on proposed pso-nn model. Fuel **318**, 123644 (2022)
125. M. Aghaaminiha, R. Mehrani, T. Reza, S. Sharma, Comparison of machine learning methodologies for predicting kinetics of hydrothermal carbonization of selective biomass. Biomass Convers. Bioresour. **13**, 9855 (2023)
126. O.S. Djandja, A.A. Salami, Z.C. Wang, J. Duo, L.X. Yin, P.G. Duan, Random forest-based modeling for insights on phosphorus content in hydrochar produced from hydrothermal carbonization of sewage sludge. Energy **245**, 123295 (2022)
127. O.S. Djandja, P.-G. Duan, L.-X. Yin, Z.-C. Wang, J. Duo, A novel machine learning-based approach for prediction of nitrogen content in hydrochar from hydrothermal carbonization of sewage sludge **232**, 121010 (2021)
128. J. Li, X.Z. Zhu, Y.A. Li, Y.W. Tong, Y.S. Ok, X.N. Wang, 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 (2021)
129. W. Jiang, X.J. Xing, S. Li, X.W. Zhang, W.Q. Wang, Synthesis, characterization and machine learning based performance prediction of straw activated carbon. J. Clean. Prod. **212**, 1210 (2019)
130. M. Meena, S. Shubham, K. Paritosh, N. Pareek, V. Vivekanand, Production of biofuels from biomass: predicting the energy employing artificial intelligence modelling. Bioresour. Technol. **340**, 125642 (2021)
131. F. Cheng, E.R. Belden, W.J. Li, M. Shahabuddin, R.C. Paffenroth, M.T. Timko, Accuracy of predictions made by machine learned models for biocrude yields obtained from hydrothermal liquefaction of organic wastes. Chem. Eng. J. **442**, 136013 (2022)
132. T. Katongtung, T. Onsree, N. Tippayawong, Machine learning prediction of biocrude yields and higher heating values from hydrothermal liquefaction of wet biomass and wastes. Bioresour. Technol. **344**, 126278 (2022)
133. P.V. Gopirajan, K.P. Gopinath, G. Sivaranjani, J. Arun, Optimization of hydrothermal liquefaction process through machine learning approach: process conditions and oil yield. Biomass Convers. Bioresour. **13**, 1213 (2023)
134. A. Shafizadeh, H. Shahbeig, M.H. Nadian, H. Mobli, M. Dowlati, V.K. Gupta, W.X. Peng, S.S. Lam, M. Tabatabaei, M. Aghbashlo, Machine learning predicts and optimizes hydrothermal liquefaction of biomass. Chem. Eng. J. **445**, 136579 (2022)
135. L. Leng, W. Zhang, T. Liu, H. Zhan, J. Li, L. Yang, J. Li, H. Peng, H. Li, Machine learning predicting wastewater properties of the aqueous phase derived from hydrothermal treatment of biomass. Bioresour. Technol. **358**, 127348 (2022)
136. S.A. Kalogirou, Artificial intelligence for the modeling and control of combustion processes: a review. Prog. Energy Combust. **29**, 515 (2003)
137. S. Sezer, F. Kartal, U. Özveren, Prediction of combustion reactivity for lignocellulosic fuels by means of machine learning. J. Therm. Anal. Calorim. **147**, 9793 (2022)
138. M. Karimi, M. Khosravi, R. Fathollahi, A. Khandakar, B. Vaferi, Determination of the heat capacity of cellulosic biosamples employing diverse machine learning approaches. Energy Sci. Eng. **10**, 1925 (2022)
139. P. Tóth, A. Garami, B. Csordás, Image-based deep neural network prediction of the heat output of a step-grate biomass boiler. Appl. Energy **200**, 155 (2017)
140. H.B. Bi, C.X. Wang, Q.Z. Lin, X.D. Jiang, C.L. Jiang, L. Bao, Combustion behavior, kinetics, gas emission characteristics and artificial neural network modeling of coal gangue and biomass via tg-ftir. Energy **213**, 118790 (2020)
141. W. Jiang, X.J. Xing, X.W. Zhang, M.X. Mi, Prediction of combustion activation energy of naoh/koh catalyzed straw pyrolytic carbon based on machine learning. Renew. Energy **130**, 1216 (2019)
142. F. Kartal, U. Özveren, Prediction of activation energy for combustion and pyrolysis by means of machine learning. Therm. Sci. Eng. Prog. **33**, 101346 (2022)
143. V. Sharma, A.P.S. Chouhan, D. Bisen, Prediction of activation energy of biomass wastes by using multilayer perceptron neural network with weka. Mater. Today Proc. **57**, 1944 (2022)
144. S. Sunphorka, B. Chalermsinsuwan, P. Piumsomboon, Application of artificial neural network for kinetic parameters prediction of biomass oxidation from biomass properties. J. Energy Inst. **90**, 51 (2017)
145. M. Buyukada, Co-combustion of peanut hull and coal blends: artificial neural networks modeling, particle swarm optimization and monte carlo simulation. Bioresour. Technol. **216**, 280 (2016)
146. Z. Ni, H. Bi, C. Jiang, H. Sun, W. Zhou, J. Tian, Q. Lin, Investigation of co-combustion of sewage sludge and coffee industry residue by tg-ftir and machine learning methods **309**, 122082 (2022)
147. S.T. Wen, M. Buyukada, F. Evrendilek, J.Y. Liu, Uncertainty and sensitivity analyses of co-combustion/pyrolysis of textile dyeing sludge and incense sticks: Regression and machine-learning models. Renew. Energy **151**, 463 (2020)
148. J.O. Rico-Contreras, A.A. Aguilar-Lasserre, J.M. Mendez-Contreras, J.J. Lopez-Andres, G. Cid-Chama, Moisture content prediction in poultry litter using artificial intelligence techniques and Monte Carlo simulation to determine the economic yield from energy use. J. Environ. Manag. **202**, 254 (2017)
149. N. Li, G. Lu, X. Li, Y. Yan, Prediction of nox emissions from a biomass fired combustion process through digital imaging, non-negative matrix factorization and fast sparse regression. In: *2015 IEEE International Instrumentation and Measurement Technology Conference (I2MTC) Proceedings*, pp. 176–180 (2015)
150. N. Li, G. Lu, X.L. Li, Y. Yan, Prediction of nox emissions from a biomass fired combustion process based on flame radical imaging and deep learning techniques. Combust. Sci. Technol. **188**, 233 (2016)
151. L. Böhler, M. Fallmann, G. Görtler, J. Krail, F. Schittl, M. Kozek, Emission limited model predictive control of a small-scale biomass furnace **285**, 116414 (2021)
152. L. Böhler, G. Görtler, J. Krail, M. Kozek, Carbon monoxide emission models for small-scale biomass combustion of wooden pellets **254**, 113668 (2019)

153. L. Böhler, J. Krail, G. Görtler, M. Kozek, Fuzzy model predictive control for small-scale biomass combustion furnaces **276**, 115339 (2020)
154. A.S. Noushabadi, A. Dashti, F. Ahmadijokani, J.G. Hu, A.H. Mohammadi, Estimation of higher heating values (hhvs) of biomass fuels based on ultimate analysis using machine learning techniques and improved equation. Renew. Energy **179**, 550 (2021)
155. H. You, Z. Ma, Y. Tang, Y. Wang, J. Yan, M. Ni, K. Cen, Q. Huang, Comparison of ann (mlp), anfis, svm, and rf models for the online classification of heating value of burning municipal solid waste in circulating fluidized bed incinerators. Waste Manag. **68**, 186 (2017)
156. X. Chen, Y. Zhang, B. Xu, Y. Li, A simple model for estimation of higher heating value of oily sludge **239**, 121921 (2022)
157. J.O. Ighalo, C.A. Igwegbe, A.G. Adeniyi, Multi-layer perceptron artificial neural network (mlp-ann) prediction of biomass higher heating value (hhv) using combined biomass proximate and ultimate analysis data **8**, 3177 (2022)
158. P. Khunphakdee, K. Korkerd, C. Soanuch, B. Chalermsinsuwan, Data-driven correlations of higher heating value for biomass, waste and their combination based on their elemental compositions. Energy Rep. **8**, 36 (2022)
159. S.H. Samadi, B. Ghobadian, M. Nosrati, Prediction of higher heating value of biomass materials based on proximate analysis using gradient boosted regression trees method **43**, 672 (2021)
160. A. Ghaderi, A.A. Shahri, S. Larsson, An artificial neural network based model to predict spatial soil type distribution using piezocone penetration test data (cptu). Bull. Eng. Geol. Environ. **78**, 4579 (2019)
161. C.L. Williams, T.L. Westover, R.M. Emerson, J.S. Tumuluru, C.L. Li, Sources of biomass feedstock variability and the potential impact on biofuels production. Bioenerg. Res. **9**, 1 (2016)
162. D.H. Wolpert, W.G. Macready, No free lunch theorems for search, technical report SFI-TR-95-02-010, Santa Fe Institute (1995)
163. M.V. Shcherbakov, A. Brebels, N.L. Shcherbakova, A.P. Tyukov, T.A. Janovsky, V.A.E. Kamaev, A survey of forecast error measures. **24**, 171 (2013)