Machine learning and circular bioeconomy

Building new resource efficiency from diverse waste streams

Tsui, To Hung; van Loosdrecht, Mark C.M.; Dai, Yanjun; Tong, Yen Wah

DOI
10.1016/j.biortech.2022.128445

Publication date
2023

Document Version
Final published version

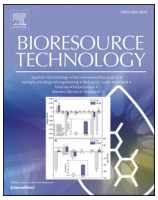
Published in
Bioresource Technology

Bioresource Technology 369 (2023) 128445

Contents lists available at ScienceDirect

Bioresource Technology

journal homepage: www.elsevier.com/locate/biortech



Machine learning and circular bioeconomy: Building new resource efficiency from diverse waste streams

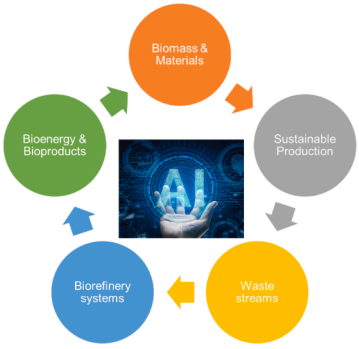
To-Hung Tsui [a,b], Mark C.M. van Loosdrecht [c], Yanjun Dai [d], Yen Wah Tong [a,b,e,*]

[a] Environmental Research Institute, National University of Singapore, 1 Create Way, 138602, Singapore
[b] Energy and Environmental Sustainability for Megacities (E2S2) Phase II, Campus for Research Excellence and Technological Enterprise (CREATE), 1 Create Way, Singapore, 138602, Singapore
[c] Department of Biotechnology, Delft University of Technology, the Netherlands
[d] School of Mechanical Engineering, Shanghai Jiaotong University, 800 Dongchuan Road, Shanghai, 200240, China
[e] Department of Chemical and Biomolecular Engineering, National University of Singapore, 4 Engineering Drive 4, 117585, Singapore

HIGHLIGHTS

- Machine learning applications for four critical biorefinery systems are analyzed.
- Characteristics and limitations of common algorithms are summarized.
- Advancements against modeling techniques of mechanistic approaches are studied.
- Collective efforts for next-stage machine learning applications are highlighted.

GRAPHICAL ABSTRACT



ARTICLE INFO

Keywords:
Biorefinery
Multiscale modeling
Resource recovery
Supply chain
Sustainability

ABSTRACT

Biorefinery systems are playing pivotal roles in the technological support of resource efficiency for circular bioeconomy. Meanwhile, artificial intelligence presents great potential in handling scientific tasks of high-dimensional complexity. This review article scrutinizes the status of machine learning (ML) applications in four critical biorefinery systems (i.e. composting, fermentation, anaerobic digestion, and thermochemical conversions) as well as their advancements against traditional modeling techniques of mechanistic approach. The contents cover their algorithm selections, modeling challenges, and prospective improvements. Perspectives are sketched to further inform collective efforts on crucial aspects. The multidisciplinary interchange of modeling knowledge will enable a more progressive digital transformation of sustainability efforts in supporting sustainable development goals.

* Corresponding author at: Department of Chemical and Biomolecular Engineering, National University of Singapore, 4 Engineering Drive 4, Singapore 117585, Singapore.
E-mail address: chetyw@nus.edu.sg (Y.W. Tong).

https://doi.org/10.1016/j.biortech.2022.128445
Received 3 October 2022; Received in revised form 29 November 2022; Accepted 2 December 2022
Available online 5 December 2022

1. Introduction

The significance of a circular bioeconomy has been put forward by the European Union to support climate actions as a part of the Paris agreement (Stegmann et al., 2020). By its definition, it aims to divert the disposal of waste streams and re-create their social-economic values, through upcycling their circularity within bioresource systems. Biofuels and bioproducts recovery from waste streams present a great potential to provide sustainability pursuits for energy and resources at a regional scale (Tsui & Wong, 2019). The resource recovery can be achieved via biological conversion methods (including fermentation, anaerobic digestion, and composting) in mild environments or via much faster thermochemical processes. The emerging policy agenda across countries are closely relevant to many researchers and decision-makers in exploring new strategies through cascading use of biomass, contributing to achieving the United Nations' sustainable development goals. Despite this, there is a wide range of upcoming challenges that would determine the degree of success. Particularly, integrated biorefinery systems involve high-dimensional complexity that varies across different spatial and temporal circumstances, which adds complications to technological and policy formulations.

Artificial intelligence technologies are playing increasing roles in engineering design and scientific research. Traditionally, approaches of mechanistic modeling have been mathematically developed to simulate the fixed assumptions of complex physical/ chemical / biological interaction within biorefinery systems, as well as computational fluid dynamics and heat/mass transfer inside their reactor designs (Tsui et al., 2016; Clauser et al., 2021). The existing difficulties/ inefficiencies to model complex systems are making mechanistic assumptions, determining mathematical structures of various processes, and calibrating a large number of model parameters. Thus, when it comes to the details of implementation, the complication of diverse feedstocks, targeted biofuels/ bioproducts, and pathways of biorefinery technologies have been more often managed by experimental methods, though they are known to be time-consuming, expensive, and manpower-intensive (Li et al., 2021; Cruz et al., 2022). With the recent advance in Machine Learning (ML) algorithms, it provides researchers with more accessible opportunities of developing new system efficiency in a solution approach that can directly build on available data.

Recently, there has been a surge in ML applications for biorefinery systems. They are happening against the backdrop of digital transformations required in each industry and academic discipline to address rising sustainability challenges of resource efficiency. Without a doubt, ML can expand the task scale which potentially provides new progress in modeling techniques of system prediction/ optimization. It can be asked whether this is finally the moment or if much more effort is still missing for such a regime shift in circular bioeconomy. Therefore, this review article aims to provide an updated examination of the fast-growing roles of ML and existing mechanistic modeling techniques on the four critical biorefinery systems (i.e. anaerobic digestion, composting, fermentation, and thermochemical conversion). The overall computational concepts and algorithms of ML are briefly summarized. The core review contents focus on the multidisciplinary interchange of modeling knowledge under the wide range of scientific tasks advanced by ML, where their targeted biofuels/bioproducts recovery, input parameters, and optimized predictive algorithms are summarized. Based on the review findings, perspectives are sketched to convey research needs and future direction.

2. An overview of machine learning

Every ML application begins with a precise scientific/ technological task. Once the scope of ML tasks is defined, data collection can be proceeded, followed by sequential computational steps (Fig. 1). Data preparation is crucial in the high-quality development of predictive algorithms. Particularly, biorefineries applying microbiological methods that consider systems biology can have a large number of features, whereas microarray datasets can contain up to sixty thousand features with several hundred samples. The data features could be prone to overfitting issues. Before ML computation, the process of feature selection (manual or automatic) allows for eliminating irrelevant features and retaining subsets of the most influential features for model training.

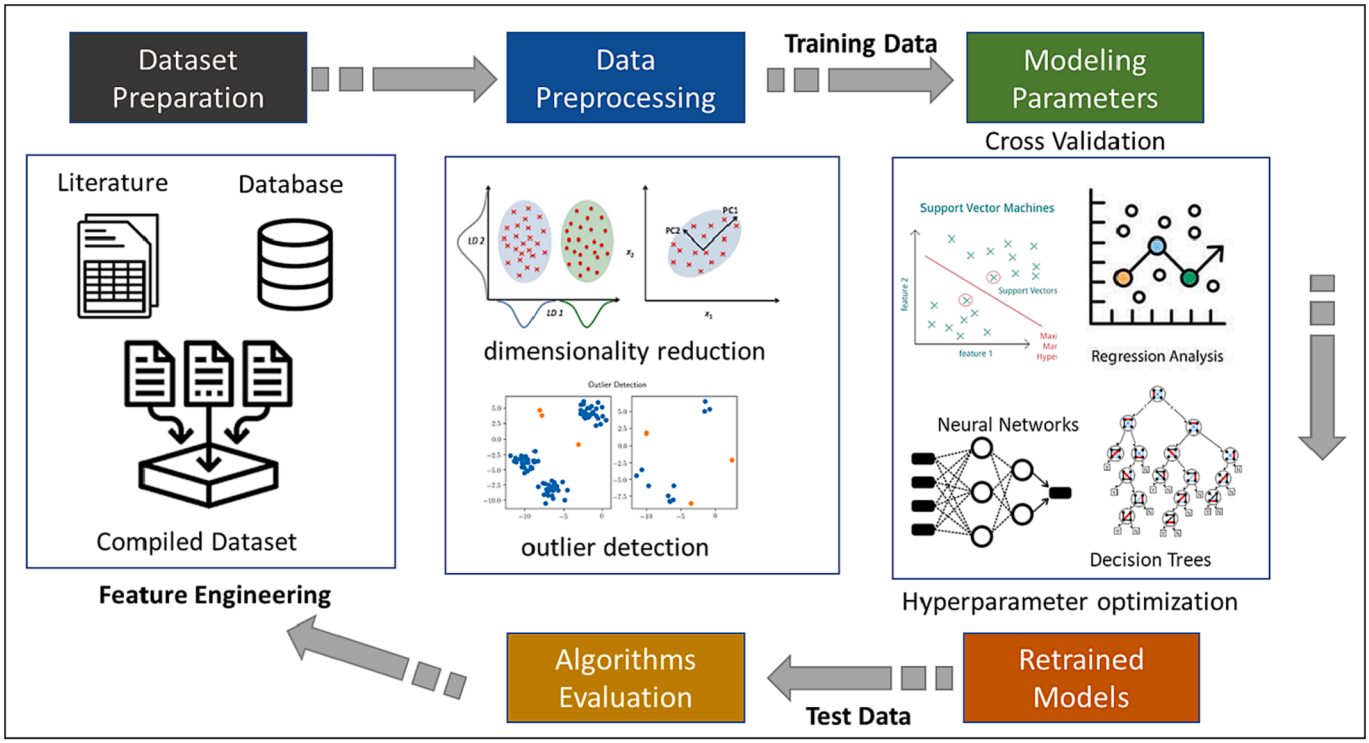


Fig. 1. Summary of computational steps.

Feature engineering would cost more analytical efforts, but it can often improve the training data with better feature representation. The overall ML computational steps involve cycles of (i) model training – analysis of data patterns by algorithms; (ii) model validation – hyperparameters tuning to optimize the predictive performance; and (iii) model assessment – testing by used and unused datasets (Fig. 1). There are different kinds of assessment metrics (such as confusion matrix and Gini coefficient), which depends on the type of ML model and the implementation plan. The brief introduction of common algorithms below provides a quick entry for readers with less ML background. As further illustrated in sections 3 and 4, algorithm selection could be challenging in the current research landscape of resource recovery, and it is determined by a wide range of factors (e.g. algorithm principle for data, modeling objective, computational constraints).

2.1. Regression methods

Regression analyses (e.g. the simplest form as the least squares method) have existed for many years to examine correlations between target outputs and independent variables. Their general limitations are the simplicity of inputs/outputs, so there could be a severe impact of outliers on performance. Common types of regression techniques include polynomial/logistic regression and their processing methods. For data not following linear relationships, polynomial methods can be applied for correlation by polynomial equations of the nth degree. Logistic methods can be used for discrete problems, and shrinkage methods (e.g. ridge regression) can be further used to handle data suffering from multicollinearity. Kernel smoothing is a non-parametric technique, whereas the computation process applies linear classifiers to non-linear needs by mapping non-linear data into a higher-dimensional space (Hastie et al., 2009). Compared to other ML algorithms, the overall advantages of regression methods are less complex to perform and with fast computational speed.

2.2. Artificial neural networks

Artificial neural networks (ANNs) are based on the biomimicry of how the human brain processes information and passes signals to other neurons in the next layer. During the model development, the neuron network (of hidden and output layers) learns data patterns by accordingly adjusting the weights of each neuron and feeding back its predictions. For solving operations on the weighted sum of the neurons, activation functions (e.g. softmax function, sigmoid function, hyperbolic tangent function) have been developed to improve the performance (Karlik and Olgac, 2011). ANNs are appealing to model non-linear relationships, and the advantages of high prediction accuracy are often observed even in high-dimensional inputs. Some examples of deep-learning architectures include recurrent neural networks, long short-term memory, convolutional neural network, spatiotemporal graph, and convolutional networks. The frameworks provide an efficient computational approach to handling time-series applications on biorefinery systems. However, their drawbacks are often described by their black-box nature of poor interpretability. Regarding this, model-agnostic methods (e.g. feature permutation and global surrogate models) can provide a more accessible way to interpret the ML computation. Since the development of accurate neural networks requires a large data size, the computational costs are generally higher.

2.3. Decision tree methods

Decision tree methods are known for handling both classification and regression problems. Based on the hierarchical if/else questions, the tree-structure computation classifies the data inputs and derives decisions from branches. Random forest (RF) and gradient boosting are two common ensemble methods. In terms of their fundamental difference, gradient boosting makes use of weak learners to improve prediction as a sequential method, whereas RF averages the results of decision trees to overcome each limitation. In terms of hyperparameter tuning, gradient boosting models tend to outperform an RF, but RF has fewer hyperparameters and can do satisfactory predictions even under sub-optimal tuning. Compared to other ML algorithms, tree-based models can be trained rapidly with lower computational costs, but they are also prone to overfitting. Pruning is a strategy to relieve the overfitting issue by removing excess branches. Through the introduction of a cost-complexity criterion, the general principle of pruning is based on the compensation between the goodness of fit and generalization capability (Schmidhuber, 2015). Due to the hierarchical nature, a high variance could be observed, when the training data is subjected to small changes. Another general characteristic of decision tree methods is their average performance on regression tasks.

2.4. Support vector machines

Support vector machines (SVMs) can handle regression and classification problems, and they are known for their capability for binary classification tasks. In principle, SVMs map training examples in space thereby distancing the width between the categories, and support vectors of data points result in the hyperplane construction. For solving non-linear problems, data variables can be mapped into a high-dimension space by kernel functions (including radial basis functions and nth-degree polynomials). The computational tasks include pattern analysis of general types (e.g. clusters, classifications, and correlations). Least-squares SVMs are a well-known class of kernel-based methods. For many real-world applications across fields, SVMs have performed well in prediction (Somvanshi et al., 2016). For high-dimensional datasets of a limited size, they were also shown to yield good predictions, compared to other ML algorithms. However, SVMs are less ideal for large sample volumes and indirect probability estimation during cross-validation. Attributed by the way of support vectors in constructing hyperplanes, SVMs are generally shown to be less sensitive to outliers.

3. Optimization of resource efficiency and modeling techniques

Biorefinery experiments often need long-term trials, and the methodological procedures also require sophisticated facilities. To illustrate the status of ML applications on biorefinery systems, this section summarizes the specific examples of the latest research and how ML can optimize efficiencies in exploring resource recovery from waste streams. Each sub-section is started with a summary of these biorefinery processes and mechanistic modeling, followed by a discussion of their ML methods aiding the application purposes (e.g. predictions, optimization, or other scientific tasks).

3.1. Composting

Aerobic composting is an effective approach to recycling agricultural wastes into nutrient-rich biofertilizers. Over the years, experimental researchers have been studying the factors (e.g. inoculum sources, C/N ratio, air supply, and pH) controlling compost maturity and other mechanistic phenomena (e.g. greenhouse gas emission, nutrients loss) (Ren et al., 2022; Sun et al., 2020). Composting systems can be operated in open environments (e.g. through agitated windrow patterns and static solid beds) or in-vessel bioreactors (of fixed, agitated, or rotating forms). During the temperature-rising phase, microbial activities emit a large quantity of heat themselves. The thermophilic environments (up to 70–90 °C) can contribute to a greater rate and extent of organic degradation. In terms of resource conservation, the high-temperature conditions would lead to nutrient loss. In-depth modeling is important to understanding feedstock inputs (and other processing parameters) for energy transfer and nutrient conversion. By doing so, organics transformation can be optimized by in-vessel systems (through temperature control, duration of microbial phases, features of bulking agent, turning

frequency, etc.).

Quantitative models can assist in understanding how system features (i.e. feedstock characteristics and control measures) affect composting of different priorities (e.g. length of the composting regime, compost characteristics, pollution control). For example, the optimization principle of aeration techniques can build on the simulation of oxygen consumption rates and patterns by microbial activities. For in-vessel composting of multiple substrates, six models were developed for individually simulating the rate of change in oxygen, moisture, carbon dioxide, temperature, microbial biomass, and feedstock over the time scale. Mathematical models (mainly built on empirical equations, Monod-type equations, and first-order kinetics) can be used to describe the intermediate conversion patterns of biomass. Monod's growth kinetics simulate biological-mixture decomposition and it could fit the kinetics requirements of major degradation. However, the application of Monod-type equations is less popular and it is often constrained by the experimental practicability of excessive variables. In contrast, the first-order kinetic equation is the most common application for composting simulation, although the method was less suitable to model organics conversion at constant temperature parameters. The detailed evaluation of biofertilizers currently mainly relies on plant growth testing and genomic sequencing. At the current stage, the modeling techniques that consider end-product compost regarding its safety, quality, and reuse performance were still not put into much attention.

Currently, the research of ML on aerobic composting is still in the early stage. Table 1 summarizes the recent ML applications on aerobic composting, and the scope of scientific tasks can be divided into three aspects. Firstly, the agricultural interests of bioproduct qualities (including predictive variables of compost maturity, carbon dioxide emission, humification degree, and nutrient contents) (Ding et al., 2022; Li et al., 2022a; Hosseinzadeh et al., 2020). RF and ANN are commonly observed to have the best prediction performance, and the accuracy of R^2 was usually > 0.9. Secondly, the monitoring of contaminant residues such as pesticides, hormones, and antimicrobials (Kyakuwaire et al., 2019). The transfer of antimicrobial-resistant bacteria/ genes from animal agriculture is a widespread consequence of antimicrobial use. Thus, the recycling process through composting has raised safety concerns in a circular bioeconomy. To monitor their effective degradation, researchers have applied ML to study the tetracyclines reduction from chicken manure (Alavi et al., 2019). However, the recent progress of ML applications is far from the microbiological details to support the examination down to the gene level as well as further mitigation monitoring. For other contaminants such as bio-plastic degradation, Yamawaki et al. (2021) applied RF and XGBoost (extreme gradient boosting) to simulate the weight-loss ratio. However, the prediction accuracy was up to 0.38. Given that the analysis was based on a limited size of experimental data and there are no other similar applications (also based on 13C-CP/MAS spectra and initial crystallinity) for further comparisons, it cannot generalize implications at this review stage on whether existing ML algorithms have limitations to handle the prediction of contamination and their correlated impacts. Thirdly, the optimization needs of system design. The interpretive and optimization methods of ML can be applied to analyze conversion patterns in composting. Soto-Paz et al. (2020) used the particle swarm optimization method together with ANN (accuracy of 0.97) to assist in the co-composting design of sugarcane filter cakes. Moncks et al. (2022) applied ML to develop a sensor adjustment method for the automatic monitoring of moisture content during composting. The accuracy (R^2: 0.9939) was further verified by gravimetric analysis and the prediction obtained by the sensor node.

3.2. Fermentation

Fermentation biorefineries can convert biomass into chemical building blocks and biopolymers (Zhang et al., 2021). Common building blocks include fatty acids (such as lactic, itaconic, butyric, and succinic acids) and alcohols (such as ethanol, 1,3-Propanediol, 2,3-Butanediol, glycerol, etc). Some popular examples of biopolymers from waste streams include alginate, polyhydroxyalkanoate, cellulose, glycogen, cyanophycin, xanthan, and dextran. The biopolymer types from the microbial processes can be summarized into four major categories (i.e. polyamides, polysaccharides, polyesters, and polyanhydrides). Common considerations for system designs include the availability of engineered or wild-type microbes, control methods of their metabolic pathways as well as nutrient requirements for fermentation. The biosynthesis efficiency can also be controlled by networks of regulatory pathways subjected to external stimuli, and the microbiological responses are a type of cell protection mechanism through carbon /energy storage. An efficient simulation of the fermentation environments (such as the design of extracellular conditions and nutrient configurations) is crucial for screening/ prediction of technological options to inform economic feasibility.

To analyze and optimize genome-scale pathways, mechanistic approaches build on modeling metabolic network reconstruction to study

Table 1
Key findings from the literature on composting.

Biomass	Scientific Tasks	Major parameters	Algorithms	Best accuracy	Reference
Kitchen waste	Prediction of composting maturity	pH, C/N ratio, TOC, GI, EC, OM, nitrogen fractions, composting time, temperature, moisture	KNN, LR, DT, SVR, RF	RF; R^2:0.975	Ding et al. (2022)
Compost	Sensor adjustment for moisture monitoring	Sample temperature, air temperature, humidity, sample moisture, days	LR, MLP, IBK	IBK; R^2: 0.9075	Moncks et al. (2022)
Green waste	Prediction of carbon dioxide emission	CO2, TOC, TN, C/N ratio, cellulose, hemicellulose, lignin	AdaBoost, Bagging, GB, RF, kNN, DT	RF; R^2:0.88	Li et al. (2022a)
Sugarcane filter cake	Optimized design of co-composting	Mixing ratio, Turning frequency, Temperature, pH, O2, TOC, TN, TP, Respirometric index	ANN-PSO	ANN; R^2: 0.97	Soto-Paz et al. (2020)
Cardboard, Boxwood leaves, Sawdust	Prediction of total nitrogen & phosphorus conservation	TN, pH, EC, C/N, NH4/NO3, TP, water soluble carbon, DEH enzyme	LR, ANN	ANN; R^2: 0.999	Hosseinzadeh et al. (2020)
Chicken manure, Bagasse	Prediction of tetracyclines reduction	Time, Tetracycline hydrochloride, oxytetracycline hydrochloride, chlortetracycline hydrochloride, bagasse fraction	ANN	ANN; R^2: 0.99	Alavi et al. (2019)
Bioplastic, Compost	Prediction of weight-loss ratio in bioplastic degradation	moisture content, degradation period, initial crystallinity, and features of 13C-CP/MAS spectra	RF, XGBoost	RF: R^2: 0.38	Yamawaki et al. (2021)
Chicken manure, Cow dung, Sewage sludge, Garden waste, Rice straw	Prediction of humification degree	categories at genus level, humic acids, fulvic acids, OM, TOC	RF, LR	RF; R^2:0.9694	Yang et al. (2022)

Remarks: Instance Based Learner (IBK), Particle Swarm Optimization(PSO), Decision Tree (DT), LR (Linear regression), ML(Multilayer perceptron), GB(Gradient Boost), AdaBoost (Adaptive Boosting).

genetic information (such as associated phenotype of gene-protein-reaction) within the system biology (Bordbar et al., 2014; Zhang et al., 2022a, Zhang et al., 2022b). Flux balance analysis and metabolic flux analysis are two common methods to study the reactions within the metabolic network. For example, flux balance analysis has also been used to understand of modulation of carbon flux distribution for the biosynthesis of ethanol and butanol (Kaushal et al., 2018). Metabolic flux analysis was applied to study pH influences on intracellular activities of β-lactamase producing *Bacillus licheniformis* (Çalik and İleri, 2007). Compared to data-driven methods, mechanistic approaches based on existing knowledge allows the extraction of deeper patterns in the underlying mechanism. More mechanistic models can subsequently be developed for context-specific interests (Long et al., 2015). For example, the integration of omics data can help study the metabolism under different conditions (Rai & Saito, 2016), where the data could also be used for validation and searching for solution space. The modeling methods have been widely applied for many bioproducts/ bioprocessing, including lipids, chemicals, food, and recombinant proteins (Calmels et al., 2019; Huang et al., 2020; Parichehreh et al., 2019; Aminian-Dehkordi et al., 2020). Conventional methods for achieving optimal fermentation conditions are through uniform design, orthogonal experimental design, and response surface methodology (RSM).

The development of integrated mechanistic models is increasingly challenging due to the fast-growing data size and dimensions. Table 2 summarizes the recent ML applications on fermentation biorefineries (including microbial lipid, bioethanol photo-fermentation, bacterial cellulose, etc.). To optimize fermentation conditions, researchers mainly apply ANN and SVM to build models. The predictive advantage of ANN against the conventional RSM method was investigated in the case of lipase production (Nelofer et al., 2012). Patil et al. (2017) similarly observed the advantage of ANN (over SVM and also RSM) for recovery prediction of arginine deiminase by *Pseudomonas putida*. In other similar comparisons, ANN often demonstrates a better performance of both R^2 and adjusted- R^2 values (Sydney et al., 2020; Monroy et al., 2018). In view of potentially obtaining local optimum and slow convergence, genetic algorithms can be together applied for searching optimum solution sets. In an application study of dark fermentation using Box-Behnken design, a hybrid approach that combines an ANN and RSM was applied to accelerate convergence and identification of critical parameters (Wang et al., 2021a). In a recent study of lipid fermentation from cellulosic ethanol wastewater, ANN and SVM were compared for their predictive performances on organic degradation, lipid, and biomass yields (Zhang et al., 2020). Regarding the scale of omics datasets, ANN possesses an appealing advantage in supporting the data analysis. For training models from smaller samples in similar applications, SVM is often more capable of locating optimal solutions than ANN. For other ML algorithms, RF and SVM were reported to predict the output targets (of solid recovery and cellulose enrichment) from cellulose-rich materials (Phromphithak et al., 2021). In this investigation, the best performance was achieved by RF, and the R^2 coefficients were 0.94 and 0.84, respectively. RF was also recently applied for fermentation prediction in microbial electrolysis cells (Wang et al., 2021b). While the most recent ML applications focus on the direct optimization of operating parameters (Dong and Chen, 2019; Ahmad et al., 2021; Vinitha et al., 2022), it has great potential for broader guidance (e.g. pathways control by gene-annotated strain in metabolic engineering) (Hannigan et al., 2019). It deserves more ML research efforts (e.g. gene editing targets for rational strain designs).

3.3. Anaerobic digestion

Anaerobic digestion is a recognized approach for the large-scale recycling of organic waste into bioenergy and digestate which is a potential organic fertilizer. Compared to the previous two biological methods (i.e. composting and fermentation), anaerobic digestion systems rely more on the balanced functional structure of diverse microbes to degrade organic matters, where a disturbance in the synergistic balance can directly affect reaction stability (Tsui et al., 2021; Lee et al., 2022a,b). In terms of system designs, the physiology and engineering requirements of methanogenic microorganisms fundamentally differ from acidogenic and acetogenic microbes. Different reactor configurations (e.g. continuously stirred tank reactor, leach bed reactor, fixed bed reactor, sequencing batch reactor, phase-separated system, and up-flow anaerobic sludge bed) have been developed for different operational and optimization needs (Tsui et al., 2018; Mao et al., 2021a,b).

Quantitative models are essential for understanding and optimizing

Table 2
Key findings from the literature on fermentation.

Biomass	Scientific Tasks	Major parameters	Algorithms	Best accuracy	Reference
Cheese whey	Prediction of hydrogen fermentation volume	pH, COD, HRT, Fe, Ni, Biomass proportion, Ethanol, Acetate, Butyrate	GB, SVR, RF, AdaBoost, MLP, LR, RR	GB; R^2: 0.985	Hosseinzadeh et al. (2020)
Tea, Sugar	Recovery prediction of bacterial cellulose from Kombucha	Mass and concentration of feedstocks, Mass of inoculum, pH, Duration, Temperature, Types of tea extracts	LR, XGB, PL	XGB; R^2: 0.9048	Priyadharshini et al. (2022)
Date pulp waste	Prediction of lactic acid production	pH, HRT, batch/ cyclic mode, enzymatic, and non-enzymatic pretreated samples	ANN	ANN; R^2:0.99	Ahmad et al. (2021)
Cellulosic ethanol	Prediction of microbial lipid fermentation	Biomass concentration, Time, Glucose concentration	BP-ANN, SVM	SVM; R^2: 0.996	Zhang et al. (2020)
Corn stalk	Prediction of biogas fermentation	Weight, Ultrasonic duration, Single/dual-frequency, Alkali pretreatment time	LS-SVM	LS-SVM; R^2: n.a.	De Clercq et al. (2019)
Mixed streams	Prediction of cellulose recovery	Cellulose enrichment factor, Solid recovery, Lignocellulosic characteristic, Pretreatment condition, ILS identity, Catalyst loading	SVM, RF, GB	RF; R^2: 0.94	Phromphithak et al. (2021)
Lignocellulosic hydrolysate	Identification of important genus in hydrogen electro-fermentation	Taxonomical information (Anode &cathode), Electrochemical performance	RF	RF; R^2: n.a.	Wang et al. (2021a, b)
Mixed streams	Optimization of bioethanol production	Biomass characteristics, Enzymatic saccharification process parameters, Glucose yield, Fermentation process parameters, Ethanol yield	SVM	SVM; R^2: 0.9762	Vinitha et al. (2022)
Sugarcane vinasse	Prediction of hydrogen production rate	Lactate, Acetate, Propionate, Butyrate, Time	ANN	ANN; R^2:0.987	Sydney et al. (2020)
Acetate, Butyrate, Sodium glutamate	Prediction of photo-fermentations rate	Time, Iron, Vitamin, Molybdenum, Light intensity, pH	ANN	ANN; R^2:0.939	Monroy et al. (2018)

Remarks: PL (Polynomial regression); BP(Backpropagation); RR(Ridge regression).

the implementation of anaerobic digesters, especially in evaluating energy balance and mitigation measures for improved economic sustainability (Tsui et al., 2022a, Tsui et al., 2022b). Simplified models are primarily based on empirical and kinetic methods. Empirical oxide-reduction reactions theoretically assume the complete organic degradation into carbon dioxide, methane, and other simple compounds. Since the methods do not incorporate factors (such as microbial growth and inhibition), they are good at the fast estimation of simple substrates (e.g. lactose and sucrose) but not process monitoring/ control. Fitzhugh model, cone model, transfer function, and modified Gompertz model are the most common kinetic models to simulate methane yields (Tsui et al., 2020). Anaerobic Digestion Model No. 1 (ADM1) is a comprehensive method that incorporates a wide range of kinetic equations. It has been proven the scale-up feasibility for simulating the impacts of substrate effects and operating modes on digester (Fatolahi et al., 2020). However, due to the complexity of physicochemical reactions and microbial conversions, the incomplete understanding of possible mechanistic relationships may hinder the prediction needs, and it also requires significant efforts for parameter calibration.

Data-driven ML technique is more independent of solving interactions involved in the ADM1 model, since the prediction could be entirely built on available data. Table 3 summarizes the most recent ML application of anaerobic digestion. ML algorithms are able to handle high-dimension data, but the selection of appropriate algorithms for anaerobic digestion has been shown to be critical to achieving the modeling purpose. Several ML algorithms (e.g. ANN, RF, SVM, k-nearest neighbors) and adaptive neuro-fuzzy interference systems (ANFIS) have been often applied for modeling the non-linear relationships in AD applications (Wang et al., 2020; Alejo et al., 2018; Park et al., 2021; Li et al., 2022b; Zareei and Khodaei, 2017). However, as shown in Table 3, there is no dominant advantage of a single algorithm having such a generalization ability for different scientific tasks. The different performances could be due to the variable format in their datasets as well as the design variation of biological systems. For the reactor configuration of upflow anaerobic sludge bed, a good prediction accuracy (R^2: 0.9793) of methane yields was achieved by the ANN model (Antwi et al., 2017). In the application of biogas prediction, ANN models are more effective on a controlled laboratory scale, while RF and XGBoost seem to be more accurate for industrial digesters of complex feedstocks (De Clercq et al., 2019 & De Clercq et al., 2020; Tufaner and Demirci, 2020). In anaerobic co-digestion of lignocellulose wastes (e.g maize straw and spent mushroom compost), ANFIS models were shown to have a better prediction of biogas performance (Najafi & Ardabili, 2018). Apart from biogas prediction/optimization, researchers also applied ML to study changes in effluent quality (e.g ammonia) in two-phase reactors system (Alejo et al., 2018). The beneficial use of zero-valent iron an additive for promoting methane enhancement by syntrophic communities was analyzed by ML application (Xu et al., 2021). Besides, Long et al. (2021) investigated genomic data for identifying key microbial species and control strategies, where the best prediction accuracy of R^2: 0.82 was achieved by RF. Though the datasets were obtained from limited experimental studies, the study provided an important exploration of genomic data for the prediction of complex interactions. Factually, the characteristic of the microbial community is known to be more determined (than short-term operating parameters) on the performance of anaerobic digestion (Vendruscolo et al., 2020).

3.4. Thermochemical conversions

Through the thermochemical processes of torrefaction/pyrolysis/gasification, the product outputs from biomass include syngas, bio-oil, and biochar. A hydrothermal method is a type of variation that emphasizes application for wet biomass (Song et al., 2021; Li et al., 2020b). Compared to the methods of biological conversion, thermochemical conversion requires more energy input in the presence of gasifying agents, and the conversion rate is much faster. Key controlling parameters include heating rate, temperature range, retention time, and catalyst applications. The reaction chemistry and fluid dynamics of thermochemical processes are sensitive to reaction conditions and feedstock properties. For example, bio-oil and syngas are the major products at low residence times (with high heating rates). In general, gasification (temperatures of >600 °C) favor secondary cracking reactions and syngas production which can be used for energy generation or chemical synthesis. At a lower temperature range (400 to 550 °C), bio-oil is generally observed in higher yield. A lower temperate (with an extended residence time) in torrefaction and slow pyrolysis could increase the amount of biochar recovery.

Modeling efforts for thermochemical processing include reactor configuration, process monitoring, and product upgrading. Operators can control process parameters (e.g. gasifying agents, biomass characteristics, pressure, and temperature) to deliver their ideal outputs. The approaches mainly include kinetic methods, thermodynamic

Table 3
Key findings from the literature on anaerobic digestion.

Biomass	Scientific Tasks	Major parameters	Algorithms	Best accuracy	Reference
Poultry manure	Ammonia prediction in two-phase system	TVS, COD, initial TAN	ANN, SVM	SVM; R^2: 0.898	Alejo et al. (2018)
Wheat straw	Co-digestion design	C/N ratio, temperature, retention time	LR, ANN, ANFIS	ANFIS; R^2: 0.9996	Najafi & Ardabili. (2018)
Mixed streams	Input study for industrial-scale system	Waste types (Food waste, percolate, chicken litter, fish waste, etc.)	LR, SVM, RF, XGBoost, kNN	kNN; R^2: 0.87	De Clercq et al. (2019)
Diverse streams	Identification of key microbial species for control strategies	Genomic data, VFAs, temperature, OLRs, HRT	GLMNET, RF, kNN, ANN, XGBoost	RF; R^2:0.82	Long et al. (2021)
Sewage sludge & Swine manure	Bioenergy enhancement by zero-valent iron	TS, vS COD, ISR, pH, temperature, dosage, particle size	RF, XGBoost, ANN	XGBoost; R^2: n.a	Xu et al. (2021)
Food waste	Prediction in continuous system	pH, alkalinity, COD, VFA	RF, XGboost, SVR, RNN	RNN; R^2: 0.9731	Park et al. (2021)
Mixed streams	Identification of key feedstock composition for biogas prediction	TC, TN, C/N ratio, cellulose, xylan, lignin, glucan content, temperature	RF, GLMNET, SVM, kNN	GLMNET; R^2: 0.73	Wang et al. (2020)
Mixed streams	Optimization of operating parameters	C/N, VS/TS, HRT, OLR, pH, SCOD, temperature, biochar dosage	Ensemble model	GBR &SVR; R^2:0.82& 0.86	Li et al. (2022b)
Mixed streams	Biogas prediction for industrial-scale digesters	Loading rates, waste types	RF, EN, XGBoost	XGBoost; R^2:0.88	De Clercq et al. (2020)
Cow manure & Maize straw	Optimization of mixing intensity	C/N ratio, TS, mixing intensity	ANFIS	ANFIS; R^2: 0.99	Zareei & Khodaei. (2017)
Bovine & Swine Slurry	Network analysis of microbial communities	Relative abundance of each taxon	RF	RF; R^2: n.a.	Vendruscolo et al., 2020

Remarks: ANFIS (adaptive neuro-fuzzy interference system), LR (Logistic regression), EN (Elastic net), RNN (Recurrent neural network), GB (Gradient boosting).

equilibrium, and computational fluid dynamics (CFD). Kinetic methods can simulate variables in temperature profiles and outputs composition under non-steady phases. Due to the microscopic complexity (e.g gas–solid contact, particle size evolution), good accuracy is usually difficult to achieve, and research efforts are ongoing to combine rate laws and CFD models for improved computation. Depending on the methodology (e.g Eq-sing or Eq-separate method), thermodynamic equilibrium can simulate maximum product output under ideal conditions (e.g. fully mixed conditions) and also actual yields (through empirical adjustments and reaction factors) (Cerinski et al., 2021). To consider the non-equilibrium product distributions, kinetic methods are necessary to assist the prediction. The overall predictive capability of thermochemical models builds on as much as consideration of all reactions and transport phenomena, but details are ignored to simplify assumptions of some cases. For instance, conversion reactions of hydrocarbon species (e.g. tar formation) are ignored in gasification systems.

As discussed, one challenge of thermodynamic equilibrium models is to handle a large number of mechanistic assumptions, while a reduced accuracy could be observed when the process conditions (e.g. low-temperature gasification) derail from equilibrium (Safarian et al., 2019). Researchers have applied ML algorithms (e.g. RF, SVM, ANN, and GB) to predict the gaseous products, bio-oil, and char yields from diverse waste streams. Through the parameter inputs of elemental analysis and chemical constituents, ML algorithms were compared to predict the heating value from different biomass (with the best accuracy of $R^2 > 0.94$ by RF) (Xing et al., 2019). Table 4 summarizes the recent publication of ML applications on thermochemical conversion. The algorithmic optimization together with ML models has shown the advantage of automated solutions searching against conventional methods (e.g. trial-and-error-based and RSM methods). With the help of feature selection by genetic algorithm, ML algorithms were compared for the prediction of bio-oil yield (best accuracy of $R^2 > 0.98$ by RF) (Ullah et al., 2021). For ML application of other biofuels, the recovery of methyl levulinate from waste peanut shells (using catalytic conversion at a lower temperature) was investigated by ANN (optimized by a genetic algorithm) (Li et al., 2020a). Direct hydrogen production as a biofuel has also been of interest to researchers (Tang et al., 2020; Li et al., 2022c). Traditionally, the non-linear systems relationships are difficult to be managed by building lots of assumptions. For the next stage of research efforts, more modeling explorations on environmental-friendly methods (e.g. suitable use of catalysts) and reactor designs (e.g. solar and microwave-assisted gasification) are important (Foong et al., 2020). Furthermore, the combined use of ML algorithms and optimization techniques shows great potential for their multi-objective optimization (e.g lower exhaust emission and higher utilization efficiency of engine performances).

4. Perspectives on research needs and future direction

The above sections examine the recent ML applications in each biorefinery system as well as the status of mechanistic approaches for mathematical modeling. As discussed, mechanistic approaches alone cannot often be squarely handled for all modeling needs, and it has resulted in most researchers/ engineers tending to use simplified models. In the past, artificial intelligence technologies could never become a standard tool in biorefinery or the larger context of circular bioeconomy, but it is clear that recent ML developments have triggered more interest by researchers, as reflected by the recent surge of publications on the topic. This review article illustrates the wide range of feasibility for ML modeling and process control of bioproduct/ biofuel recovery. ML can serve as an important modeling tool for solving non-linear relationships, but the performance of different algorithms varied in those scientific tasks. The review findings indicate there is no single algorithm perfect (or sufficiently robust) for all modeling needs of targeted resource efficiencies. Thus, the updated summary of their

Table 4
Key findings from the literature on thermochemical conversions.

Biomass	Scientific Tasks	Major parameters	Algorithms	Best accuracy	Reference
Agriculture & forest waste, Algae	Prediction of gaseous products	AC-FC-V, CHNO, Highest temperature, Heating rate, Particle size, Sweep gas flow rate	RF, SVM	RF; R^2: 0.85–0.87	Tang et al. (2021)
Lignocellulosic biomass, Herbaceous plant, Algae-based feedstock	Prediction of bio-oil yield	Particle size, Nitrogen flow rate, AC-FC-V, CHNO, Heating rate, Highest Temperature	SVM, RF, DT, MLR	RF; R^2: 0.98	Ullah et al. (2021)
Lignocellulose biomass	Prediction of biochar yield and carbon contents	Pyrolysis conditions, Lignin, Cellulose, Hemicellulose, AC-FC-V, CHNO, Particle size	RF	RF; R^2:0.8548	Zhu et al. (2019)
Fermentation biomass	Prediction of remaining residuals	Heating rate, Temperature, Weight loss at various heating rates	SVM	SVM; R^2:0.9999	Shahbeig and Nosrati (2020)
Mixed streams	Prediction of product distribution and bio-oil heating value during fast pyrolysis	Fluidization number, Pyrolysis temperature, AC-FC-V, CHNO, Particle size	ANN, SVM	ANN; R^2: 0.988	Chen et al. (2018)
Waste peanut shells	Optimization of methyl levulinate yield	Ratio of waste peanut shells to methanol, Metal sulfate loading, Reaction temperature, Reaction time	ANN-GA	ANN; R^2: 0.89	Li et al. (2020a)
Mixed streams	Prediction of bio-oil yield and hydrogen contents	Ash content, CHNO, Highest temperature, Heating rate, Liquid oil production, Particle size, Nitrogen flow rate	RF, MLR	RF; R^2: 0.92 & 0.79	Tang et al. (2020)
Lignocellulosic biomass	Prediction of solid products from biomass torrefaction	Moisture content, CHNO, Ash content, Sample size, Residence time, Temperature, Fractions of reacting gas	Lasso, Ridge regression, KRR, DT, AdaBoost, GB, RF, ET, kNNs, SVM	GB; R^2:0.9	Onsree and Tippayawong (2021)
Mixed streams	Optimization of hydrothermal gasification conditions	Biomass characteristics, Temperature, Pressure, Catalyst loading, Solvent to biomass ratio, Residence time, Product gas yield	Random Forest	RF; R^2:0.9878 &0.9865	Gopirajan et al. (2021)
Fruits, Agri-wastes, Briquettes/pellets, Industry wastes, Forest wastes	Prediction of biomass HHV from ultimate or proximate analysis	Mass fractions of fixed carbon, Volatile matter, Ash, CHNO	ANN, SVM, RF	RF; R^2: 0.94	Xing et al. (2019)

Remarks: MLR(Multi-Linear regression), GA (Genetic algorithms), ET (Extremely-randomized trees), KRR(Kernel ridge regression).

algorithmic performances has become more crucial for reference by researchers. Based on the above findings, prioritization can be made to explore algorithms according to their proven applicability.

Currently, data quality could be a major challenge for the next-stage ML application. During the review process, it was observed that using homogeneous datasets from the same research group often led to better prediction performance. And for studies collecting data from literature of experimental research, there is often a lack of clear consistency or clarity of data processing reported by researchers. One good example of data concerns is anaerobic digestion. Experimental studies have been observed for inconsistencies in their analytical methodology (e.g whether the BMP protocol was applied). The uninformed practices could lead to significant differences in models' training and the results of model comparison. Therefore, a lack of data protocol for each biorefinery system is a crucial factor hindering concrete guidance of what approaches are potentially more applicable for the data training and follow-up applications. It will be valuable to understand the impacts of data quality (including experimental designs and data processing steps) on computational improvement. In parallel efforts, experimentalists should be informed about data specification for improving experimental designs more systematically.

Algorithms with simpler structures are often stereotyped as having a less ideal performance for complex relationships, while those complex structures (e.g. ANN) might achieve better performance but less interpretability. While approaches (such as partial dependence analysis) were developed to support a partial interpretation of internal ML mechanisms, only very limited studies have further applied them in recent biorefinery publications. It was also seldom addressed by researchers whether their improved performance comes with increasing computational costs (e.g. larger size of training data and more data processing). Researchers need to consider the tradeoffs between accuracy needs and computational costs. For example, simple models may present advantages for real-time process control as compared to complex algorithms. Developing new strategies of ensemble learning models and more robust evaluation approaches will be important to meet the increasing needs of different scientific tasks. The field of ML had started to grow for some years, but in the biorefinery community, with very few exceptions, a lag of about 3–5 years was observed in their application status. For future direction (already happening in artificial intelligence frontiers), the combination of rapid computation and reliable data sources can offer opportunities for the creation of digital twins and therefore better process/ tracking control (e.g. high-frequency adaptive optimization and blockchain technologies), leading to more innovation systems in circular bioeconomy. The intergenerational well-being from global climate actions needs more innovations for sustainability pursuits, and these will require more scientific & technological progress from collective efforts.

5. Conclusions

Recent progress in machine learning (ML) provides more accessible opportunities for digitalization to biorefinery communities, contributing to a more progressive development of a circular bioeconomy. The joint innovation will unfold new potentials of sustainability efforts as a systems approach to supporting sustainable development goals. This review article systematically examines the status of recent ML applications for four critical biorefinery systems (including composting, fermentation, anaerobic digestion, and thermochemical conversions) as well as their advancement of scientific/ technological tasks. Perspectives are sketched to convey research needs and future direction.

CRediT authorship contribution statement

To-Hung Tsui: Conceptualization, Writing – original draft, Writing – review & editing. **Mark C. M. van Loosdrecht:** Writing – review & editing. **Yanjun Dai:** Funding acquisition, Project administration. **Yen Wah Tong:** Supervision, Funding acquisition, Project administration.

Declaration of Competing Interest

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

Data availability

Data will be made available on request.

Acknowledgements

This research is supported by the National Research Foundation, Prime Minister's Office, Singapore under its Campus for Research Excellence and Technological Enterprise (CREATE) programme. The authors express gratefulness for the kind invitation to this meaningful special issue.

References

Ahmad, A., Banat, F., Taher, H., 2021. Comparative study of lactic acid production from date pulp waste by batch and cyclic–mode dark fermentation. Waste Manag. 120, 585–593.
Alavi, N., Sarmadi, K., Goudarzi, G., Babaei, A.A., Bakhshoodeh, R., Paydary, P., 2019. Attenuation of tetracyclines during chicken manure and bagasse co-composting: Degradation, kinetics, and artificial neural network modeling. J. Environ. Manage. 231, 1203–1210.
Alejo, L., Atkinson, J., Guzmán-Fierro, V., Roeckel, M., 2018. Effluent composition prediction of a two-stage anaerobic digestion process: machine learning and stoichiometry techniques. Environ. Sci. Pollut. Res. 25 (21), 21149–21163.
Aminian-Dehkordi, J., Mousavi, S.M., Marashi, S.A., Jafari, A., Mijakovic, I., 2020. A systems-based approach for cyanide overproduction by Bacillus megaterium for gold bioleaching enhancement. Front. Bioeng. Biotechnol. 8, 528.
Antwi, P., Li, J., Boadi, P.O., Meng, J., Shi, E., Deng, K., Bondinuba, F.K., 2017. Estimation of biogas and methane yields in an UASB treating potato starch processing wastewater with backpropagation artificial neural network. Bioresour. Technol. 228, 106–115.
Bordbar, A., Monk, J.M., King, Z.A., Palsson, B.O., 2014. Constraint-based models predict metabolic and associated cellular functions. Nat. Rev. Genet. 15 (2), 107–120.
Çalik, P., İleri, N., 2007. pH influences intracellular reaction network of β-lactamase producing Bacillus licheniformis. Chem. Eng. Sci. 62 (18–20), 5206–5211.
Calmels, C., McCann, A., Malphettes, L., Andersen, M.R., 2019. Application of a curated genome-scale metabolic model of CHO DG44 to an industrial fed-batch process. Metab. Eng. 51, 9–19.
Cerinski, D., Ferreiro, A.I., Baleta, J., Costa, M., Zimbardi, F., Cerone, N., Wang, J., 2021. Modelling the biomass updraft gasification process using the combination of a pyrolysis kinetic model and a thermodynamic equilibrium model. Energy Rep. 7, 8051–8061.
Chen, X., Zhang, H., Song, Y., Xiao, R., 2018. Prediction of product distribution and bio-oil heating value of biomass fast pyrolysis. Chem. Eng. Process.-Process Intensif. 130, 36–42.
Clauser, N.M., Felissia, F.E., Area, M.C., Vallejos, M.E., 2021. A framework for the design and analysis of integrated multi-product biorefineries from agricultural and forestry wastes. Renew. Sustain. Energy Rev. 139, 110687.
Cruz, I.A., Chuenchart, W., Long, F., Surendra, K.C., Andrade, L.R.S., Bilal, M., Ferreira, L.F.R., 2022. Application of machine learning in anaerobic digestion: Perspectives and challenges. Bioresour. Technol. 345, 126433.
De Clercq, D., Jalota, D., Shang, R., Ni, K., Zhang, Z., Khan, A., Yuan, K., 2019. Machine learning powered software for accurate prediction of biogas production: A case study on industrial-scale Chinese production data. J. Clean. Prod. 218, 390–399.
De Clercq, D., Wen, Z., Fei, F., Caicedo, L., Yuan, K., Shang, R., 2020. Interpretable machine learning for predicting biomethane production in industrial-scale anaerobic co-digestion. Sci. Total Environ. 712, 134574.
Ding, S., Huang, W., Xu, W., Wu, Y., Zhao, Y., Fang, P., Lou, L., 2022. Improving kitchen waste composting maturity by optimizing the processing parameters based on machine learning model. Bioresour. Technol. 127606.
Dong, C., Chen, J., 2019. Optimization of process parameters for anaerobic fermentation of corn stalk based on least squares support vector machine. Bioresour. Technol. 271, 174–181.
Fatolahi, Z., Arab, G., Razaviarani, V., 2020. Calibration of the Anaerobic Digestion Model No. 1 for anaerobic digestion of organic fraction of municipal solid waste under mesophilic condition. Biomass Bioenergy 139, 105661.
Foong, S.Y., Liew, R.K., Yang, Y., Cheng, Y.W., Yek, P.N.Y., Mahari, W.A.W., Lam, S.S., 2020. Valorization of biomass waste to engineered activated biochar by microwave pyrolysis: Progress, challenges, and future directions. Chem. Eng. J. 389, 124401.

Gopirajan, P.V., Gopinath, K.P., Sivaranjani, G., Arun, J., 2021. Optimization of hydrothermal gasification process through machine learning approach: Experimental conditions, product yield and pollution. J. Clean. Prod. 306, 127302.
Hannigan, G.D., Prihoda, D., Palicka, A., Soukup, J., Klempir, O., Rampula, L., Bitton, D. A., 2019. A deep learning genome-mining strategy for biosynthetic gene cluster prediction. Nucleic Acids Res. 47 (18), e110–e.
Hastie, T., Tibshirani, R., Friedman, J., 2009. Kernel smoothing methods. In: The Elements of Statistical Learning. Springer, New York, NY, pp. 191–218.
Hosseinzadeh, A., Baziar, M., Alidadi, H., Zhou, J.L., Altaee, A., Najafpoor, A.A., Jafarpour, S., 2020. Application of artificial neural network and multiple linear regression in modeling nutrient recovery in vermicompost under different conditions. Bioresour. Technol. 303, 122926.
Huang, Z., Xu, J., Yongky, A., Morris, C.S., Polanco, A.L., Reily, M., Yoon, S., 2020. CHO cell productivity improvement by genome-scale modeling and pathway analysis: Application to feed supplements. Biochem. Eng. J. 160, 107638.
Karlik, B., Olgac, A.V., 2011. Performance analysis of various activation functions in generalized MLP architectures of neural networks. Internat. J. Artif. Intell. Expert Syst. 1 (4), 111–122.
Kaushal, M., Chary, K.V.N., Ahlawat, S., Palabhanvi, B., Goswami, G., Das, D., 2018. Understanding regulation in substrate dependent modulation of growth and production of alcohols in Clostridium sporogenes NCIM 2918 through metabolic network reconstruction and flux balance analysis. Bioresour. Technol. 249, 767–776.
Kyakuwaire, M., Olupot, G., Amoding, A., Nkedi-Kizza, P., Ateenyi Basamba, T., 2019. How safe is chicken litter for land application as an organic fertilizer?: A review. Int. J. Environ. Res. Public Health 16 (19), 3521.
Lee, J.T., Dutta, N., Zhang, L., Tsui, T.T., Lim, S., Tio, Z.K., Tong, Y.W., 2022a. Bioaugmentation of Methanosarcina thermophila grown on biochar particles during semi-continuous thermophilic food waste anaerobic digestion under two different bioaugmentation regimes. Bioresour. Technol. 360, 127590.
Lee, J.T., Lim, E.Y., Zhang, L., Tsui, T.H., Tian, H., Yan, M., Tong, Y.W., 2022b. Methanosarcina thermophila bioaugmentation and its synergy with biochar growth support particles versus polypropylene microplastics in thermophilic food waste anaerobic digestion. Bioresour. Technol. 360, 127531.
Li, P., Du, Z., Chang, C., Zhao, S., Xu, G., Xu, C.C., 2020b. Efficient catalytic conversion of waste peanut shells into liquid biofuel: an artificial intelligence approach. Energy Fuel 34 (2), 1791–1801.
Li, Y., Li, S., Sun, X., Hao, D., 2022c. Prediction of carbon dioxide production from green waste composting and identification of critical factors using machine learning algorithms. Bioresour. Technol. 360, 127587.
Li, J., Pan, L., Suvarna, M., Tong, Y.W., Wang, X., 2020a. Fuel properties of hydrochar and pyrochar: Prediction and exploration with machine learning. Appl. Energy 269, 115166.
Li, J., Zhu, X., Li, Y., Tong, Y.W., Ok, Y.S., Wang, X., 2021. 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.
Li, J., Li, L., Tong, Y.W., Wang, X., 2022a. Understanding and optimizing the gasification of biomass waste with machine learning. Green Chemical Engineering.
Li, J., Zhang, L., Li, C., Tian, H., Ning, J., Zhang, J., Wang, X., 2022b. Data-driven based in-depth interpretation and inverse design of anaerobic digestion for CH4-rich biogas production. ACS ES&T Engineering 2 (4), 642–652.
Long, M.R., Ong, W.K., Reed, J.L., 2015. Computational methods in metabolic engineering for strain design. Curr. Opin. Biotechnol. 34, 135–141.
Long, F., Wang, L., Cai, W., Lesnik, K., Liu, H., 2021. Predicting the performance of anaerobic digestion using machine learning algorithms and genomic data. Water Res. 199, 117182.
Mao, L., Tsui, T.H., Zhang, J., Dai, Y., Tong, Y.W., 2021a. System integration of hydrothermal liquefaction and anaerobic digestion for wet biomass valorization: Biodegradability and microbial syntrophy. J. Environ. Manage. 293, 112981.
Mao, L., Tsui, T.H., Zhang, J., Dai, Y., Tong, Y.W., 2021b. Mixing effects on decentralized high-solid digester for horticultural waste: Startup, operation and sensitive microorganisms. Bioresour. Technol. 333, 125216.
Moncks, P.C.S., Corrêa, É.K., Guidoni, L.L.C., Moncks, R.B., Corrêa, L.B., Lucia Jr, T., Marques, F.S., 2022. Moisture content monitoring in industrial-scale composting systems using low-cost sensor-based machine learning techniques. Bioresour. Technol. 359, 127456.
Monroy, I., Guevara-López, E., Buitrón, G., 2018. Biohydrogen production by batch indoor and outdoor photo-fermentation with an immobilized consortium: a process model with Neural Networks. Biochem. Eng. J. 135, 1–10.
Najafi, B., Ardabili, S.F., 2018. Application of ANFIS, ANN, and logistic methods in estimating biogas production from spent mushroom compost (SMC). Resour. Conserv. Recycl. 133, 169–178.
Nelofer, R., Ramanan, R.N., Rahman, R.N.Z.R.A., Basri, M., Ariff, A.B., 2012. Comparison of the estimation capabilities of response surface methodology and artificial neural network for the optimization of recombinant lipase production by E. coli BL21. J. Ind. Microbiol. Biotechnol. 39 (2), 243–254.
Onsree, T., Tippayawong, N., 2021. Machine learning application to predict yields of solid products from biomass torrefaction. Renew. Energy 167, 425–432.
Parichehreh, R., Gheshlaghi, R., Mahdavi, M.A., Elkamel, A., 2019. Optimization of lipid production in Chlorella vulgaris for biodiesel production using flux balance analysis. Biochem. Eng. J. 141, 131–145.
Park, J.G., Jun, H.B., Heo, T.Y., 2021. Retraining prior state performances of anaerobic digestion improves prediction accuracy of methane yield in various machine learning models. Appl. Energy 298, 117250.
Patil, M.D., Dev, M.J., Tangadpalliwar, S., Patel, G., Garg, P., Chisti, Y., Banerjee, U.C., 2017. Ultrasonic disruption of Pseudomonas putida for the release of arginine deiminase: Kinetics and predictive models. Bioresour. Technol. 233, 74–83.
Phromphithak, S., Onsree, T., Tippayawong, N., 2021. Machine learning prediction of cellulose-rich materials from biomass pretreatment with ionic liquid solvents. Bioresour. Technol. 323, 124642.
Priyadharshini, T., Nageshwari, K., Vimaladhasan, S., Prakash, S.P., Balasubramanian, P., 2022. Machine learning prediction of SCOBY cellulose yield from Kombucha tea fermentation. Bioresour. Technol. Rep. 18, 101027.
Rai, A., Saito, K., 2016. Omics data input for metabolic modeling. Curr. Opin. Biotechnol. 37, 127–134.
Ren, X., Wang, Z., Zhao, M., Xie, J., Zhang, Z., Yang, F., Ding, Y., 2022. Role of selenite on the nitrogen conservation and greenhouse gases mitigation during the goat manure composting process. Sci. Total Environ. 155799.
Safarian, S., Unnþórsson, R., Richter, C., 2019. A review of biomass gasification modelling. Renew. Sustain. Energy Rev. 110, 378–391.
Schmidhuber, J., 2015. Deep learning in neural networks: An overview. Neural Netw. 61, 85–117.
Shahbeig, H., Nosrati, M., 2020. Pyrolysis of biological wastes for bioenergy production: Thermo-kinetic studies with machine-learning method and Py-GC/MS analysis. Fuel 269, 117238.
Somvanshi, M., Chavan, P., Tambade, S., Shinde, S.V., 2016. In: A Review of MAchine LeArning Techniques Using Decision Tree and Support Vector MAchine. IEEE, pp. 1–7.
Song, B., Lin, R., Lam, C.H., Wu, H., Tsui, T.H., Yu, Y., 2021. Recent advances and challenges of inter-disciplinary biomass valorization by integrating hydrothermal and biological techniques. Renew. Sustain. Energy Rev. 135, 110370.
Soto-Paz, J., Oviedo-Ocaña, E.R., Manyoma, P.C., Marmolejo-Rebellón, L.F., Torres-Lozada, P., Barrena, R., Komilis, D., 2020. Influence of mixing ratio and turning frequency on the co-composting of biowaste with sugarcane filter cake: a mixture experimental design. Waste Biomass Valoriz. 11 (6), 2475–2489.
Stegmann, P., Londo, M., Junginger, M., 2020. The circular bioeconomy: Its elements and role in European bioeconomy clusters. Resources, Conservation & Recycling: X 6, 100029.
Sun, Y., Ren, X., Pan, J., Zhang, Z., Tsui, T.H., Luo, L., Wang, Q., 2020. Effect of microplastics on greenhouse gas and ammonia emissions during aerobic composting. Sci. Total Environ. 737, 139856.
Sydney, E.B., Duarte, E.R., Burgos, W.J.M., de Carvalho, J.C., Larroche, C., Soccol, C.R., 2020. Development of short chain fatty acid-based artificial neuron network tools applied to biohydrogen production. Int. J. Hydrogen Energy 45 (8), 5175–5181.
Tang, Q., Chen, Y., Yang, H., Liu, M., Xiao, H., Wu, Z., Naqvi, S.R., 2020. Prediction of bio-oil yield and hydrogen contents based on machine learning method: effect of biomass compositions and pyrolysis conditions. Energy Fuel 34 (9), 11050–11060.
Tang, Q., Chen, Y., Yang, H., Liu, M., Xiao, H., Wang, S., Naqvi, S.R., 2021. Machine learning prediction of pyrolytic gas yield and compositions with feature reduction methods: effects of pyrolysis conditions and biomass characteristics. Bioresour. Technol. 339, 125581.
Tsui, T.H., Chen, L., Hao, T., Chen, G.H., 2016. A super high-rate sulfidogenic system for saline sewage treatment. Water Res. 104, 147–155.
Tsui, T.H., Ekama, G.A., Chen, G.H., 2018. Quantitative characterization and analysis of granule transformations: Role of intermittent gas sparging in a super high-rate anaerobic system. Water Res. 139, 177–186.
Tsui, T.H., Wong, J.W., 2019. A critical review: emerging bioeconomy and waste-to-energy technologies for sustainable municipal solid waste management. Waste Disposal & Sustainable Energy 1 (3), 151–167.
Tsui, T.H., Wu, H., Song, B., Liu, S.S., Bhardwaj, A., Wong, J.W., 2020. Food waste leachate treatment using an Upflow Anaerobic Sludge Bed (UASB): Effect of conductive material dosage under low and high organic loads. Bioresour. Technol. 304, 122738.
Tsui, T.H., Zhang, L., Lim, E.Y., Lee, J.T., Tong, Y.W., 2021. Timing of biochar dosage for anaerobic digestion treating municipal leachate: Altered conversion pathways of volatile fatty acids. Bioresour. Technol. 335, 125283.
Tsui, T.H., Zhang, L., Zhang, J., Dai, Y., Tong, Y.W., 2022a. Methodological framework for wastewater treatment plants delivering expanded service: Economic tradeoffs and technological decisions. Sci. Total Environ. 823, 153616.
Tsui, T.H., Zhang, L., Zhang, J., Dai, Y., Tong, Y.W., 2022b. Engineering interface between bioenergy recovery and biogas desulfurization: Sustainability interplays of biochar application. Renew. Sustain. Energy Rev. 157, 112053.
Tufaner, F., Demirci, Y., 2020. Prediction of biogas production rate from anaerobic hybrid reactor by artificial neural network and nonlinear regressions models. Clean Techn. Environ. Policy 22 (3), 713–724.
Ullah, Z., Naqvi, S.R., Farooq, W., Yang, H., Wang, S., Vo, D.V.N., 2021. A comparative study of machine learning methods for bio-oil yield prediction–A genetic algorithm-based features selection. Bioresour. Technol. 335, 125292.
Vendruscolo, E.C.G., Mesa, D., Rissi, D.V., Meyer, B.H., de Oliveira Pedrosa, F., de Souza, E.M., Cruz, L.M., 2020. Microbial communities network analysis of anaerobic reactors fed with bovine and swine slurry. Sci. Total Environ. 742, 140314.
Vinitha, N., Vasudevan, J., Gopinath, K.P., 2022. Bioethanol production optimization through machine learning algorithm approach: biomass characteristics, saccharification, and fermentation conditions for enzymatic hydrolysis. Biomass Convers. Biorefin. 1–13.
Wang, L., Long, F., Liao, W., Liu, H., 2020. Prediction of anaerobic digestion performance and identification of critical operational parameters using machine learning algorithms. Bioresour. Technol. 298, 122495.
Wang, L., Long, F., Liang, D., Xiao, X., Liu, H., 2021a. Hydrogen production from lignocellulosic hydrolysate in an up-scaled microbial electrolysis cell with stacked bio-electrodes. Bioresour. Technol. 320, 124314.
Wang, Y., Yang, G., Sage, V., Xu, J., Sun, G., He, J., Sun, Y., 2021b. Optimization of dark fermentation for biohydrogen production using a hybrid artificial neural network

(ANN) and response surface methodology (RSM) approach. Environ. Prog. Sustain. Energy 40 (1), e13485.
Xing, J., Luo, K., Wang, H., Gao, Z., Fan, J., 2019. A comprehensive study on estimating higher heating value of biomass from proximate and ultimate analysis with machine learning approaches. Energy 188, 116077.
Xu, W., Long, F., Zhao, H., Zhang, Y., Liang, D., Wang, L., Liu, H., 2021. Performance prediction of ZVI-based anaerobic digestion reactor using machine learning algorithms. Waste Manag. 121, 59–66.
Yamawaki, R., Tei, A., Ito, K., Kikuchi, J., 2021. Decomposition factor analysis based on virtual experiments throughout Bayesian optimization for compost-degradable polymers. Appl. Sci. 11 (6), 2820.
Yang, H.Y., Zhang, S.B., Meng, H.H., Zhao, Y., Wei, Z.M., Zheng, G.R., Wang, X., 2022. Predicting the humification degree of multiple organic solid waste during composting using a designated bacterial community. Waste Manage. 150, 257–266.
Zareei, S., Khodaei, J., 2017. Modeling and optimization of biogas production from cow manure and maize straw using an adaptive neuro-fuzzy inference system. Renew. Energy 114, 423–427.
Zhang, L., Chao, B., Zhang, X., 2020. Modeling and optimization of microbial lipid fermentation from cellulosic ethanol wastewater by Rhodotorula glutinis based on the support vector machine. Bioresour. Technol. 301, 122781.
Zhang, L., Tsui, T.H., Loh, K.C., Dai, Y., Tong, Y.W., 2021. Effects of plastics on reactor performance and microbial communities during acidogenic fermentation of food waste for production of volatile fatty acids. Bioresour. Technol. 337, 125481.
Zhang, L., Jiang, Z., Tsui, T.H., Loh, K.C., Dai, Y., Tong, Y.W., 2022a. A Review on Enhancing Cupriavidus necator Fermentation for Poly (3-hydroxybutyrate) (PHB) Production From Low-Cost Carbon Sources. Front. Bioeng. Biotechnol. 10.
Zhang, L., Tsui, T.H., Loh, K.C., Dai, Y., Tong, Y.W., 2022b. Acidogenic fermentation of organic wastes for production of volatile fatty acids. In: Biomass, Biofuels, Biochemicals. Elsevier, pp. 343–366.
Zhu, X., Li, Y., Wang, X., 2019. Machine learning prediction of biochar yield and carbon contents in biochar based on biomass characteristics and pyrolysis conditions. Bioresour. Technol. 288, 121527.