Systematic Review of Machine-learning Techniques to Support Development of Lignocellulose Biorefineries

A. Jurinjak Tušek,[a] A. Petrus,[b] A. Weichselbraun,[b] R.-P. Mundani,[b] S. Müller,[b] I. Barkow,[b] A. Bucić-Kojić,[c] M. Planinić,[c] and M. Tišma[c,*]

[a]University of Zagreb Faculty of Food Technology and Biotechnology, Pierottijeva 6, 10000 Zagreb, Croatia

[b]Swiss Institute for Information Science, University of Applied Sciences of the Grisons, Ringstrasse 34, 7000 Chur, Switzerland

[c]Josip Juraj Strossmayer University of Osijek, Faculty of Food Technology Osijek, Franje Kuhača 18, Osijek, HR-31000, Croatia

doi: https://doi.org/10.15255/CABEQ.2023.2273

Review
Received: November 13, 2023
Accepted: July 23, 2024

Lignocellulosic biorefineries (LBRs) are platforms for the production of a variety of bio-based products such as biofuels, biomaterials, biochemicals, food, and feed using lignocellulosic biomass (LB) as feedstock. LBRs are still rare worldwide. Their commercialization depends on challenges associated with the entire feedstock supply chain, efficiency, sustainability, and scale-up of pretreatment methods, as well as isolation and purification of value-added products. Each step within LBRs requires the development of new technologies or the improvement of existing ones, considering all three sustainability dimensions, environmental, social, and economic. Machine learning (ML) methods are widely used in various industrial fields, including biotechnology. The merging of biotechnology and ML has driven scientific progress and opened new opportunities for the development of LBRs as well. In this review, ML methods and their efficiency, used in biotechnology (metabolic engineering, bioprocess development, and environmental engineering), are presented, followed by their application in various phases of LB valorization.

Keywords
lignocellulosic biomass, lignocellulosic biorefinery, machine learning, sustainability

Introduction

The linear economy operates on the "take-make-waste" model, where raw materials are extracted from the earth, products are manufactured and consumed, and then disposed of as waste. Production processes generate significant waste streams, including gas emissions that contribute to the greenhouse effect and climate change. In contrast, the circular (bio)economy aims to recycle materials and minimize waste accumulation. While the circular economy includes various feedstocks, such as fossil-based materials, the bioeconomy specifically focuses on biomass, with recent emphasis on residual LB. A bioeconomy based on the innovative and cost-effective use of residual LB for the production of bio-based products should be driven by integrated LBRs[1–3]. These biorefineries are often compared to traditional petrochemical refineries in terms of feedstocks, building block composition, processes, and the chemical intermediates produced at a commercial scale[1]. Although LB is chemically complex, it is no more so than petroleum[4,5]. However, unlike petroleum, it is available worldwide. The entire supply chain in LBRs depends on many factors, such as the seasonal availability of feedstock, transportation and storage challenges due to its large volume, and chemical complexity, among others[6]. To ensure sustainability, all three main components of LB (lignin, cellulose, and hemicellulose) must be utilized in environmentally friendly and cost-effective ways for the co-production of multi-

*Correspondence: marina.tisma@ptfos.hr; Tel.: 00385 31 224 358

ple bio-based products in LBRs, which poses significant challenges. High capital and operating expenditures, irregularities in the biomass supply chain, immature technologies (such as pretreatment, isolation and purification, catalytic or biocatalytic conversion, etc.), and scale-up difficulties are the reasons why LBRs remain relatively rare worldwide[7].

Due to its complexity, the biorefinery concept presents challenges for industry, decision-makers, and investors in identifying the most promising options and assessing technological and economic risks[1]. The initial phase in LBRs involves pretreatment. LB pretreatment methods can generally be categorized into physical, chemical, physicochemical, and biological approaches[8,9]. Selecting a pretreatment method requires finding a cost-effective and environmentally friendly approach that avoids the formation of inhibitors[7]. Additionally, challenges arise in scaling up processes to extract, and isolate, and convert polymers from biomass into desired products sustainably. One novel and sustainable extraction method for LBRs is the use of deep eutectic solvents, as recently reviewed by Sharma *et al*.[10] A life cycle assessment should be employed as a decision support tool early in the product development stage within LBRs[11,12]. New technologies are needed to monitor, automatically control, and accurately predict each step of the biorefinery process. In this context, ML is of great interest, as it can help save labor and time in experiments, as highlighted in numerous recent publications[13–19]. ML, a branch of artificial intelligence, was developed to address the need for efficient tools to analyze large datasets due to the automation of experimental equipment and the enormous increase in computer resources, which have generated numerous datasets[20]. Using statistical methods, ML enables computers to learn from data and make judgments without being explicitly programmed[21], allowing systems to improve performance over time by learning from experience. Consequently, ML has become crucial for the development of various industries, new products and services, data analysis, and visualization[22], including the advancement of biorefineries[13–19].

The proposed review aims to provide researchers and decision-makers in the field of LBR development with insights into the application of ML methods in various biorefinery steps. The first part discusses general information on LBRs, emphasizing technological challenges. The second part focuses on the analysis of ML algorithms and describes the development of ML models, including the importance of selecting model input variables. The third part reviews traditional applications of ML in several fields of biotechnology. Finally, the review examines the use of ML in various process steps within LBRs and outlines future research opportunities and perspectives.

Concept of lignocellulose biorefineries

The concept of biorefineries emerged in response to increasing fossil resource prices, their uncertain availability, and global environmental concerns. Biorefineries vary widely depending on the feedstocks used, the type of intermediates generated (such as syngas or sugar), the conversion processes employed (thermochemical, biochemical, two-platform), and the stage of technological development (conventional, advanced, etc.)[4,23]. They can also be categorized based on the different generations of feedstock used: first-, second-, third-, and fourth-generation biorefineries[24–26]. Some specific characteristics of different types of biorefineries are detailed in Table 1.

LBRs are notable for their ability to produce multiple products from LB through several process stages, as presented in Fig. 1a. These biorefineries can be either energy- or material-driven[23]. The conversion of lignocellulosic biomass into high-value chemicals involves several key steps, including pretreatment, hydrolysis, and the subsequent conversion of the resulting sugars and other components into desired products (Fig. 1b)[27,28].

In addition to biological processes, various thermochemical processes have been developed for biomass utilization. Key information about these thermochemical processes[29–33] is summarized in Table 2.

LB constitutes more than 90 % of all biomass and does not compete with food resources[34]. This includes energy crops, grasses, and biological residues from various industrial processes, such as harvest residues, wood industry residues, food industry residues, and household biowaste, etc.[34–37]

LB is primarily composed of lignin, cellulose, and hemicellulose present in varying amounts and ratios depending on the type and origin of the biomass. It also contains smaller amounts of pectin, protein, extractives, and inorganic compounds. The complex and variable chemical structure of LB poses a barrier to its widespread use in producing bio-based products[3,38]. Determining the chemical composition of LB is a crucial first step in selecting the appropriate pretreatment method[39]. Conventional methods for this purpose are often time-consuming and not environmentally friendly, highlighting the need for rapid and accurate alternatives[40,41]. The choice of feedstock is a critical determinant in utilizing LB for chemical production, significantly influencing both the efficiency and sustainability of the process[42,43]. Feedstock composition, which in-

Table 1 – *Overview of various types of biorefineries*[24–26]

Biorefinery type	Feedstock	Typical product	Advantages	Disadvantages
First-generation	Food crops (corn, sugarcane, or soybeans)	Bioethanol Biodiesel	Established technology with well-understood processes. Provides an alternative use for agricultural produce.	Competes with food supply. Limited feedstock availability.
Second-generation	Non-food biomass, including agricultural residues (e.g., straw, corn stover), woody biomass, and dedicated energy crops like switchgrass or miscanthus	Cellulosic ethanol Bio-based chemicals	Does not compete with food supply. Utilizes a broader range of feedstocks.	More complex and costly technology. Requires advanced pretreatment and processing methods.
Third-generation	Microalgae and other aquatic biomass as feedstocks	Algal biofuels Algae-derived chemicals	High productivity and growth rates of algae. Can be grown on non-arable land and using non-potable water. Potential for wastewater treatment.	High water and nutrient requirements. Challenges in harvesting and processing algae economically.
Fourth-generation	Focuses on advanced biotechnological methods, including genetic engineering and synthetic biology, to optimize feedstock production and conversion processes	Genetically modified organisms Carbon capture and utilization	Potential for significantly higher efficiencies and lower environmental impacts. Tailored production of specific high-value chemicals and fuels.	Ethical and regulatory concerns regarding GMOs. High initial research and development costs.

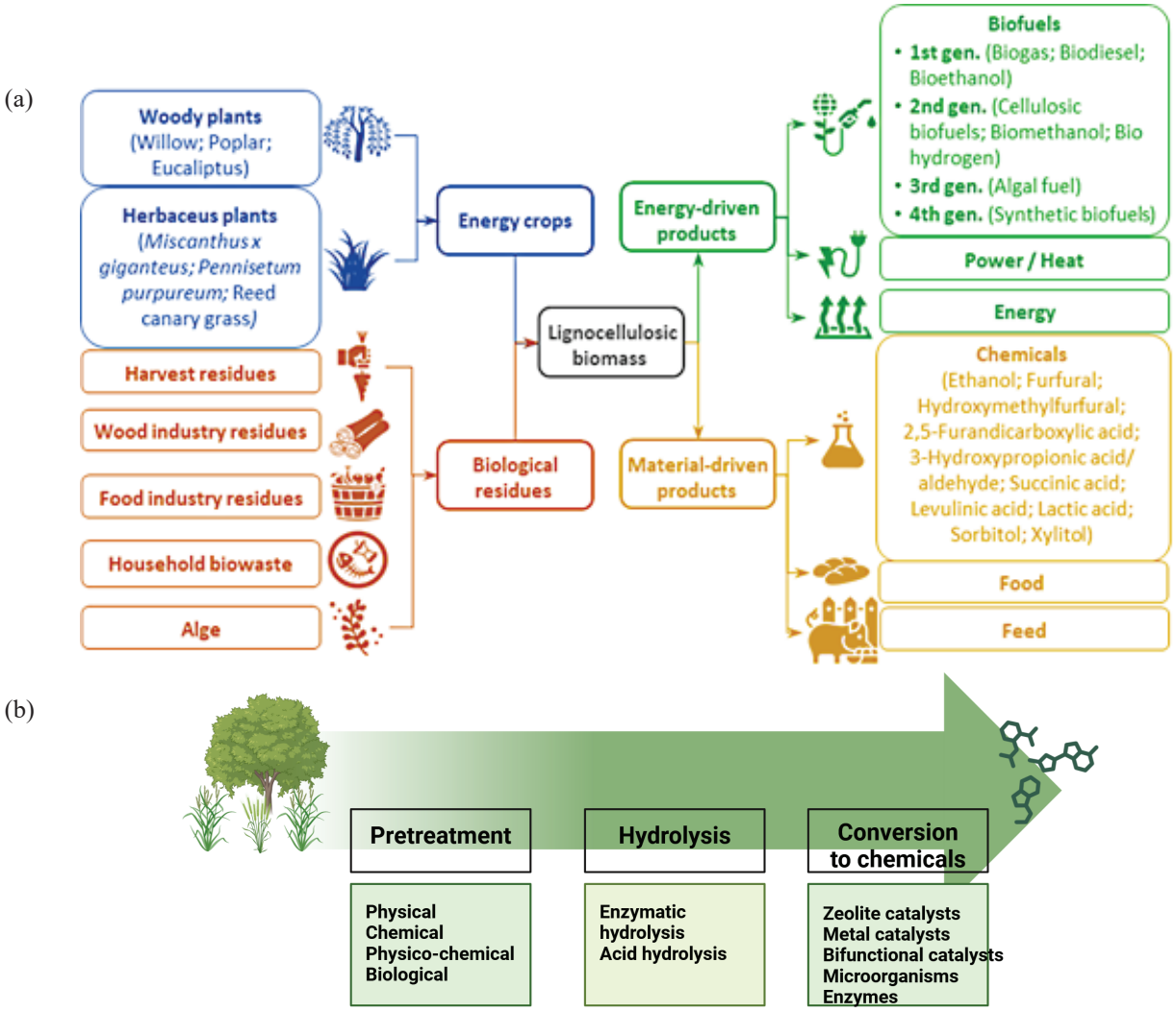

Fig. 1 – *(a) Biorefinery concept: Types of lignocellulosic biomass (LB) used for the production of a variety of energy- and material-driven products (b) Steps of conversion of lignocellulosic biomass into high-value chemicals*

Table 2 – *Thermochemical process used in biomass utilization*[29–33]

Process	Mechanism	Product	Advantages	Disadvantages
Torrefaction	Involves heating biomass in an inert or low-oxygen environment to temperatures typically ranging from 200 °C to 300 °C.	Torrefied biomass or biocoal	Improved energy density Biomass hydrophobicity Uniformity Reduced transportation costs	Energy consumption Capital and operational costs
Pyrolysis	Thermal degradation of various biomass (either a solid or liquid) in the absence of oxygen, producing bio-oil, biochar, and gases as fuels in a sustainable and green manner. Is usually conducted under elevated temperatures ranging from 400 to 600 °C.	Biochar Bio-oil Syngas	Versatility in products Waste management Carbon sequestration Reduction of emissions	High capital and operational costs Complexity in process control Bio-oil stability
Gasification	A process that converts organic materials (biomass) into a combustible gas mixture called syngas (synthesis gas). This process occurs at high temperatures (typically 700–1000 °C) in an environment with a controlled amount of oxygen, steam, or air, which is insufficient for complete combustion.	The syngas produced primarily consists of carbon monoxide (CO), hydrogen (H_2), carbon dioxide (CO_2), methane (CH_4), and other trace gases	High energy efficiency Versatile syngas applications Renewable energy source Waste reduction	High capital costs Complex process control Feedstock variability Limited scalability
Liquefaction	A process that converts solid biomass into liquid fuels and chemicals through thermochemical or biochemical methods. Two primary methods for biomass liquefaction are: hydrothermal liquefaction (HTL), and fast pyrolysis.	Biofuels or as intermediates for producing higher-value chemicals	High energy density products Efficient conversion Compatibility with existing infrastructure Valorization of waste	High capital and operational costs Complex process Environmental concerns Hydrogen and catalysts
Carbonization	Converts organic materials into carbon-rich solids, primarily biochar, through the application of heat in a low-oxygen or oxygen-free environment. The process involves heating biomass to temperatures typically between 300 °C and 700 °C, resulting in the thermal decomposition of organic materials.	Biochar	Biochar production Waste management Energy production Soil health improvement	High energy input Capital and operational costs By-product management Scale-up challenges Potential release of pollutants

cludes varying proportions of cellulose, hemicellulose, and lignin, determines the suitability of specific biomass types for particular chemical pathways. For instance, hardwoods, softwoods, and agricultural residues like corn stover or wheat straw exhibit distinct structural and chemical characteristics that affect their conversion efficiency[44,45]. Hardwoods, which typically have higher cellulose content, are more suitable for bioethanol production, while agricultural residues, rich in hemicellulose, are more suited for producing bio-based chemicals such as xylitol or furfural[46,47]. Moreover, the inherent variability in feedstocks affects both the yield and quality of the final products. Consistency in feedstock quality ensures predictable processing outcomes, which is essential for industrial-scale operations. Variations in moisture content, ash content, and the presence of inhibitory compounds can complicate the pretreatment and conversion processes, leading to inefficiencies and reduced product quality[48,49]. Selecting feedstocks that are locally available and abundant can reduce transportation costs and carbon footprint, enhancing the overall sustainability of the biomass conversion process. The economic viability of LB conversion also depends on the cost and availability of the chosen feedstock. Agricultural residues and forestry by-products are often considered waste materials and can be sourced at lower costs compared to dedicated energy crops. However, the seasonal availability of some feedstocks necessitates robust supply chain management to ensure a consistent input to biorefineries[50,51]. In summary, selecting the appropriate feedstock is fundamental to optimizing the yield, quality, and economic feasibility of chemicals produced from LB. A thorough understanding of the chemical composi-

tion and availability of different biomass types allows for tailored processing strategies that maximize efficiency while minimizing environmental impact and costs[52].

Deciding on the location for biorefineries is crucial as it significantly influences both investment and operating costs. The location also has environmental implications, particularly through transportation and logistics activities for supply chain sourcing. Once a biorefinery's location is established, the supply chain must be designed[53], including decisions on which crops to harvest, when to harvest them, and how to transport them (whether crops or residues) to the biorefinery. For economic feasibility, biorefineries must fully utilize the LB. This means adopting strategies that use all the building blocks of lignocellulose for the production of bio-based products. Additionally, the waste streams generated during the pretreatment and fractionation of LB, as well as during the extraction, isolation, and catalytic or biocatalytic production of the products from lignin, cellulose, or hemicellulose, should also be effectively utilized. The optimal approach involves the cascade utilization of lignocellulose[54].

For instance, combining thermal and chemical pretreatment methods can result in the formation of highly toxic chemical compounds, which inhibit the microorganisms used in further microbial conversion of pretreated biomass into products[55]. Guo *et al*.[56] summarized various pretreatment methods, detailing the inhibitors, their mechanisms of action, and their removal, along with information on their sources, whether lignin, hemicellulose or cellulose. When selecting an appropriate pretreatment method, considerations should include sustainability, energy consumption, investment costs, and the overall efficiency of the processes[9,40]. The selection process is further complicated by the difficulty in assessing the efficiency of various options due to limited data representation[39]. Another challenge lies in choosing upstream methods, such as the extraction and/or isolation of desired polymers, as well as selecting methods (catalytic, biocatalytic, or microbial using advanced metabolic engineering technologies, etc.) to convert the process residues into products[3,7,57,58].

There remain significant technological and economic challenges in utilizing LB to produce chemicals with high selectivity and yield. It is often suggested, for economic reasons, to integrate bioenergy production with the production of chemicals and/or other bio-based products[23]. Sustainability in integrated LBRs requires developing new technologies or improving existing ones for each step of the process. When producing chemicals from LB, tradeoffs between yield and quality involve a delicate balance of feedstock selection, pretreatment, enzymatic hydrolysis, fermentation, catalytic conversion, and downstream processing[59–61]. Optimizing these factors requires a holistic approach that considers both technical and economic aspects to achieve a feasible and sustainable process. Fig. 2 provides a systematic overview of the tradeoffs between yield and quality in the production of chemicals from LB. Given the substantial amount of data generated by automation and computerization in this field, there is a pressing need for rapid and effective data analysis.

Machine learning in lignocellulose biorefineries

Basic concept of machine learning

According to Schmidt *et al*.[62], the advancement of information technology has enabled the efficient implementation of ML algorithms for tasks such as classification, regression, clustering, or dimensionality reduction of high-dimensional input data. Coupling ML algorithms with specific process or phenomena databases has shown considerable promise in identifying complex relationships implicit in the data[63].

The term ML refers to a subset of artificial intelligence that uses a series of methodologies or algorithms that enable computers to automate data-driven model development by systematically discovering patterns in statistically significant data[64,65]. Artificial intelligence broadly encompasses technologies that mimic cognitive processes such as learning and problem solving, often associated with human intelligence. ML focuses on algorithms that improve their performance as they are exposed to more data, while deep learning, a subset of ML, uses multi-layered neural networks to learn from large datasets.

According to Basha and Rajput[66], and Wang *et al*.[67], solving problems using ML involves three key steps: (i) framing the learning problem in an algorithm that the computer can process, (ii) choosing an evaluation method to assess the quality or accuracy of the ML system's predictions, often a classifier, and (iii) optimizing the process. ML algorithms (Fig. 3) are classified into taxonomies based on the algorithm's expected output[68]: (a) Supervised learning, (b) Unsupervised learning, (c) Semi-supervised learning, (d) Reinforcement learning, and (e) Transduction. Commonly applied ML methods include K-Nearest Neighbours (K-NN), Naïve Bayes Classification (NBC), Decision Tree (DT) Classification, Random Forest (RF), Gradient-Boosted Decision Trees (GBDT), Support Vector Machines (SVMs), and Artificial Neural Networks (ANN)[69,70].

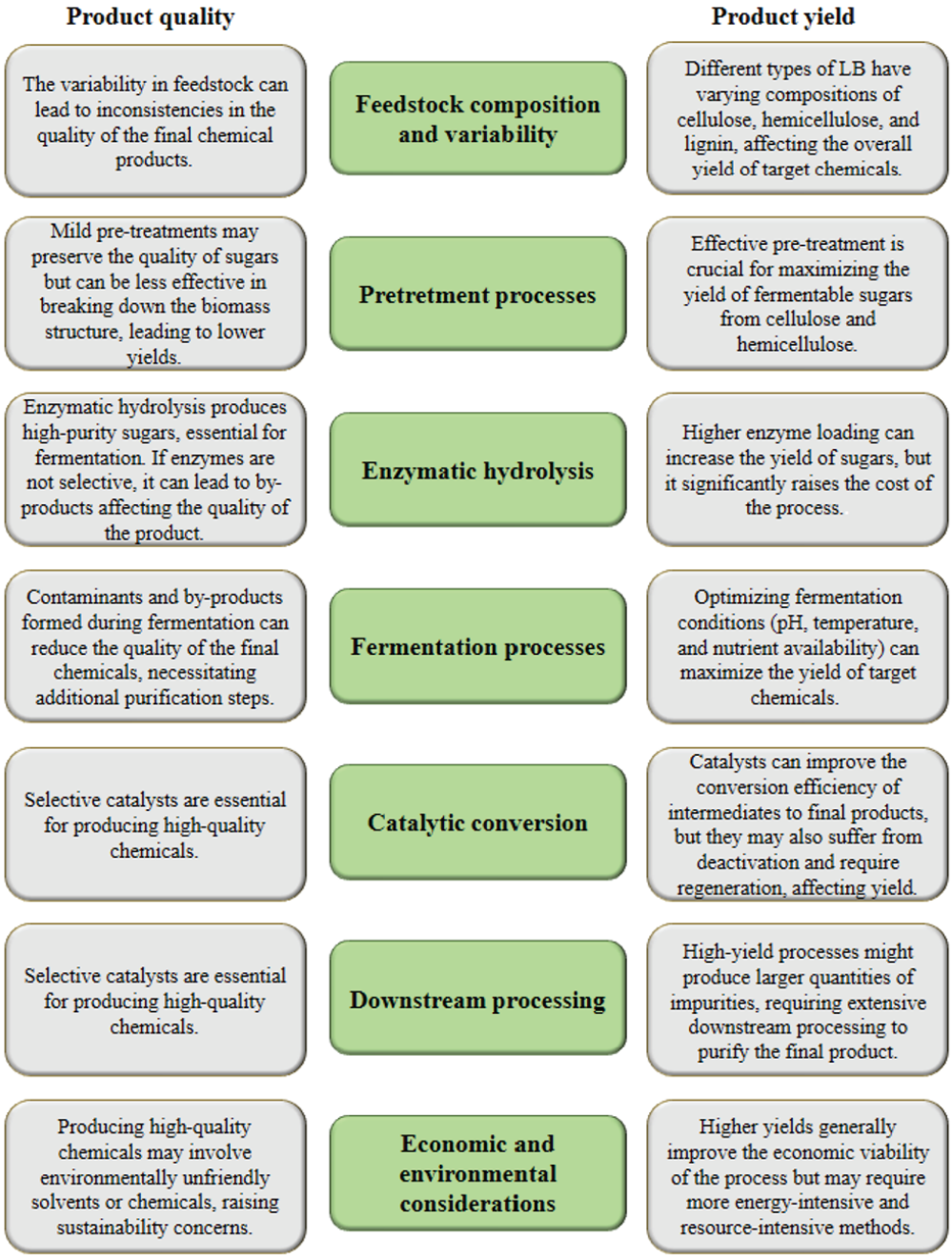

Fig. 2 – *Systematic overview of the tradeoffs between yield and quality in the production of chemicals from LB*[59–61]

The K-NN algorithm categorizes components based on the closest training samples in the feature space[71,72]. K-NN is an instance-based learning or lazy-learning method, where the function is estimated locally, and full computation is delayed until classification. When there is little deep information about the data distribution, K-NN is the most basic and simplest classification algorithm[71,72]. Because of its simplicity, ease of implementation, and effectiveness, K-NN is a widely used classification technique. It is one of the top 10 data mining algorithms and is widely used in a variety of sectors.

NBC is a classification algorithm that predicts the most likely class by combining frequency and value combinations in the given dataset[73,74]. Determining which domain knowledge is beneficial for selecting information features in model data classification involves several steps and strategies. The goal is to leverage domain expertise to identify features that are most likely to improve model performance while avoiding irrelevant or misleading information[75]. For example, Exploratory Data Analysis (EDA)[76] can be used to understand the distribution, relationships, and patterns in the data. Furthermore,

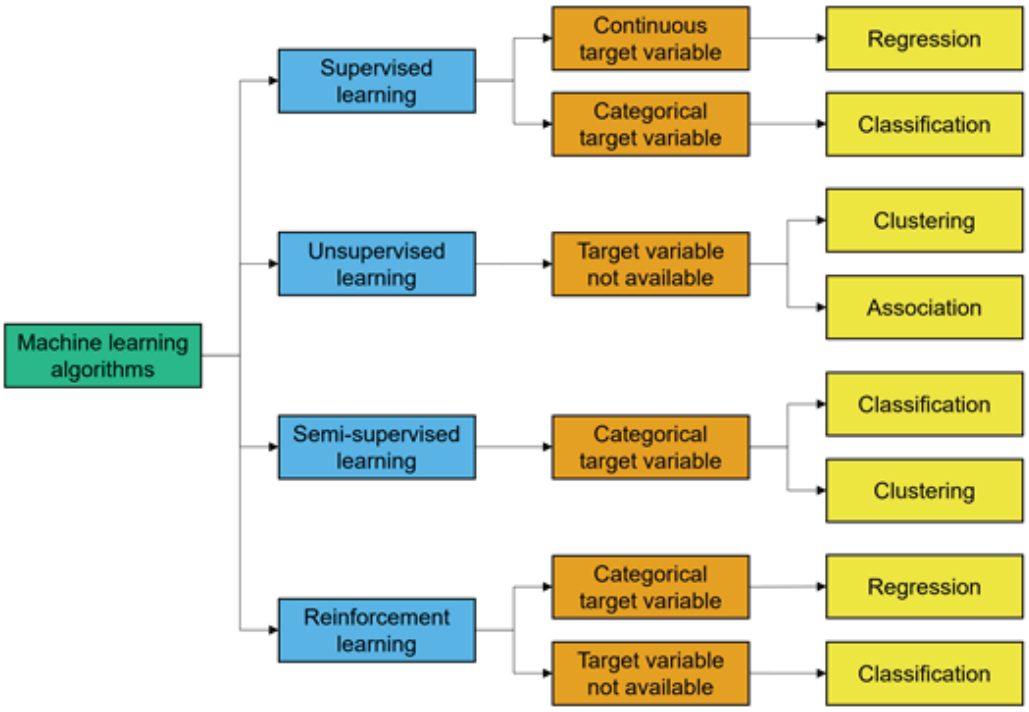

Fig. 3 – *Machine-learning algorithms classification*

correlation analysis[77] calculates correlation coefficients to identify relationships between features and the target variable, while filter methods use statistical techniques[78] (e.g., chi-square test, mutual information) to select features that have a strong relationship with the target variable. DTs are among the most straightforward methods for data categorization because of their ease of interpretation and application. DTs split the input space into sub-spaces associated with a classifier and can be used to represent numerical (ordered) or categorical (unordered) properties[79]. RF may be applied to classification and regression in ML. By using RF, the algorithm predicts a value or category through combining results from a number of decision trees. It uses ensemble learning, the process of merging numerous classifiers to solve a complicated issue and enhance the model's performance[80]. GBDT involves converting a straightforward cut-based analysis into a multivariate method. The idea behind a GBDT is to not dismiss events that fail a criterion right away, but rather to see whether other criteria may assist to appropriately classify these occurrences[81]. SVM may be employed for either regression or classification applications. The SVM algorithm's objective is to achieve a hyperplane in an N-dimensional space (N = the number of characteristics) that effectively classifies the data points[82,83]. ANN are nonlinear models inspired by biological neural networks[84]. ANNs consist of interconnected neurons, where each neuron functions as a parallelized processor capable of extensive data processing and classification[85]. In the same way that mammalian neurons learn from historical events and failures to attain goal outcomes, the structure of an ANN is altered depending on the data presented within the learning process. The ANN structure includes input, hidden, and output layers. Artificial neurons receive inputs and generate outputs that may be sent to numerous other neurons[86]. Over the years, specialized ANN architectures have emerged. Recurrent Neural Networks (RNN)[87] such as Long Short-Term Memory (LSTM)[88], for example, are particularly well-suited for processing sequential information. In contrast, applications in the areas of image processing often draw upon Convolutional Neural Networks (CNN)[89] since they excel in recognizing patterns within the input data.

CNNs have also been used to extend natural language processing models and speech processing[90]. Transformer-based architectures such as Bidirectional Encoder Representations from Transformers (BERT)[91], DistilBERT[92], and Generative Pretrained Transformer 3 (GPT-3)[93] allow unsupervised pretraining on massive datasets, reducing the data required for supervised task-specific fine-tuning. They also consume the entire input sequence at once, and use an attention mechanism to contextualize positions within the input sequence. Transformers, therefore, address major shortcomings of RNN architectures, such as lower parallelization due to dependencies between computation steps, and the vanishing gradient problem.

The development of an ML model typically includes several phases, as shown in Fig. 4. The first phase, "Data Analysis" involves problem identification, data collection, and feature analysis. This first phase is crucial as the quality of the model's output depends on the input data[18]. It includes identifying missing values, data outliers, converting categorical data into numeric data, and selecting suitable properties that contribute to model accuracy. The second phase, "Building the ML Model", involves dividing the data into training and testing sets, selecting an appropriate ML algorithm, and developing the ML model based on the training dataset. The third phase, "Tuning Model Parameters", includes evaluating the model's applicability and hyperparameter tuning. It is important to mention that in the third step, feature engineering is not required for deep learning models. The process of selecting a set of ideal hyperparameters for a learning algorithm is known as hyperparameter tuning. A hyperparameter is a model parameter whose value is determined prior to beginning the learning process. The fourth phase of ML model development involves model selection based on performance metrics, and if the performance is not satisfactory, the process of building and tuning the model is repeated. Ensuring the accuracy and representativeness of a generalized machine learning model, while avoiding overfitting, involves a combination of good practices in model validation, evaluation, and monitoring[94,95]. Various techniques can be used to achieve this.

For example, k-fold cross-validation divides the dataset into k subsets. The model is trained k times, each time using k-1 subsets for training and the remaining subset for validation. This process ensures that every data point is used for both training and validation, providing a robust estimate of model performance[96].

Holdout validation[97] splits the dataset into three parts: training, validation, and test sets. The training set is used to build the model, the validation set is used to tune hyperparameters and for initial model evaluation, and the test set is used for the final evaluation of the model's performance. The test set should not be used during model training or hyperparameter tuning.

Hyperparameter tuning is used to optimize the hyperparameters using techniques like grid search, random search, or Bayesian optimization, ideally in conjunction with cross-validation to avoid overfitting on the validation set[98].

Application of machine learning methods in biotechnology

Advancements in biotechnology increasingly rely on the significant use of big data gathered through sophisticated analytical equipment[99–101]. Coupling ML algorithms with available databases has shown considerable potential in identifying complex connections in biotechnological processes[102]. As extensively described by Mowbray *et al.*[63], ML algorithms have applications in bioreactor engineering, optimization and control of microbial fuel cells, development of microfluidic and soft sensors, development of new biomaterials, modeling and optimization of biofuels and bioenergy processes, environmental engineering, metabolic engineering, and cell culture and protein engineering.

This review presents ML methods and their efficiency, used in metabolic engineering (Table 3), bioprocess development (Table 4), and environmental engineering (Table 5).

Due to their structure, ANNs are suitable for modeling highly nonlinear processes with even a small training dataset. Zhou *et al.*[103] developed an ANN model to optimize heterologous β-carotene and violacein biosynthesis pathways in *Saccharomyces cerevisiae*, implementing the examination of a small fraction of combinatorial space to accurately adjust the expression level of each gene in the analyzed pathway. Similarly, Zhang *et al.*[104] used genotype data, growth profiles, and biosensor data to optimize the aromatic amino acid tryptophan pathway in *S. cerevisiae* with the Automated Recommendation Tool (ART) and the EVOLVE algorithm. Both methods were able to describe the train-

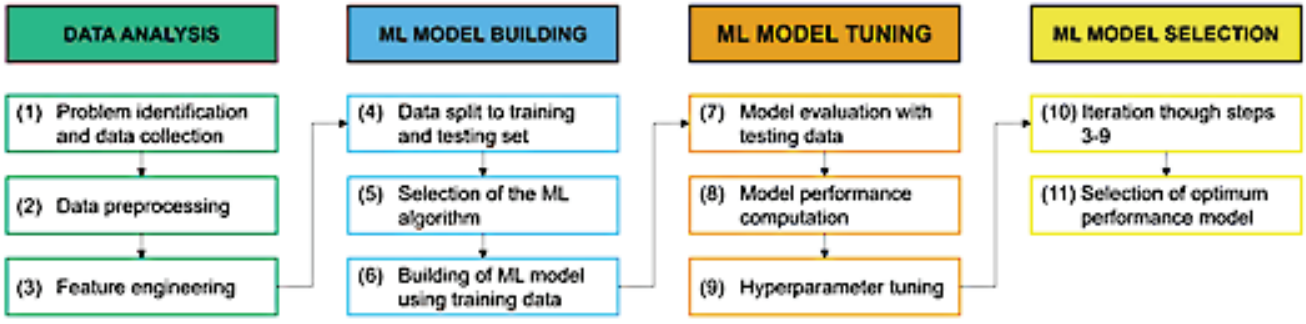

Fig. 4 – *Development of ML models (adopted from[99,100])*

Table 3 – *Application of machine-learning methods in metabolic engineering*

	ML method*	Application	ML method efficiency	Reference
1	ANN model coupled with YeastFAb	Optimization of beta carotene biosynthetic pathway in *S. cerevisiae* to produce violacein.	Model was developed and verified for the existence of a strain that showed a 2.42-fold titer improvement in violacein production among 3125 possible designs.	103
2	FNN	Effects of culture media on gene regulation in *E. coli*.	Model predicts the effects of culture media (up to 10 nutrients) on gene regulation with accuracy of 0.867.	113
3	GNN	Prediction of enzyme-substrate pairs.	Model ensured enzyme-substrate pared selection accuracy above 91 %.	114
4	LG, RF, gradient-boosted DT, and SVM	Prediction of enzyme activity for a set of bacterial nitrilases based on docking descriptors.	Each of the ML models performed similarly (average ROC = 0.9, average accuracy = ~82 %).	115
5	RF	Modeling of the human mesenchymal stromal cells expansion by predicting the population doubling time based on individual donor characteristics.	Mean absolute prediction errors range from 0.8 to 0.99.	116
6	NB, K-NN, QDA, DT, AB, RF, and NN	Automatic classification of living and dead microalgae based on the multispectral fluorescence microscopy images.	RF model classifies the cells with 82 % accuracy.	117
7	LRM, RF, scalable TBS, NN, K-NN, and SVM with radial kernel	Classification of feed substrates (acetate, carbohydrates and wastewater) in microbial fluid cells based on genomic data divided into four groups.	NN model ensured data grouping with accuracy around 93 %.	118
8	MPLS and evolving PLS model	Prediction of end-point concentration of the monoclonal antibodies in CHO cell cultures based on metabolomics data of the culture.	PLS model with variables selection predicts the end point concentration of mAb with R^2 greater than 0.9.	119
9	TBDN	Identification of small molecules for their mass spectrometric spectra.	53 % precision for molecules with a molecular mass of 500 Da or lower.	120
10	TNN	Protein-specific *de novo* molecule design.	Efficiency of the proposed method in terms of predicted binding affinity of generated ligands to the target protein, percentages of valid diverse structure, drug-likeness metrics and synthetic accessibility.	121

*Adaptive Boosting (AB); Artificial Neural Network (ANN); Decision Tree (DT); Feedforward Neural Network (FNN); Graph Neural Networks (GNN); K-Nearest Neighbors Algorithm (K-NN); Logistic Regression (LG); Logistic Regression Multiclass (LRM); Multiple Partial Least Squares Regression (MPLS); Naive Bayes (NB); Neural Network (NN); Partial Least Squares Regression (PLS); Quadratic Discriminant Analysis (QDA); Random Forest (RF); Support Vector Machine (SVM); Transformer Neural Network Approach (TNN); Transformer-Based Deep Neural Network (TBDN); Tree Boosting System (TBS).

ing set data with high precision. The authors reported that the ML tools enabled the engineering of the complex aromatic amino acid metabolism in yeast with a tryptophan titer increase of up to 74 % and a process productivity increase of up to 43 %.

However, despite the growing number of studies highlighting the advantages of ML, utilizing the data for further analysis can be challenging due to experimental conditions and the lack of metadata[105]. Tulsyan *et al.*[106] presented an efficient Bayesian non-parametric approach for modeling and optimizing biopharmaceutical batch processes with limited data. Onel *et al.*[107] applied Support Vector Machine-based selection algorithm to monitor the penicillin production process based on big data gathered through online measurements (fermentation volume, dissolved O_2 concentration, dissolved CO_2 concentration, temperature in reactor, pH, feed rate, feed temperature, agitator power, cooling/heating medium flow rate, heating medium temperate, hot/

Table 4 – *Application of machine learning in bioprocess development*

	ML method*	Application	ML method efficiency	Reference
1	ANN	Optimization of acetic acid production considering glucose and ethanol concentrations and incubation period.	Optimal process conditions yield 4.88 g/100 mL of acetic acid.	[122]
2	RF, Xgboost, K-NN, and ANN	Prediction of medium-chain carboxylic acid concentration and production rate based on environmental and operational parameters and genomic data.	RF model had the highest prediction accuracy of 0.83, 0.87, and 0.89 when the operational parameters, genomic data, and combined dataset were used as input parameters, respectively.	[123]
3	ELM, ANN, and RF	Analysis and simulation of anaerobic digestion of dry fermentation.	ELM best predicted biogas production with R^2 of 0.9574 and MAE of 0.678.	[124]
4	Partly supervised reinforcement learning algorithm	Control of semi-continuous process of yeast fermentation (Input: substrate concentration in feed and the dilution rate; output: biomass and substrate concentration).	Mean square error of 1.5745 for biomass concentration prediction.	[125]
5	Asynchronous advantage actor-critic algorithm belonging to deep reinforcement learning strategy	Fed-batch control of the bioreactor for the production of cyanobacterial phycocyanin.	52.1 % increase in product yield and 20.1 % increase in cyanobacterial-phycocyanin concentration compared to reactor without control.	[126]
6	GBRM	Evaluation of *Candida antarctica* growth kinetics for lipase production.	Reduction in time and resources by cca. 50 %	[127]
7	PCA and PLS	Detection and quantification of harmful cyanobacterial species in microalgal bioreactor based on Low Resolution Raman Spectra.	Detection and quantification of algal and cyanobacterial species at concentrations of 103 cells/mL.	[128]
8	PCA, PLS, NN, RF, SVM, and CR	Prediction of polyhidroxybutirate (PHB) concentration produced by *Cupriavidus necator* based on near-infrared spectra (NIR).	PLS performance showed the best prediction based on the raw spectra (R^2 0.66 and RMSEP 0.38 g L^{-1}).	[129]
9	SVM	Identification of the flow regime in a bubble column reactor based on the data collected by optical probe (Inputs: dimensionless variance of bubble time and time scale).	Linear SVM model predicted the flow regime with 98.1 % accuracy.	[130]
10	K-NN	Optimisation of mycophenolic acid production with *Penicillium brevicompactum*.	1.64-fold higher production efficiency.	[131]
11	MLR, MLP, and SMO	Modeling of ultrasound-mediated *Escherichia coli* cell disruption (Inputs: cell mass concentration, sonication time, duty cycle and acoustic power; Output: release of nitrilase and cytosolic proteins and extent of cell disruption.	Maximum cell disruption of 92 % was achieved under optimized process conditions.	[132]
12	DT classification	Optimization of algal biomass productivity and lipid content.	DT method detected 11 combinations of input variables contributing to higher production yield.	[133]

*Artificial Neural Network (ANN); Cubist Regression (CR); Decision Tree (DT); Extreme Gradient Boosting (XGBoost); Extreme Learning Machine (ELM); Gradient Boosting Regression Model (GBRM); K-Nearest Neighbor (K-NN); Multi-layer Perceptron (MLP); Multiple Linear Regression (MLR), Neural Network (NN), Partial Least Squares Regression (PLS); Principal Component Analysis (PCA); Random Forest (RF); Sequential Minimal Optimization (SMO); Support Vector Machine (SVM).

Table 5 – *Application of machine learning in environmental engineering (waste and wastewater treatment)*

	ML method*	Application	ML method efficiency	Reference
1	Eight ML models	Simulation of micropollutants' behaviour in forward osmosis.	ANFIS in forecasting micropollutants removal (R = 0.99; RMSE = 0.56 %).	134
2	CNN	Development of the multivariable identification model based on a compacted cascade neural network to identify membrane fouling.	Models described membrane permeability, integrity, and life with accuracy > 99 %.	135
3	GA	Prediction of effluent concentrations and biogas production in wastewater treatment processes.	Models described effluent and biogas concentrations with accuracy > 99 %.	136
4	Model based on HGSA and ANFIS	Simulation of electrochemical oxidation process (reaction time, pH, salt concentration, and voltage).	ANFIS model described COD and TOC removal efficiency with RMSE 2.63.	137
5	Feed-forward back-propagation-based ANN	Selenite removal based on influent selenite concentration and hydraulic retention time in fungal pelleted airlift bioreactor for wastewater treatment.	ANN model with high values of the correlation coefficient ($0.96 \leq R \leq 0.98$), low root mean square error ($1.72 \leq$ RMSE ≤ 2.81), mean absolute percentage error ($1.67 \leq$ MAPE ≤ 2.67).	138
6	DCCN	Modeling and prediction of real municipal wastewater treatment in anaerobic membrane bioreactors	Accuracy rate of up to 97.44 %.	139
7	MLP NNM	Prediction of ammonium, total phosphorus and total nitrogen removal in anaerobic-anoxic-oxic reactor.	R^2 for ammonium removal were in the range from 0.763 to 0.876.	140
8	RNN	Simulation of the long-term operation of an osmotic membrane bioreactor-clarifier.	RMSE of 3 % and 2 % of the average conductivity values for OMBR and OMBRC, respectively.	141
9	RNN	Early prediction of osmotic membrane bioreactor system to reduce the environmental impacts of wastewater.	R^2 of 0.92 and 0.93 for the prediction of water flux and membrane fouling simulations.	142
10	Combination of ANN, ANFIS, and SVR	Prediction of process parameters that describe biomass properties, operation parameters, and effluent properties of the biological nutrients removal wastewater treatment process.	Average correlation coefficient for the model outputs was 69 %.	143
11	NNM	Prediction of nitrogen removal rate in multi-soil-layering bioreactor at different loading rates.	R^2 for nitrogen removal rate was greater than 0.93.	144
12	Hybrid ML method based on RF and short-term MA	Redefending the key factors and improving the production of output data form the biogas plant.	20 % increased accuracy compared to traditional analytical models.	145
13	GBM, SVM, RF, AdaBoost	Modeling of hydrogen production from wastewater during dark fermentation based on Fe concentration.	R^2 for used methods were 0.893, 0.885, 0.902 and 0.889.	146, 147
14	SARIMAX, RF, SVM, GTB, ANFIS, LSTM	Prediction of total phosphorus concentration at the outflow from the wastewater treatment plant based on real-time gathered data (24 h x 365 days) for 10 wastewater parameters.	SARIMAX showed the lowest mean square error and the highest coefficient of determination.	148
15	ANN, SVR, ANFIS	Prediction of the aerobic granulated sludge process performance based on the lab-scale reactors gathered data.	Model predicted process performance with R^2 = 95.7 %.	143
17	DNN, RTRF, ABR, and GBR	Modeling efficiency of membrane bioreactor for wastewater treatment.	The highest R^2 (0.847, 0.792 and 0.851) were obtained for GBR.	149
18	Q-LA	Optimization of the hydraulic retention times in anaerobic and aerobic reactions of biological phosphorous removal.	Stable output concentrations at optimum process conditions.	150

*Adaptive Boosting Regression (ABR); Adaptive Neuro-Fuzzy Inference System (ANFIS); Artificial Neural Networks (ANN); Cascade Neural Network (CNN); Densely Connected Convolutional Network (DCCN); Deep Neural Network (DNN); Genetic Algorithm (GA); Gradient Boosting Machine (GBM); Gradient Boosting Regression (GBR); Gradient Tree Boosting (GTB); Hunger Games Search Algorithm (HGSA); Long Short-Term Memory (LSTM); Multilayer Perceptron (MLP); Neural Network Model (NNM); Q-Learning Algorithm (Q-LA); Random Forest (RF); Recurrent Neural Network (RNN); Regression Tree (RT); Seasonal Autoregressive Integrated Moving Average (SARIMAX), Short-Term Memory Analysis (Short-Term MA); Support Vector Machine (SVM); Support Vector Regression (SVR).

cold switch, base flow rate and acid flow rate). The fault and time-specific models were trained utilizing three separate period methods, and obtained results showing that the suggested methodology possessed potential for online diagnosis in batch operating processes.

In environmental engineering, ML methods are used for modeling and optimizing membrane bioreactors[108,109]. When working with real-time applications of ML, i.e., membrane bioreactor technology, membrane fouling presents the biggest challenge, therefore significant efforts have been made in developing reliable and accurate membrane fouling models that can be used on the industrial scale. Hamedi *et al*.[109] demonstrated the potential of artificial neural networks (ANNs), gene expression programming (GEP), and least square support vector machine (LSSVM) modeling for prediction of the fouling resistance. The obtained results showed that the LSSVM model was the most suitable for the prediction of fouling resistance with the lowest mean squared error (0.0002), maximum absolute percentage error (3.18), minimum absolute percentage error (0.01), and the highest coefficient of determination (0.99). Chen *et al*.[108] developed radial basis function ANN models for prediction of interfacial energy related to membrane fouling based on contact angles of three probe liquids, zeta potential, and separation distance achieving high accuracy (R^2 > 0.99). Hazrati *et al*.[110] used feed-forward ANN with a back propagation algorithm for predicting the chemical oxygen demand (COD) and transmembrane pressure in petrochemical wastewater treatment in a membrane bioreactor. Mixed liquor suspended solids, hydraulic retention time, and time were used as the modeling inputs, achieving a total correlation coefficient of 0.999, demonstrating that ANNs possess great potential for including multiple variables in highly nonlinear interactions, such as those occurring in membrane bioreactors during wastewater treatment. Bioreactors are also commonly used for the treatment of waste gases, and ML methodologies are employed for optimizing these processes. For instance, Baskaran *et al*.[111] developed an efficient ANN model (R^2 = 0.992) for predicting the performance of an airlift bioreactor used in the biological removal of trichloroethylene by *Rhodococcus opacus.* Similarly, Baskaran *et al*.[112] employed an ANN model (R^2 = 0.992) for predicting the efficiency of a continuous stirred tank bioreactor in the biological removal of trichloroethylene by *Rhodococcus opacus*.

Application of machine learning in lignocellulose biorefineries

According to Wang *et al*.[151], ML can be used for classification, regression, and optimization tasks within biorefinery operations. Ge *et al*.[15] presented a comprehensive overview of various ML techniques for analyzing LB, highlighting their use cases, including advantages and challenges in fields such as high-value utilization transformation, chemical characterization, process simulation, and the preparation of functional materials from lignocellulose. Ascher *et al*.[146] stressed the importance of considering correlated features to avoid losing critical variables, noting that both numerical and categorical variables are essential. Reshmy *et al*.[3] highlighted the importance of fine-tuning biomass pretreatment strategies to achieve optimal techno-economic feasibility. In general, ML tools have shown significant potential in predicting the chemical composition of LB, analyzing and optimizing LB pretreatment and treatment processes, and in analyzing and optimizing the biofuel production process. This review presents various ML methods, including their efficiency that are used for determining chemical composition, and optimizing pretreatment, treatment and biofuel production from LB (Table 6).

While the predictive performance of most ML models are generally accurate (R^2 > 0.85), interpreting and comparing various models remains challenging, with no single model being universally satisfactory. The data presented suggests that ANN modelling is the most common ML method used in various stages of LBRs. The application of ANN and ANFIS in modeling LCB valorization processes has been reviewed by Pradhan *et al*.[152] The nonlinear and adaptive nature of ML methods makes them well-suited for handling large databases. Recent developments show that novel and hybrid machine-learning algorithms are continuously being developed and refined to further improve predictive efficiency. Table 7 provides a brief overview of the data matrices used in ML modeling within lignocellulose biorefineries, illustrating their size and complexity.

Conclusions and future perspectives

Machine learning (ML) offers numerous advantages across various applications, including lignocellulosic biorefineries (LBRs). ML can identify patterns within data, forecast future trends, automatically create new features, cluster data, and routinely detect outliers, all of which are essential for enhancing the efficiency and productivity of LBR processes. In the context of LBRs, the application of ML is crucial due to the multitude of variables that affect process efficiency. Highly nonlinear mathematical models are required for accurate description, prediction, in-silico analysis, and optimization. ML ensures a thorough analysis of large datasets, although a comprehensive experimental dataset

Table 6 – *Application of machine-learning methods in lignocellulosic biorefinery*

	Step within lignocellulosic biorefinery	ML method*	ML method efficiency	Reference
1	*Chemical analysis:* Non-invasive analysis of cellulose, hemicellulose, and lignin concentrations in corn stover	CNN models with Bayesian training algorithm based on the FTIR absorption peak parameters.	CNN models with R^2 = 0.98–0.99 and RMSE of 0.11.	153
2	*Chemical analysis:* Rapid analysis of cellulose and lignin concentrations in 30 crops	BiPLS with PCA and SVM for establishing concentrations based on NIR spectra.	Models with R^2 = 0.91–0.99.	154
3	*Chemical analysis:* Analysis of cellulose, lignin, pentosane, and holocellulose concentrations in jute fibers	ANN models for predicting chemical composition of jute fibers based on FT-NIR spectra.	Prediction efficiency of ANN varied from 72–99 % for calibration, validation, and test datasets.	155
4	*Chemical analysis:* Analysis of cellulose, hemicellulose, and lignin in 178 samples of LB	ANN models for predicting cellulose, hemicellulose, and lignin concentration of LB based on proximate analysis data.	ANN model predicted all major biochemical components with R^2> 0.96.	156
5	*Chemical analysis:* Analysis of hemicellulose and lignin in rice husk, redwood, pine wood, rubber wood biomass	PCA, PLS-DA, SVM, RBFNN, and ELM to quantitatively distinguish biomass pellets based on laser-induced breakdown spectroscopy data.	RBFNN model showed a 100 % average recognition accuracy in calibration and 96.8 % average recognition accuracy in prediction sets, respectively.	157
6	*Chemical analysis:* Estimation of holocellulose content	RR, LR, and ENR, classical ML algorithms (SVR, DT, and RF), and advanced GBM algorithms (LightGBM, CatBoost, and XGBoost) to build the holocellulose content based on the Raman spectra.	CatBoost and XGBoost could predict holocellulose content with high accuracy of test R^2 above 0.93 and test RMSE less than 0.29 %.	158
7	*Pretreatment:* Pretreatment of mixed vegetable waste by organic and inorganic acid	ANN model for optimization of the mixed vegetable pretreatment.	ANN model with R^2=0.997 and adjusted correlation coefficient (R^2=0.987).	159
8	*Pretreatment:* Chemical pretreatment of cassava peels	ANN, RF and DT to model fermentable sugar concentration and combined severity factor index from pretreated starch-based LB-cassava peels.	For the fermentable sugar concentration R^2>0.99 was achieved with DT. For the CSF index, R^2>0.93 with ANN model.	160
9	*Pretreatment:* LB pretreatment with deep eutectic solvents	PCA, PLS, LG, OGB, ANN, and RF to analyze the mechanisms and interactions between LB and eutectic solvents.	The most significant effect on variations in lignin extraction were reaction temperature, hydrophilicity in the DES characteristic parameters, and the hemicellulose content of raw lignocellulose components.	161
10	*Pretreatment:* Enzymatic hydrolysis of rice straw	ANN and RSS for pretreatment and enzymatic hydrolysis of rice straw.	ANN model with R^2 > 0.99	162
11	*Pretreatment:* Hydrothermal pretreatment of lignin from side-stream waste	BO to relate hydrothermal pretreatment conditions with lignin structural characteristics based on 2D NMR.	ML model predicted lignin yield with 5 % error.	163
12	*Pretreatment:* Microwave-acid and enzymatic hydrolysis of LB	ANN and PSO for estimation of glucose and xylose yields.	Hybrid PSO-ANN model with R^2 > 0.99 for both glucose and xylose prediction.	164
13	*Pretreatment:* Enzymatic degradation of apple pomace using commercial enzymatic preparations	ANN model to predict release profiles of glucose and reducing sugars.	ANN model predicted output variables with R^2 > 0.99 for validation dataset.	165
14	*Pretreatment:* LB pretreatment with ionic liquid solvents	RF algorithm to predict cellulose-rich materials production.	RF predicted output variables with R^2 > 0.9 for validation dataset.	166

	Step within lignocellulosic biorefinery	ML method*	ML method efficiency	Reference
15	*Pretreatment:* Dilute inorganic acids hydrolysis of LB	ANN model for simultaneous prediction of the derived phenolic contents and glucose yield in corn stover hydrolysate before microbial fermentation.	ANN model predicted output variables with $R^2 > 0.9$ for validation dataset.	167
16	*Treatment:* Hydrothermal liquefaction of wet biomass	RF, k-NN and extreme GB for prediction of biocrude yields.	Extreme GB model gave the best prediction accuracy.	168
17	*Treatment:* Gasification of LB	ML methods to model biomass and waste gasification and predict syngas yield and composition, syngas lower heating value, and syngas tar content, as well as char yield.	$R^2 = 0.90$ when averaged across ten key gasification outputs.	146
18	*Treatment:* Hydrothermal liquefaction of LB	NNR, GAM, SVR, and GPR to model hydrothermal liquefaction products based on biomass composition and reaction conditions.	GPR ensured the highest accuracy, with a correlation coefficient higher than 0.926 and a mean absolute error lower than 0.031.	169
19	*Treatment:* Hydrothermal liquefaction of LB	SVMLK to evaluate the importance of hydration parameters on biocrudes production.	SVMLK efficiently predicted biocrudes composition.	170
20	*Treatment:* Pyrolysis of LB	RF, K-NN, DT, Gaussian Naïve Bayes, GB, and MPC for classification of LB.	The K-NN classifier performed the best for classifications using raw mass spectroscopy data.	171
21	*Treatment:* Pyrolysis – prediction of pyrolytic products yields	RF, GBDT, XGBoost, and Adaboost algorithms applied to predict bio-oil yield during pyrolysis based on moisture content, carbon content, and final heating temperature.	RF ensured the highest precision for bio-oil yield, biochar, and pyrolytic gas yields.	13
22	*Treatment:* Pyrolysis – prediction of pyrolytic products yields for 46 different types of biomass	MLR, DT, RF, SVM, and K-NN to develop predictive models for estimating biochar yield and specific surface area based on agricultural LB content data and pyrolysis conditions data.	RF algorithm ensured the highest precision biochar yields and specific surface area of the produced biochar.	172
23	*Treatment:* Pyrolysis – biomass pyrolysis kinetics	Activation energy of biomass pyrolysis calculated via three frequently used model-free kinetic methods were collected from literature and modeled by ANN, RF, and SVM.	Optimized RF model presented satisfactory accuracy and significant potential for making a quick prediction of activation energy with R^2 of 0.9110.	173
24	Biofuels production process from LB	New OERNN based prediction model for biofuel production prediction.	Model ensured biofuel production process enhancement by 50 %.	174
25	Bioethanol production process from sugarcane bagasse biomass	ANN, CT, RF to describe the effects of temperature, time, biomass, and inoculum size on ethanol fermentation by simultaneous hydrolysis and fermentation process.	ANN and RF models ensured R^2>0.91 for the validation dataset.	175
26	Bioethanol production from sugar cane lignocellulosic biomass	ANN model and PC Optimization algorithm to optimize industrial bioethanol production based on the dataset including 3400 experimental values.	ANN model predicted the bioethanol concentration at the end of the process with high accuracy.	176
27	Bioethanol production from LB	RF, EGB, and ANN to predict the ethanol yield and NaOH consumption based on characteristics of the biomass before and after pretreatment, hydrolysis parameters, and fermentation parameters.	Models described ethanol yield with accuracy of ~0.85 and NaOH consumption with accuracy greater than 0.8	177

	Step within lignocellulosic biorefinery	ML method*	ML method efficiency	Reference
28	Bioethanol production process based on different ionic liquid type, enzymatic preparation, enzyme dose, time and temperature of pretreatment, and type of yeast for fermentation	ANN and RF.	Hybrid model with R^2 of 0.96 for bioethanol concentration prediction.	[178]
29	Bioethanol production from LB based on the volatile composition of the LB	DNN.	A six-layer DNN ensured good accuracy for learning and validation.	[179]
30	Methane production efficiency from LB	10 ML methods.	The best model was KNN (R^2 >0.75) with the leave-one-out method.	[180]
31	Methane production efficiency from LB	RF, XGBoost and K-NN.	The best model was RF model for prediction of specific methane yields with R^2 of 0.85 and RMSE of 0.06.	[181]

*Adaptive Boost (Adaboost); Artificial Neural Network (ANN); Backward Interval PLS (Bipls); Bayesian Optimization (BO); Classification Trees (CT); Convolutional Neural Network (CNN); Decision Tree (DT); Deep Neural Network Model (DNN); Extreme Gradient Boosting (EGB); Extreme Gradient Boosting (XGBoost); Extreme Learning Machine (ELM); Gaussian Process Regression (GPR); Generalized Additive Model (GAM); Gradient Boosting (GB); Multiple Linear Regression (MLR); Multilayer Perceptron Classifiers (MPC); Neural Network Regression (NNR); Optimal Elman Recurrent Neural Network (OERNN); Partial Least Squares (PLS); Linear Regression (LG); Optimized Gradient Boosting (OGB); Partial Least Squares Discrimination Analysis (PLS-DA); Particle Swarm (PC); Principal Component Analysis (PCA); Radial Basis Function Neural Network (RBFNN); Random Forest (RF); Gradient Boosting Decision Tree (GBDT); Response Surface Methodology (RSM); Support Vector Machine Linear Kernel Method (SVMLK); Support Vector Machines (SVM); Support Vector Regression (SVR).

Table 7 – *Information on datasets used in ML modeling within lignocellulose biorefineries*

List of variables	Number of samples	Availability of the dataset in open data frame	URL address for assessing data	Reference
Input variables: trichloroethylene inlet concentration, residence time, trichloroethylene inlet loading rate Output variable: trichloroethylene concentration at the inlet of continuous stirred tank reactor	77	No		[112]
Input variables: DNA sequences of a target gene and medium composition Output variable: substrate for single enzymes	4389	No		[113]
Input variables: enzyme-substrate pairs Output variable: direction of gene regulation	274,030	Yes	https://github.com/AlexanderKroll/ESP	[114]
Input variables: multispectral fluorescence microscopy and flow cytometry Output variable: alive or dead cell number	Not specified	No		[117]
Input variables: genomic data, temperature, HRT, OLR, waste type, pH and operation model Output variable: medium-chain carboxylic acid	752	No		[123]
Input variables: microwave power, exposure time and solid loading Output variable: glucose and xylose yields	54	No		[164]

is still needed in this field. The potential of ML in biotechnology and its promising future for LBR development can be harnessed in several key areas:

(i) Chemical and physical characterization: Fast and accurate methods for determining LB characteristics can benefit both producers and buyers;

(ii) Pretreatment methods: Accurate prediction models for physical, chemical, biological, chemo-enzymatic, thermal, and chemo-thermal pretreatments can help select optimal methods for increasing biomass surface accessibility and permeability, enhancing conversion rates without producing inhibitors;

(iii) Multi-product production: Models that predict the best multi-product revenue options can ensure economic viability;

(iv) Green solvents: Predictive models for extracting products using novel solvents like NADES can enhance process efficiency;

(v) Catalytic conversion: Models for the catalytic or biocatalytic conversion of polymers to targeted products and fermentable sugars can optimize yields;

(vi) Fermentation and chemical conversion: Accurate models for fermenting or chemically converting fermentable sugars can lead to new product development;

(vii) Techno-economic analysis: Predictive models for techno-economic analysis can guide investment and operational decisions[182–184].

Despite its advantages, ML has certain drawbacks. The opacity of decision-making processes often labels ML algorithms as "black boxes," making it difficult to understand their judgments. Additionally, ML can introduce discrimination and bias, requires specialized knowledge in computer science, mathematics, and statistics, and involves substantial investment in talent and technology. The lack of human involvement in automated tasks can also lead to detachment from the work. Understanding ML's impact on the three pillars of sustainability—social, economic, and environmental—is essential for its effective application in LBRs. Assembling a strong interdisciplinary team with a clear vision is crucial[182–184]. The selection of optimal ML methodologies, tailored to specific process features, ensures detailed insights into critical process variables and their interactions, ultimately advancing the development of lignocellulosic biorefineries.

Literature

1. *de Jong, E., Jungmeier, G.*, Chapter 1 – Biorefinery Concepts in Comparison to Petrochemical Refineries, in Pandey, A., Höfer, R., Taherzadeh, M., Nampoothiri, K.M., Larroche, C. (Eds.), Industrial Biorefineries & White Biotechnology, Elsevier, Amsterdam, 2015, pp 3–33.
2. *Qiao, J., Cui, H., Wang, M., Fu, X., Wang, X., Li, X., Huang, H.*, Integrated biorefinery approaches for the industrialization of cellulosic ethanol fuel, Bioresour. Technol. **360** (2022) 127516. doi: https://doi.org/10.1016/j.biortech.2022.127516
3. *Reshmy, R., Philip, E., Madhavan, A., Sirohi, R., Pugazhendhi, A., Binod, P., Kumar Awasthi, M., Vivek, N., Kumar, V., Sindhu, R.,* Lignocellulose in future biorefineries: Strategies for cost-effective production of biomaterials and bioenergy, Bioresour. Technol. **344** (2022) 126241. doi: https://doi.org/10.1016/j.biortech.2021.126241
4. *Kamm, B., Kamm, M.*, Biorefinery – systems, Chem. Biochem. Eng. Q. **18** (2004) 1.
5. *Wang, W., Lee, D.-J.*, Lignocellulosic biomass pretreatment by deep eutectic solvents on lignin extraction and saccharification enhancement: A review, Bioresour. Technol. **339** (2021) 125587. doi: https://doi.org/10.1016/j.biortech.2021.125587
6. *Singh, N., Singhania, R. R., Nigam, P. S., Dong, C.-D., Patel, A. K., Puri, M.*, Global status of lignocellulosic biorefinery: Challenges and perspectives, Bioresour. Technol. **344** (2022) 126415. doi: https://doi.org/10.1016/j.biortech.2021.126415
7. *Chandel, A. K., Garlapati, V. K., Singh, A. K., Antunes, F. A. F., Da Silva, S. S.,* The path forward for lignocellulose biorefineries: Bottlenecks, solutions, and perspective on commercialization, Bioresour. Technol. **264** (2018) 370. doi: https://doi.org/10.1016/j.biortech.2018.06.004
8. *Galbe, M., Wallberg, O.*, Pretreatment for biorefineries: A review of common methods for efficient utilisation of lignocellulosic materials, Biotechnol. Biofuels **12** (2019) 294. doi: https://doi.org/10.1186/s13068-019-1634-1
9. *Haldar, D., Purkait, M. K.,* A review on the environment-friendly emerging techniques for pretreatment of lignocellulosic biomass: Mechanistic insight and advancements, Chemosphere **264** (2021) 128523. doi: https://doi.org/10.1016/j.chemosphere.2020.128523
10. *Sharma, V., Tsai, M.-L., Chen, C.-W., Sun, P.-P., Patel, A. K., Singhania, R. R., Nargotra, P., Dong, C.-D.*, Deep eutectic solvents as promising pretreatment agents for sustainable lignocellulosic biorefineries: A review, Bioresour. Technol. **360** (2022) 360 127631. doi: https://doi.org/10.1016/j.biortech.2022.127631
11. *Meena, M., Shubham, S., Paritosh, K., Pareek, N., Vivekanand, V.,* Production of biofuels from biomass: Predicting the energy employing artificial intelligence modelling, Bioresour. Technol. **340** (2021) 125642. doi: https://doi.org/10.1016/j.biortech.2021.125642
12. *Gallezot, P.*, Catalytic conversion of biomass: Challenges and issues, ChemSusChem. **1** (2008) 734. doi: https://doi.org/10.1002/cssc.200800091
13. *Dong, Z., Bai, X., Xu, D., Li, W.,* Machine learning prediction of pyrolytic products of lignocellulosic biomass based on physicochemical characteristics and pyrolysis conditions, Bioresour. Technol. **367** (2023) 128182. doi: https://doi.org/10.1016/j.biortech.2022.128182
14. *Garcia, A. C., Shuo, C., Cross, J. S.*, Machine learning based analysis of reaction phenomena in catalytic lignin depolymerization, Bioresour. Technol. **345** (2022) 126503. doi: https://doi.org/10.1016/j.biortech.2021.126503
15. *Ge, H., Zheng, J., Xu, H.*, Advances in machine learning for high value-added applications of lignocellulosic biomass, Bioresour. Technol. **369** (2023) 128481. doi: https://doi.org/10.1016/j.biortech.2022.128481

16. *Khanal, S. K., Tarafdar, A., You, S.*, Artificial intelligence and machine learning for smart bioprocesses, Bioresour. Technol. **375** (2023) 128826. doi: https://doi.org/10.1016/j.biortech.2023.128826
17. *Kow, P.-Y., Lu, M.-K., Lee, M.-H., Lu, W.-B., Chang, F.-J.*, Develop a hybrid machine learning model for promoting microbe biomass production, Bioresour. Technol. **369** (2023) 128412. doi: https://doi.org/10.1016/j.biortech.2022.128412
18. *Mondal, P. P., Galodha, A., Verma, V. K., Singh, V., Show, P. L., Awasthi, M. K., Lall, B., Anees, S., Pollmann, K., Jain, R.,* Review on machine learning-based bioprocess optimization, monitoring, and control systems, Bioresour. Technol. **370** (2023)128523. doi: https://doi.org/ 10.1016/j.biortech.2022.128523
19. *Yadav, A., Sharma, V., Tsai, M.-L., Chen, C.-W., Sun, P.-P., Nargotra, P., Wang, J.-X., Dong, C.-D.*, Development of lignocellulosic biorefineries for the sustainable production of biofuels: Towards circular bioeconomy, Bioresour. Technol. **381** (2023) 129145. doi: https://doi.org/10.1016/j.biortech.2023.129145
20. *Choudhary, K., DeCost, B., Chen, C. Jain, A., Tavazza, F., Cohn, R., WooPark, C., Choudhary, A., Agrawal, A., Bilinge, S. J. L., Holm, E., Ong, S. P., Wolverton, C.*, Recent advances and applications of deep learning methods in materials science, Mater. Sci. **8** (2022) 59. doi: https://doi.org/10.1038/s41524-022-00734-6
21. *Beardall, W. A. V., Stan, G.-B., Dunlop, M. J.,* Deep learning concepts and applications for synthetic biology, GEN Biotechnol. **4** (2022) 360. doi: https://doi.org/10.1089/genbio.2022.0017
22. *Sarker, I. H.*, Machine learning: Algorithms, real-world applications and research directions, SN Comput. Sci. **2** (2021) 160. doi: https://doi.org/10.1007/s42979-021-00592-x
23. *Takkellapati, S., Li, T., Gonzalez, M. A.*, An overview of biorefinery-derived platform chemicals from a cellulose and hemicellulose biorefinery, Clean Techn. Environ. Policy **20** (2018) 1615. doi: https://doi.org/10.1007/s10098-018-1568-5
24. *Martín, M., Taifouris, M., Galán, G.*, Lignocellulosic biorefineries: A multiscale approach for resource exploitation, Biores. Technol. **385** (2023) 129397. doi: https://doi.org/10.1016/j.biortech.2023.129397
25. *K. N, Y., T. M, M. U., S, K., Sachdeva, S., Thakur, S., S, A. K., J, R. B.*, Lignocellulosic biorefinery bechnologies: A perception into recent advances in biomass fractionation, biorefineries, economic hurdles and market outlook, Fermentation **9** (2023) 238. doi: https://doi.org/10.3390/fermentation9030238
26. *Moncada, J., Tamayo, J. A., Cardona, C. A.,* Integrating first, second, and third generation biorefineries: Incorporating microalgae into the sugarcane biorefinery, Chem. Eng. Sci. **118** (2014) 126. doi: https://doi.org/10.1016/j.ces.2014.07.035
27. *Maurya, D. P., Singla, A., Negi, S.,* An overview of key pretreatment processes for biological conversion of lignocellulosic biomass to bioethanol, 3 Biotech. **5** (2015) 597. doi: https://doi.org/10.1007/s13205-015-0279-4
28. *Amândio, M. S. T., Rocha, J. M. S., Xavier, A. M. R. B.*, Enzymatic hydrolysis strategies for cellulosic sugars production to obtain bioethanol from *Eucalyptus globulus* bark, Fermentation **9** (2023) 241. doi: https://doi.org/10.3390/fermentation9030241
29. *Tumuluru, J. S., Ghiasi, B., Soelberg, N. R., Sokhansanj, S.*, Biomass torrefaction process, product properties, reactor types, and moving bed reactor design concepts, Front. Energy Res. **9** (2021) 728140. doi: https://doi.org/10.3389/fenrg.2021.728140
30. *Escalante, J., Chen, W.-H., Tabatabaei, M., Hoang, A. T., Kwon, E. E., Lin, K.-Y. A., Saravanakumar, A.,* Pyrolysis of lignocellulosic, algal, plastic, and other biomass wastes for biofuel production and circular bioeconomy: A review of thermogravimetric analysis (TGA) approach, Renew. Sustain. Energy Rev. **169** (2022) 112914. doi: https://doi.org/10.1016/j.rser.2022.112914
31. *Tezer, Ö., Karabağ, N., Öngen, A., Çolpan, C. Ö., Ayol, A.*, Biomass gasification for sustainable energy production: A review, Int. J. Hydrogen Energy. **47** (2022) 15419. doi: https://doi.org/10.1016/j.ijhydene.2022.02.158
32. *Bontaş, M. G., Diacon, A., Călinescu, I., Rusen E.*, Lignocellulose biomass liquefaction: Process and applications development as polyurethane foams, Polymers **15** (2023) 563. doi: https://doi.org/10.3390/polym15030563
33. *Petrović, J., Ercegović, M., Simić, M., Koprivica, M., Dimitrijević, J., Jovanović, A., Janković Pantić, J.*, Hydrothermal carbonization of waste biomass: A review of hydrochar preparation and environmental application, Processes **12** (2024) 207. doi: https://doi.org/10.3390/pr12010207
34. *Yu, S., Yang, X., Li, Q., Zhang, Y., Zhou, H.*, Breaking the temperature limit of hydrothermal carbonization of lignocellulosic biomass by decoupling temperature and pressure, Green Energy Environ. **8** (2023) 1216. doi: https://doi.org/10.1016/j.gee.2023.01.001
35. *Yu, S., Dong, X., Zhao, P., Luo, Z., Sun, Z., Yang, X., Li, Q., Wang, L., Zhang, Y., Zhou, H.,* Decoupled temperature and pressure hydrothermal synthesis of carbon sub-micron spheres from cellulose, Nat. Commun. **13** (2022) 3616. doi: https://doi.org/10.1038/s41467-022-31352-x
36. *García, A., González Alriols, M., Labidi, J.*, Evaluation of different lignocellulosic raw materials as potential alternative feedstocks in biorefinery processes, Ind. Crops Prod. **53** (2014) 102. doi: https://doi.org/10.1016/j.indcrop.2013.12.019
37. *Konwar, L. J., Mikkola, J.-P., Bordoloi, N., Saikia, R., Chutia, R. S., Kataki, R.*, Sidestreams From Bioenergy and Biorefinery Complexes as a Resource for Circular Bioeconomy, in Bhaskar, T., Pandey, A., Mohan, S. V., Lee, D.-J., Khanal, S. K. (Eds.), Waste Biorefinery, Elsevier, Amsterdam, 2018, pp 85–125.
38. *Tišma, M., Bucić-Kojić, A., Planinić, M.*, Bio-based products from lignocellulosic waste biomass: A state of the art, Chem. Biochem. Eng. Q. **35** (2021) 139. doi: https://doi.org/10.15255/CABEQ.2021.1931
39. *Gandam, P. K., Chinta, M. L., Pabbathi, N. P. P., Velidandi, A., Sharma, M., Kuhad, R. C., Tabatabaei, M., Aghbashlo, M., Baadhe, R. R., Gupta, V. K.*, Corncob-based biorefinery: A comprehensive review of pretreatment methodologies, and biorefinery platforms, J. Energy Inst. **101** (2022) 290. doi: https://doi.org/10.1016/j.joei.2022.01.004
40. *Tišma, M., Žnidaršič-Plazl, P., Šelo, G., Tolj, I., Šperanda, M., Bucić-Kojić, A., Planinić, M.*, *Trametes versicolor* in lignocellulose-based bioeconomy: State of the art, challenges and opportunities, Bioresour. Technol. **330** (2021) 124997. doi: https://doi.org/10.1016/j.biortech.2021.124997

41. *Ai, N., Jiang, Y., Omar, S., Wang, J., Xia, L., Ren, J.,* Rapid measurement of cellulose, hemicellulose, and lignin content in *Sargassum horneri* by near-infrared spectroscopy and characteristic variables selection methods, Molecules **27** (2022) 335.
doi: https://doi.org/10.3390/molecules27020335
42. *Li, Y., Tao, L., Nagle, N., Tucker, M., Chen, X., Kuhn, E. M.*, Effect of feedstock variability, feedstock blends, and pretreatment conditions on sugar yield and production costs, Front. Energy Res. **9** (2022) 792216.
doi: https://doi.org/10.3389/fenrg.2021.792216
43. *Wood, I. P., Garcia-Gutierrez, E., Wellner, N., Waldron, K. W.*, Feedstock selection for polymer and chemical production: feedstock-specific recalcitrance, Faraday Discuss. **202** (2017) 391.
doi: https://doi.org/10.1039/C7FD00044H
44. *El Hage, M., Louka, N., Rezzoug, S.-A., Maugard, T., Sablé, S., Koubaa, M., Debs, E., Maache-Rezzoug, Z.,* Bioethanol production from woody biomass: Recent advances on the effect of pretreatments on the bioconversion process and energy yield aspects, Energies **16** (2023) 5052.
doi: https://doi.org/10.3390/en16135052
45. *Tan, J., Li, Y., Tan, X., Wu, H., Li, H., Yang, S.*, Advances in pretreatment of straw biomass for sugar production, Front. Chem. **9** (2021) 696030.
doi: https://doi.org/10.3389/fchem.2021.696030
46. *Mujtaba, M., Fraceto, L. F., Fazeli, M., Mukherjee, S., Savassa, S. M., de Medeiros, G. A., Pereira, A. E. S., Mancini, S. D., Lipponen, J., Vilaplana, F.*, Lignocellulosic biomass from agricultural waste to the circular economy: A review with focus on biofuels, biocomposites and bioplastics, J. Clean. Prod. **402** (2023) 136815.
doi: https://doi.org/10.1016/j.jclepro.2023.136815
47. *Vasić, K., Knez, Ž., Leitgeb, M.*, Bioethanol production by enzymatic hydrolysis from different lignocellulosic sources, Molecules **26** (2021) 753.
doi: https://doi.org/10.3390/molecules26030753
48. *Preethi, G. M., Kumar, G., Karthikeyan, O. P., Varjani, S., Banu, R. J.*, Lignocellulosic biomass as an optimistic feedstock for the production of biofuels as valuable energy source: Techno-economic analysis, environmental impact analysis, breakthrough and perspectives, Environ. Technol. Innov. **24** (2021) 102080.
doi: https://doi.org/10.1016/j.eti.2021.102080
49. *Isikogor, F. H., Becer, C. R.*, Lignocellulosic biomass: A sustainable platform for the production of bio-based chemicals and polymers, Polym. Chem. **6** (2015) 4497.
doi: https://doi.org/10.1039/C5PY00263J
50. *De Mayer, A., Cattrysse, D., Orshoven, J. V.*, Considering biomass growth and regeneration in the optimisation of biomass supply chains, Renew. Energ. **87** (2016) 990.
doi: https://doi.org/10.1016/j.renene.2015.07.043
51. *Hansen, J. K., Roni, M. S., Nair, S. K., Hartley, D. S., Griffel, L. M., Vazhnik, V., Mamun, S.,* Setting a baseline for integrated landscape design: Cost and risk assessment in herbaceous feedstock supply chains, Biomass Bioeng. **130** (2019) 105388.
doi: https://doi.org/10.1016/j.biombioe.2019.105388
52. *Abolore, R. S., Jasiwal, S., Jaiswal, A. K.*, Green and sustainable pretreatment methods for cellulose extraction from lignocellulosic biomass and its applications: A review, Carbohydr. Polym. Technol. Appl. **7** (2024) 100396.
doi: https://doi.org/10.1016/j.carpta.2023.100396
53. *Serrano, A., Faulin, J., Astiz, P., Sánchez, M., Belloso, J.,* Locating and designing a biorefinery supply chain under uncertainty in navarre: A stochastic facility location problem case, Transp. Res. Procedia **10** (2015) 704.
doi: https://doi.org/10.1016/j.trpro.2015.09.024
54. *Šibalić, D., Šalić, A., Zelić, B., Tran, N. N., Hessel, V., Nigam, K. D. P., Tišma, M.*, Synergism of ionic liquids and lipases for lignocellulosic biomass valorization, Chem. Eng. J. **461** (2023) 142011.
doi: https://doi.org/10.1016/j.cej.2023.142011
55. *Jayakody, L. N., Chinmoy, B., Turner, T. L.*, Trends in valorization of highly-toxic lignocellulosic biomass derived-compounds via engineered microbes, Bioresour. Technol. **346** (2022) 126614.
doi: https://doi.org/10.1016/j.biortech.2021.126614
56. *Guo, H., Zhao, Y., Chang, J.-S., Lee, D.-J.*, Inhibitor formation and detoxification during lignocellulose biorefinery: A review, Bioresour. Technol. **361** (2022) 127666.
doi: https://doi.org/10.1016/j.biortech.2022.127666
57. *Ashokkumar, V., Venkatkarthick, R., Jayashree, S., Chuetor, S., Dharmaraj, S., Kumar, G., Chen, W.-H., Ngamcharussrivichai, C.*, Recent advances in lignocellulosic biomass for biofuels and value-added bioproducts – A critical review, Bioresour. Technol. **344** (2022) 126195.
doi: https://doi.org/10.1016/j.biortech.2021.126195
58. *Liguori, R., Faraco, V.,* Biological processes for advancing lignocellulosic waste biorefinery by advocating circular economy, Bioresour. Technol. **215** (2016) 13.
doi: https://doi.org/10.1016/j.biortech.2016.04.054
59. *Haberzettl, J., Hilgert, P., von Cossel, M.*, A critical review on lignocellulosic biomass yield modeling and the bioenergy potential from marginal land, Agronomy **11** (2021) 2397.
doi: https://doi.org/10.3390/agronomy11122397
60. *Blasi, A., Verardi, A., Lopresto, C. G., Siciliano, S., Sangiorgio, P.,* Lignocellulosic agricultural waste valorization to obtain valuable products: An overview, Recycling **8** (2023) 61.
doi: https://doi.org/10.3390/recycling8040061
61. *Benalcázar, E. A., Deynoot, B. G., Noorman, H., Osseweijer, P.,* Production of bulk chemicals from lignocellulosic biomass via thermochemical conversion and syngas fermentation: A comparative techno-economic and environmental assessment of different site-specific supply chain configurations, Biofuel. Bioprod. Bior. **11** (2017) 861.
doi: https://doi.org/10.1002/bbb.1790
62. *Schmidt, J., Marques, M. R. G., Botti, S., Marques, M. A. L.*, Recent advances and applications of machine learning in solid-state materials science, Mater. Sci. **5** (2019) 83.
doi: https://doi.org/10.1038/s41524-019-0221-0
63. *Mowbray, M., Savage, T., Wu, C., Song, Z., Cho, B. A., Del Rio-Chanona, E. A., Zhang, D.,* Machine learning for biochemical engineering: A review, Biochem. Eng. J. **172** (2021) 108054.
doi: https://doi.org/10.1016/j.bej.2021.108054
64. *Abdualgalil, B., Abraham, S.*, Applications of Machine Learning Algorithms and Performance Comparison: A Review, in Singh, K. J., Shynu, P. G. (Eds.), Proceedings of the 2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE), IEEE, Vellore, India, 2020 pp 1–6.
65. *Pineda-Jaramillo, J. D.*, A review of machine learning (ML) algorithms used for modeling travel mode choice, Revista DYNA **86** (2019) 32.
doi: https://doi.org/10.15446/dyna.v86n211.79743
66. *Basha, S. M., Rajput, D. S.*, Survey on Evaluating the Performance of Machine Learning Algorithms: Past Contributions and Future Roadmap, in Sangaiah, K. (Ed.), Deep Learning and Parallel Computing Environment for Bioengineering Systems, Elsevier, Amsterdam, 2019, pp 153–164.

67. *Wang, G., Pu, P., Shen, T.,* An efficient gene big data analysis using machine learning algorithms, Multimed. Tools Appl. **79** (2020) 9847. doi: https://doi.org/10.1007/s11042-019-08358-7
68. *Ayodele, T. O.*, Types of Machine Learning Algorithms, in Zhang, Y. (Ed.), New Advances in Machine Learning, IntechOpen Limited, London, 2010, pp. 19–48.
69. *Rashidi, H. H., Tran, N. K., Betts, E. V., Howell, L. P., Green, R.*, Artificial intelligence and machine learning in pathology: The present landscape of supervised methods, Acad. Pathol. **6** (2019) 2374289519873088. doi: https://doi.org/10.1177/2374289519873088
70. *Yuvalı, M., Yaman, B., Tosun, Ö.*, Classification comparison of machine learning algorithms using two independent CAD datasets, Mathematics **10** (2022) 311. doi: https://doi.org/10.3390/math10030311
71. *Dhanabal, S., Chandramathi, D. S.,* A review of various k-nearest neighbor query processing techniques, Int. J. Comput. Appl. **31** (2011) 14.
72. *Imandoust, S. B., Bolandraftar, M.,* Application of k-nearest neighbor (knn) approach for predicting economic events: Theoretical background, Int. J. Eng. Res. Appl. **3** (2013) 605.
73. *Devi, F. R., Sugiharti, E., Arifudin, R.*, The comparison combination of Naïve Bayes classification algorithm with fuzzy c-means and k-means for determining beef cattle quality in Semarang regency, Sci. J. Inform. **5** (2018) 194. doi: https://doi.org/10.15294/sji.v5i2.15452
74. *Bhargavi, P., Jyothi, S.*, Applying Naive Bayes data mining technique for classification of agricultural land soils, Int. J. Netw. Secur. **9** (2009) 117.
75. *Mumuni, A., Mumuni, F.,* Automated data processing and feature engineering for deep learning and big data applications: A survey, J. Autom. Intellig. 2024. doi: https://doi.org/10.1016/j.jiixd.2024.01.002
76. *Milo, T., Somech, A.*, Automating exploratory data analysis via machine learning: An overview, Proceedings of the 2020 ACM SIGMOD International Conference on Management of Data, 2020, 2617. doi: https://doi.org/10.1145/3318464.3383126
77. *Montañez-Barrera, J. A., Barroso-Maldonaldo, J. M., Bedoya-Santacruz, A. F., Mota-Babiloni, A.,* Correlated-informed neural networks: A new machine learning framework to predict pressure drop in micro-channels, Int. J. Heat Mass Transf. **194** (2022) 123017. doi: https://doi.org/10.1016/j.ijheatmasstransfer.2022.123017
78. *Alam, M. T., Ubaid, S., Shakil, Sohail, S. S., Nadeem, M., Hussain, S., Siddiqui, J.,* Comparative analysis of machine learning based filtering techniques using MovieLens dataset, Procedia Comput. Sci.**194** (2021) 210. doi: https://doi.org/10.1016/j.procs.2021.10.075
79. *Patel, H. H., Prajapati, P.*, Study and analysis of decision tree based classification algorithms, Int. J. Adv. Res. Comput. Sci. Softw. Eng. **6** (2018) 74. doi: https://doi.org/10.26438/ijcse/v6i10.7478
80. *Ali, J., Khan, R., Ahmad, N., Maqsood, I.*, Random forests and decision trees, Int. J. Comp. Sci. Issu. **9** (2012) 272.
81. *Coadou, Y.*, Boosted decision trees and applications, EPJ Web of Conferences **55** (2013) 02004. doi: https://doi.org/10.1051/epjconf/20135502004
82. *Evgeniou, T., Pontil, M.*, Support Vector Machines: Theory and Applications, in Lipo, W. (Ed.), Machine Learning and Its Applications, Springer Berlin, 2001, pp 249–257.
83. *Ladjici, A. A., Boudour, M., Rachedi, N.*, Power system applications of support vector machine in classification and regression, 3rd International Conference on Electrical Engineering, Algiers, Algeria, 2009, pp 522–527.
84. *Basheer, I. A., Hajmeer, M.,* Artificial neural networks: Fundamentals, computing, design, and application, J. Microbiol. Methods **43** (2000) 3. doi: https://doi.org/10.1016/S0167-7012(00)00201-3
85. *Wang, S., Di, J., Wang, D., Dai, X., Hua, Y., Gao, X., Zheng, A., Gao, J.*, State-of-the-art review of artificial neural networks to predict, characterize and optimize pharmaceutical formulation, Pharmaceutics **14** (2022) 183. doi: https://doi.org/10.3390/pharmaceutics14010183
86. *Abiodun, O. I., Kiru, M. U., Jantan, A., Omolara, A. E., Dada, K. V., Umar, A. M., Linus, O. U., Arshad, H., Kazaure, A. A., Gana, U.*, Comprehensive review of artificial neural network applications to pattern recognition, IEEE Access **7** (2017) 158820. doi: https://doi.org/10.1109/ACCESS.2019.2945545
87. *Li, X., Wu, X.*, Constructing Long Short-Term Memory Based Deep Recurrent Neural Networks for Large Vocabulary Speech Recognition, in Gray, D., Cochran, D. (Eds.), Proceedings of the 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP); IEEE, South Brisbane, Queensland, 2015, pp, 4520–4524.
88. *Hochreiter, S., Schmidhuber, J.*, Long short-term memory, Neural Comput. **9** (1997) 1735. doi: https://doi.org/10.1162/neco.1997.9.8.1735
89. *Krizhevsky, A., Sutskever, I., Hinton, G. E.*, ImageNet classification with deep convolutional neural networks, Commun. ACM **60** (2012) 84. doi: https://doi.org/10.1145/3065386
90. *Pouyanfar, S., Sadiq, S., Yan, Y., Tian, H., Tao, Y., Reyes, M. P., Shyu, M.-L., Chen, S.-C., Iyengar, S. S.*, A survey on deep learning: Algorithms, techniques, and applications, ACM Comput. Surv. **51** (2018) 1. doi: https://doi.org/10.1145/3234150
91. *Devlin, J., Chang, M.-W., Lee, K., Toutanova, K.*, BERT: Pre-training of deep bidirectional transformers for language understanding 2018, Proceedings of NAACL-HLT 2019, Minneapolis, Minnesota, 2019, pp 4171–4186.
92. *Sanh, V., Debut, L., Chaumond, J., Wolf, T.*, DistilBERT, a distilled version of BERT: Smaller, faster, cheaper and lighter, arXiv 2020. doi: https://doi.org/10/gp5knf
93. *Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., Hesse, C., Chen, M., Sigler, E., Litwin, M., Gray, S., Chess, B., Clark, J., Berner, C., McCandlish, S., Radford, A., Sutskever, I., Amodei, D.*, Language models are few-shot learners, 34th Conference on Neural Information Processing Systems (NeurIPS 2020), Vancouver, Canada, 2020, pp 1-25.
94. *Gobov, D., Solovei, O.,* Approaches to improving the accuracy of machine learning models in requirements elicitation techniques selection. In: Hu, Z., Dychka, I., He, M. (eds) Advances in Computer Science for Engineering and Education VI. ICCSEEA 2023. Lecture Notes on Data Engineering and Communications Technologies, vol 181 (2023). Springer, Cham. doi: https://doi.org/10.1007/978-3-031-36118-0_51
95. *Van Giffen, B., Herhausen, D., Fhse, T.*, Overcoming the pitfalls and per ils of algorithms: A classification of machine learning biases and mitigation methods, J. Bus. Res. **144** (2022) 93. doi: https://doi.org/10.1016/j.jbusres.2022.01.076

96. *Kee, E., Chong, J. J., Choong, Z. J., Lau, M.,* A comparative analysis of cross-validation techniques for a smart and lean pick-and-place solution with deep learning, Electronics **12** (2023) 2371.
doi: https://doi.org/10.3390/electronics12112371
97. *Mezzadri, G., Laloë, T., Mathy, F., Reynaud-Bouret, P.*, Hold-out strategy for selecting learning models: Application to categorization subjected to presentation orders, J. Math. Psychol. **109** (2022) 102691.
doi: https://doi.org/10.1016/j.jmp.2022.102691
98. *Stuke, A., Rinke, P., Todorović, M.*, Efficient hyperparameter tuning for kernel ridge regression with Bayesian optimization, Mach. Learn. Sci. Technol. **2** (2021) 035022.
doi: https://doi.org/10.1088/2632-2153/abee59
99. *Alshakhs, F., Alharthi, H., Aslam, N., Khan, I. U., Elasheri, M.*, Predicting postoperative length of stay for isolated coronary artery bypass graft patients using machine learning, Int. J. Gen. Med. **13** (2020) 751.
doi: https://doi.org/10.2147/IJGM.S250334
100. Free Machine Learning Diagram Available online: https://yourfreetemplates.com/free-machine-learning-diagram/ (accessed on 19 July 2023).
101. *Oliveira, A. L.*, Biotechnology, big data and artificial intelligence, Biotechnol. J. **14** (2019) 1800613.
doi: https://doi.org/10.1002/biot.201800613
102. *Remington, J. M., Ferrell, J. B., Zorman, M., Petrucci, A., Schneebeli, S. T., Li, J.*, Machine learning in a molecular modeling course for chemistry, biochemistry, and biophysics students, Biophysicist **1** (2020) 11.
doi: https://doi.org/10.35459/tbp.2019.000140
103. *Zhou, Y., Li, G., Dong, J., Xing, X., Dai, J., Zhang, C.*, MiYA, an efficient machine-learning workflow in conjunction with the YeastFab assembly strategy for combinatorial optimization of heterologous metabolic pathways in *Saccharomyces cerevisiae*, Metab. Eng. **47** (2018) 294.
doi: https://doi.org/10.1016/j.ymben.2018.03.020
104. *Zhang, J., Petersen, S. D., Radivojevic, T., Ramirez, A., Pérez-Manríquez, A., Abeliuk, E., Sánchez, B. J., Costello, Z., Chen, Y., Fero, M. J., Martin, H. G., Nielsen, J., Keasling, J. D., Jensen, M. K.,* Combining mechanistic and machine learning models for predictive engineering and optimization of tryptophan metabolism, Nat. Commun. **11** (2020) 4880.
doi: https://doi.org/10.1038/s41467-020-17910-1
105. *Cheng, Y., Bi, X., Xu, Y., Liu, Y., Li, J., Du, G., Lv, X., Liu, L.*, Artificial intelligence technologies in bioprocess: Opportunities and challenges, Bioresour. Technol. **369** (2023) 128451.
doi: https://doi.org/10.1016/j.biortech.2022.128451
106. *Tulsyan, A., Garvin, C., Undey, C.*, Machine-learning for biopharmaceutical batch process monitoring with limited data, IFAC-PapersOnLine **51** (2018) 126.
doi: https://doi.org/10.1016/j.ifacol.2018.09.287
107. *Onel, M., Kieslich, C. A., Guzman, Y. A., Floudas, C. A., Pistikopoulos, E. N.*, Big data approach to batch process monitoring: Simultaneous fault detection and diagnosis using nonlinear support vector machine-based feature selection, Comput. Chem. Eng. **115** (2018) 46.
doi: https://doi.org/10.1016/j.compchemeng.2018.03.025
108. *Chen, Y., Yu, G., Long, Y., Teng, J., You, X., Liao, B.-Q., Lin, H.*, Application of radial basis function artificial neural network to quantify interfacial energies related to membrane fouling in a membrane bioreactor, Bioresour. Technol. **293** (2019) 122103.
doi: https://doi.org/10.1016/j.biortech.2019.122103
109. *Hamedi, H., Ehteshami, M., Mirbagheri, S. A., Zendehboudi, S.*, New deterministic tools to systematically investigate fouling occurrence in membrane bioreactors, Chem. Eng. Res. Des. **144** (2019) 334.
doi: https://doi.org/10.1016/j.cherd.2019.02.003
110. *Hazrati, H., Moghaddam, A. H., Rostamizadeh, M.*, The influence of hydraulic retention time on cake layer specifications in the membrane bioreactor: Experimental and artificial neural network modeling, J. Environ. Chem. Eng. **5** (2017) 3005.
doi: https://doi.org/10.1016/j.jece.2017.05.050
111. *Baskaran, D., Sinharoy, A., Pakshirajan, K., Rajamanickam, R.*, Gas-phase trichloroethylene removal by *Rhodococcus opacus* using an airlift bioreactor and its modeling by artificial neural network, Chemosphere **247** (2020) 125806.
doi: https://doi.org/10.1016/j.chemosphere.2019.125806
112. *Baskaran, D., Sinharoy, A., Paul, T., Pakshirajan, K., Rajamanickam, R.*, Performance evaluation and neural network modeling of trichloroethylene removal using a continuously operated two-phase partitioning bioreactor, Environ. Technol. Innov. **17** (2020) 100568.
doi: https://doi.org/10.1016/j.eti.2019.100568
113. *Kwon, M. S., Adidjaja, J. J., Kim, H. U.*, Predicting the effects of cultivation condition on gene regulation in *Escherichia coli* by using deep learning, Comput. Struct. Biotechnol. J. **21** (2023) 2613.
doi: https://doi.org/10.1016/j.csbj.2023.04.010
114. *Kroll, A., Ranjan, S., Engqvist, M. K. M., Lercher, M. J.*, A general model to predict small molecule substrates of enzymes based on machine and deep learning, Nat. Commun. **14** (2023) 2787.
doi: https://doi.org/10.1038/s41467-023-38347-2
115. *Mou, Z., Eakes, J., Cooper, C. J., Foster, C. M., Standaert, R. F., Podar, M., Doktycz, M. J., Parks, J. M.,* Machine learning-based prediction of enzyme substrate scope: Application to bacterial nitrilases, Proteins **89** (2020) 336.
doi: https://doi.org/10.1002/prot.26019
116. *Mehrian, M., Lambrechts, T., Marechal, M., Luyten, F. P., Papantoniou, I., Geris, L.*, Predicting in vitro human mesenchymal stromal cell expansion based on individual donor characteristics using machine learning, Cytotherapy **22** (2020) 82.
doi: https://doi.org/10.1016/j.jcyt.2019.12.006
117. *Reimann, R., Zeng, B., Jakopec, M., Burdukiewicz, M., Petrick, I., Schierack, P., Rödiger, S.*, Classification of dead and living microalgae *Chlorella vulgaris* by bioimage informatics and machine learning, Algal Res. **48** (2020) 101908.
doi: https://doi.org/10.1016/j.algal.2020.101908
118. *Cai, W., Lesnik, K. L., Wade, M. J., Heidrich, E. S., Wang, Y., Liu, H.,* Incorporating microbial community data with machine learning techniques to predict feed substrates in microbial fuel cells, Biosens. Bioelectron. **133** (2019) 64.
doi: https://doi.org/10.1016/j.bios.2019.03.021
119. *Barberi, G., Benedetti, A., Diaz-Fernandez, P., Finka, G., Bezzo, F., Barolo, M., Facco, P.*, Anticipated cell lines selection in bioprocess scale-up through machine learning on metabolomics dynamics, IFAC-PapersOnLine **54** (2021) 85.
doi: https://doi.org/10.1016/j.ifacol.2021.08.223
120. *Shrivastava, A. D., Swainston, N., Samanta, S., Roberts, I., Wright Muelas, M., Kell, D. B.*, MassGenie: A transformer-based deep learning method for identifying small molecules from their mass spectra, Biomolecules **11** (2021) 1793.
doi: https://doi.org/10.3390/biom11121793

121. *Grechishnikova, D.,* Transformer neural network for protein-specific de novo drug generation as a machine translation problem, Sci. Rep. **11** (2021) 321. doi: https://doi.org/10.1038/s41598-020-79682-4
122. *Upadhyay, A., Kovalev, A. A., Zhuravleva, E. A., Pareek, N., Vivekanand, V*., Enhanced production of acetic acid through bioprocess optimization employing response surface methodology and artificial neural network, Bioresour. Technol. **376** (2023) 128930. doi: https://doi.org/10.1016/j.biortech.2023.128930
123. *Long, F., Fan, J., Xu, W., Liu, H.*, Predicting the performance of medium-chain carboxylic acid (MCCA) production using machine learning algorithms and microbial community data, J. Clean. Prod. **377** (2022) 134223. doi: https://doi.org/10.1016/j.jclepro.2022.134223
124. *Pei, Z., Liu, S., Jing, Z., Zhang, Y., Wang, J., Liu, J., Wang, Y., Guo, W., Li, Y., Feng, L., Zhou, H., Li, G., Han, Y., Liu, D., Pan, J.,* Understanding of the interrelationship between methane production and microorganisms in high-solid anaerobic co-digestion using microbial analysis and machine learning, J. Clean. Prod. **373** (2022) 133848. doi: https://doi.org/10.1016/j.jclepro.2022.133848
125. *Pandian, B. J., Noel, M. M.*, Control of a bioreactor using a new partially supervised reinforcement learning algorithm, J. Process Control **69** (2018) 16. doi: https://doi.org/10.1016/j.jprocont.2018.07.013
126. *Ma, Y., Noreña-Caro, D. A., Adams, A. J., Brentzel, T. B., Romagnoli, J. A., Benton, M. G.*, Machine-Learning-based simulation and fed-batch control of cyanobacterial-phycocyanin production in plectonema by artificial neural network and deep reinforcement learning, Comp. Chem. Eng. **142** (2020) 107016. doi: https://doi.org/10.1016/j.compchemeng.2020.107016
127. *Sarmah, N., Mehtab, V., Bugata, L. S. P., Tardio, J., Bhargava, S., Parthasarathy, R., Chenna, S.*, Machine learning aided experimental approach for evaluating the growth kinetics of *Candida antarctica* for lipase production, Bioresour. Technol. **352** (2022) 127087. doi: https://doi.org/10.1016/j.biortech.2022.127087
128. *Adejimi, O. E., Ignat, T., Sadhasivam, G., Zakin, V., Schmilovitch, Z., Shapiro, O. H.,* Low-resolution Raman spectroscopy for the detection of contaminant species in algal bioreactors, Sci. Total Environ. **809** (2022) 151138. doi: https://doi.org/10.1016/j.scitotenv.2021.151138
129. *Li, M., Wijewardane, N. K., Ge, Y., Xu, Z., Wilkins, M. R.*, Visible/near infrared spectroscopy and machine learning for predicting polyhydroxybutyrate production cultured on alkaline pretreated liquor from corn stover, Bioresour. Technol. Rep. **9** (2020) 100386. doi: https://doi.org/10.1016/j.biteb.2020.100386
130. *Manjrekar, O. N., Dudukovic, M. P.*, Identification of flow regime in a bubble column reactor with a combination of optical probe data and machine learning technique, Chem. Eng. Sci. **2** (2019) 100023. doi: https://doi.org/10.1016/j.cesx.2019.100023
131. *Patel, G., Patil, M. D., Tangadpalliwar, S., Nile, S. H., Garg, P., Kai, G., Banerjee, U. C.*, Machine learning modeling for ultrasonication-mediated fermentation of *Penicillium brevicompactum* to enhance the release of mycophenolic acid, Ultrasound Med. Biol. **47** (2021) 777. doi: https://doi.org/10.1016/j.ultrasmedbio.2020.11.018
132. *Bhilare, K. D., Patil, M. D., Tangadpalliwar, S., Shinde, A., Garg, P., Banerjee, U. C.,* Machine learning modelling for the ultrasonication-mediated disruption of recombinant *E. coli* for the efficient release of nitrilase, Ultrasonics **98** (2019) 72. doi: https://doi.org/10.1016/j.ultras.2019.06.006
133. *Coşgun, A., Günay, M. E., Yıldırım, R.*, Exploring the critical factors of algal biomass and lipid production for renewable fuel production by machine learning, Renew. Energ. **162** (2021) 1299. doi: https://doi.org/10.1016/j.renene.2020.09.034
134. *Viet, N. D., Jang, A.,* Machine learning-based real-time prediction of micropollutant behaviour in forward osmosis membrane (waste)water treatment, J. Clean. Prod. **389** (2023) 136023. doi: https://doi.org/10.1016/j.jclepro.2023.136023
135. *Ren, K., Jiao, Z., Wu, X., Han, H.,* Multivariable identification of membrane fouling based on compacted cascade neural network, Chin. J. Chem. Eng. **53** (2023) 37. doi: https://doi.org/10.1016/j.cjche.2022.01.028
136. *Mbamba, C. K., Batstone, D. J.*, Optimization of deep learning models for forecasting performance in the water industry using genetic algorithms, Comp. Chem. Eng. **175** (2023) 108276. doi: https://doi.org/10.1016/j.compchemeng.2023.108276
137. *Rezk, H., Olabi, A. G., Sayed, E. T., Alshathri, S. I., Abdelkareem, M. A.*, Optimized artificial intelligent model to boost the efficiency of saline wastewater treatment based on hunger games search algorithm and ANFIS, Sustainability **15** (2023) 4413. doi: https://doi.org/10.3390/su15054413
138. *Negi, B. B., Aliveli, M., Behera, S. K., Das, R., Sinharoy, A., Rene, E. R., Pakshirajan, K.,* Predictive modelling and optimization of an airlift bioreactor for selenite removal from wastewater using artificial neural networks and particle swarm optimization, Environ. Res. **2019** (2023) 115073. doi: https://doi.org/10.1016/j.envres.2022.115073
139. *Li, G., Ji, J., Ni, J., Wang, S., Guo, Y., Hu, Y., Liu, S., Huang, S.-F., Li, Y.-Y.*, Application of deep learning for predicting the treatment performance of real municipal wastewater based on one-year operation of two anaerobic membrane bioreactors, Sci. Total Environ. **813** (2022) 151920. doi: https://doi.org/10.1016/j.scitotenv.2021.151920
140. *Yaqub, M., Lee, W.*, Modeling nutrient removal by membrane bioreactor at a sewage treatment plant using machine learning models, J. Water Process Eng. **46** (2022) 102521. doi: https://doi.org/10.1016/j.jwpe.2021.102521
141. *Viet, N. D., Im, S.-J., Kim, C.-M., Jang, A.*, An osmotic membrane bioreactor–clarifier system with a deep learning model for simultaneous reduction of salt accumulation and membrane fouling, Chemosphere **272** (2021) 129872. doi: https://doi.org/10.1016/j.chemosphere.2021.129872
142. *Viet, N. D., Jang, A.*, Development of artificial intelligence-based models for the prediction of filtration performance and membrane fouling in an osmotic membrane bioreactor, J. Environ. Chem. Eng. **9** (2021) 105337. doi: https://doi.org/10.1016/j.jece.2021.105337
143. *Zaghloul, M. S., Iorhemen, O. T., Hamza, R. A., Tay, J. H., Achari, G.*, Development of an ensemble of machine learning algorithms to model aerobic granular sludge reactors, Water Res. **189** (2021) 116657. doi: https://doi.org/10.1016/j.watres.2020.116657
144. *Sbahi, S., Ouazzani, N., Hejjaj, A., Mandi, L.*, Nitrogen modeling and performance of multi-soil-layering (MSL) bioreactor treating domestic wastewater in rural community, J. Water Process. Eng. **44** (2021) 102389. doi: https://doi.org/10.1016/j.jwpe.2021.102389
145. *Chiu, M.-C., Wen, C.-Y., Hsu, H.-W., Wang, W.-C.*, Key wastes selection and prediction improvement for biogas production through hybrid machine learning methods, Sustain. Energy Technol. Assess. **52** (2022) 102223. doi: https://doi.org/10.1016/j.seta.2022.102223

146. *Ascher, S., Wang, X., Watson, I., Sloan, W., You, S.*, Interpretable machine learning to model biomass and waste gasification, Bioresour. Technol. **364** (2022) 128062. doi: https://doi.org/10.1016/j.biortech.2022.128062
147. *Hosseinzadeh, A., Zhou, J. L., Altaee, A., Li, D.,* Machine learning modeling and analysis of biohydrogen production from wastewater by dark fermentation process, Bioresour. Technol. **343** (2022) 126111. doi: https://doi.org/10.1016/j.biortech.2021.126111
148. *Ly, Q. V., Truong, V. H., Ji, B., Nguyen, X. C., Cho, K. H., Ngo, H. H., Zhang, Z.*, Exploring potential machine learning application based on big data for prediction of wastewater quality from different full-scale wastewater treatment plants, Sci. Total Environ. **832** (2022) 154930. doi: https://doi.org/10.1016/j.scitotenv.2022.154930
149. *Zhuang, L., Tang, B., Bin, L., Li, P., Huang, S., Fu, F.*, Performance prediction of an internal-circulation membrane bioreactor based on models comparison and data features analysis, Biochem. Eng. J. **166** (2021) 107850. doi: https://doi.org/10.1016/j.bej.2020.107850
150. *Pang, J.-W., Yang, S.-S., He, L., Chen, Y.-D., Cao, G.-L., Zhao, L., Wang, X.-Y., Ren, N.-Q.*, An influent responsive control strategy with machine learning: Q-learning based optimization method for a biological phosphorus removal system, Chemosphere **234** (2019) 893. doi: https://doi.org/10.1016/j.chemosphere.2019.06.103
151. *Wang, L., Long, F., Liang, D., Xiao, X., Liu, H.,* Hydrogen production from lignocellulosic hydrolysate in an upscaled microbial electrolysis cell with stacked bio-electrodes, Bioresour. Technol. **320** (2021) 124314. doi: https://doi.org/10.1016/j.biortech.2020.124314
152. *Pradhan, D., Jaiswal, S., Jaiswal, A. K.*, Artificial neural networks in valorization process modeling of lignocellulosic biomass. Biofuels Bioprod. Bioref. **16** (2022) 1849. doi: https://doi.org/10.1002/bbb.2417
153. *Pancholi, M. J., Khristi, A., Athira, K. M., Bagchi, D.*, Comparative analysis of lignocellulose agricultural waste and pre-treatment conditions with FTIR and machine learning modeling, Bioenerg. Res. **16** (2022) 123. doi: https://doi.org/10.1007/s12155-022-10444-y
154. *Liu, J., Jin, S., Bao, C., Sun, Y., Li, W.*, Rapid determination of lignocellulose in corn stover based on near-infrared reflectance spectroscopy and chemometrics methods, Bioresour. Technol. **321** (2021) 124449. doi: https://doi.org/10.1016/j.biortech.2020.124449
155. *Uddin, M. N., Ahmed, S., Ray, S. K., Islam, M. S., Quadery, A. H., Jahan, M. S.*, Method for predicting lignocellulose components in jute by transformed FT-NIR spectroscopic data and chemometrics, Nord. Pulp Pap. Res. J. **34** (2019) 1. doi: https://doi.org/10.1515/npprj-2018-0018
156. *Kartal, F., Özveren, U.*, An improved machine learning approach to estimate hemicellulose, cellulose, and lignin in biomass, Carbohydr. Polym. **2** (2021) 100148. doi: https://doi.org/10.1016/j.carpta.2021.100148
157. *Liu, X., Feng, X., He, Y.*, Rapid discrimination of the categories of the biomass pellets using laser-induced breakdown spectroscopy, Renew. Energy **143** (2019) 176. doi: https://doi.org/10.1016/j.renene.2019.04.137
158. *Gao, W., Zhou, L., Liu, S., Guan, Y., Gao, H., Hu, J.*, Machine learning algorithms for rapid estimation of holocellulose content of poplar clones based on Raman spectroscopy, Carbohydr. Polym. **292** (2022) 119635. doi: https://doi.org/10.1016/j.carbpol.2022.119635
159. *Dharmalingam, B., Tantayotai, P., Panakkal, E. J., Cheenkachorn, K., Kirdponpattara, S., Gundupalli, M. P., Cheng, Y.-S., Sriariyanun, M.,* Organic acid pretreatments and optimization techniques for mixed vegetable waste biomass conversion into biofuel production, BioeEnerg. Res. **16** (2022) 1667. doi: https://doi.org/10.1007/s12155-022-10517-y
160. *Aruwajoye, G. S., Faloye, F. D., Kassim, A., Saha, A. K., Kana, E. G.*, Intelligent modelling of fermentable sugar concentration and combined severity factor (CSF) index from pretreated starch-based lignocellulosic biomass, Biomass. Conv. Bioref. 2022. doi: https://doi.org/10.1007/s13399-022-03013-y
161. *Xu, H., Dong, C., Wang, W., Liu, Y., Li, B., Liu, F.*, Machine learning prediction of deep eutectic solvents pretreatment of lignocellulosic biomass, Ind. Crops Prod. **196** (2023) 116431. doi: https://doi.org/10.1016/j.indcrop.2023.116431
162. *Parkhey, P., Ram, A. K., Diwan, B., Eswari, J. S., Gupta, P.*, Artificial neural network and response surface methodology: A comparative analysis for optimizing rice straw pretreatment and saccharification, Prep. Biochem. Biotechnol. **50** (2020) 768. doi: https://doi.org/10.1080/10826068.2020.1737816
163. *Löfgren, J., Tarasov, D., Koitto, T., Rinke, P., Balakshin, M., Todorović, M.*, Machine learning optimization of lignin properties in green biorefineries, ACS Sustainable Chem. Eng. **10** (2022) 9469. doi: https://doi.org/10.1021/acssuschemeng.2c01895
164. *Ethaib, S., Omar, R., Mazlina, M. K. S., Radiah, A. B. D., Syafiie, S.,* Development of a hybrid PSO–ANN model for estimating glucose and xylose yields for microwave-assisted pretreatment and the enzymatic hydrolysis of lignocellulosic biomass, Neural Comput. Applic. **30** (2018) 1111. doi: https://doi.org/10.1007/s00521-016-2755-0
165. *Gama, R., Van Dyk, J. S., Burton, M. H., Pletschke, B. I.*, Using an artificial neural network to predict the optimal conditions for enzymatic hydrolysis of apple pomace, 3 Biotech. **7** (2017) 138. doi: https://doi.org/10.1007/s13205-017-0754-1
166. *Phromphithak, S., Onsree, T., Tippayawong, N.*, Machine learning prediction of cellulose-rich materials from biomass pretreatment with ionic liquid solvents, Bioresour. Technol. **323** (2021) 124642. doi: https://doi.org/10.1016/j.biortech.2020.124642
167. *Luo, H., Gao, L., Liu, Z., Shi, Y., Xie, F., Bilal, M., Yang, R., Taherzadeh, M. J.*, Prediction of phenolic compounds and glucose content from dilute inorganic acid pretreatment of lignocellulosic biomass using artificial neural network modeling, Bioresour. Bioprocess. **8** (2021) 134. doi: https://doi.org/10.1186/s40643-021-00488-x
168. *Katongtung, T., Onsree, T., Tippayawong, N.,* Machine learning prediction of biocrude yields and higher heating values from hydrothermal liquefaction of wet biomass and wastes, Bioresour. Technol. **344** (2022) 126278. doi: https://doi.org/10.1016/j.biortech.2021.126278
169. *Shafizadeh, A., Shahbeig, H., Nadian, M. H., Mobli, H., Dowlati, M., Gupta, V. K., Peng, W., Lam, S. S., Tabatabaei, M., Aghbashlo, M.,* Machine learning predicts and optimizes hydrothermal liquefaction of biomass, Chem. Eng. J. **445** (2022) 136579. doi: https://doi.org/10.1016/j.cej.2022.136579
170. *Zhou, X., Zhao, J., Chen, M., Wu, S., Zhao, G., Xu, S.,* Effects of hydration parameters on chemical properties of biocrudes based on machine learning and experiments, Bioresour. Technol. **350** (2022) 126923. doi: https://doi.org/10.1016/j.biortech.2022.126923

171. *Nag, A., Gerritsen, A., Doeppke, C., Harman-Ware, A. E.,* Machine learning-based classification of lignocellulosic biomass from pyrolysis-molecular beam mass spectrometry data, Int. J. Mol. Sci. **22** (2021) 4107. doi: https://doi.org/10.3390/ijms22084107
172. *Hai, A., Bharath, G., Patah, M. F. A., Wan Daud, W. M. A., Rambabu, K., Show, P., Banat, F.*, Machine learning models for the prediction of total yield and specific surface area of fee derived from agricultural biomass by pyrolysis, Environ. Technol. Innov. **30** (2023) 103071. doi: https://doi.org/10.1016/j.eti.2023.103071
173. *Liu, J., Jia, H., Mairaj Deen, K., Xu, Z., Cheng, C., Zhang, W.*, Application of machine learning methods for lignocellulose biomass pyrolysis: Activation energy prediction from preliminary analysis and conversion degree, Fuel **343** (2023) 128005. doi: https://doi.org/10.1016/j.fuel.2023.128005
174. *Kumar, N. P., Vijayabaskar, S., Murali, L., Ramaswamy, K.*, Design of optimal elman recurrent neural network based prediction approach for biofuel production, Sci. Rep. **13** (2023) 8565. doi: https://doi.org/10.1038/s41598-023-34764-x
175. *Fischer, J., Lopes, V. S., Cardoso, S. L., Coutinho Filho, U., Cardoso, V. L.*, Machine learning techniques aplied to lignocellulosic ethanol in simultaneous hydrolysis and fermentation, Braz. J. Chem. Eng. **34** (2017) 53. doi: https://doi.org/10.1590/0104-6632.20170341s20150475
176. *Pereira, R. D., Badino, A. C., Cruz, A. J. G.*, Framework based on artificial intelligence to increase industrial bioethanol production, Energ. Fuels **34** (2020) 4670. doi: https://doi.org/10.1021/acs.energyfuels.0c00033
177. *Long, F., Liu, H.,* An integration of machine learning models and life cycle assessment for lignocellulosic bioethanol platforms, Energy Convers. Manag. **292** (2023) 117379. doi: https://doi.org/10.1016/j.enconman.2023.117379
178. *Smuga-Kogut, M., Kogut, T., Markiewicz, R., Słowik, A.*, Use of machine learning methods for predicting amount of bioethanol obtained from lignocellulosic biomass with the use of ionic liquids for pretreatment, Energies **14** (2021) 243. doi: https://doi.org/10.3390/en14010243
179. *Konishi, M.*, Bioethanol production estimated from volatile compositions in hydrolysates of lignocellulosic biomass by deep learning, J. Biosci. Bioeng. **129** (2020) 723. doi: https://doi.org/10.1016/j.jbiosc.2020.01.006
180. *Wang, Z., Peng, X., Xia, A., Shah, A. A., Yan, H., Huang, Y., Zhu, X., Zhu, X., Liao, Q.*, Comparison of machine learning methods for predicting the methane production from anaerobic digestion of lignocellulosic biomass, Energy **263** (2023) 125883. doi: https://doi.org/10.1016/j.energy.2022.125883
181. *Sonwai, A., Pholchan, P., Tippayawong, N.*, Machine learning approach for determining and optimizing influential factors of biogas production from lignocellulosic biomass, Bioresour. Technol. **383** (2023) 129235. doi: https://doi.org/10.1016/j.biortech.2023.129235
182. *Chien, I., Enrique, A., Palacios, J., Rega, T., Keegan, D., Carter, D., Tschiatschek, S., Nori, A., Thieme, A., Richards, D., Doherty, G., Belgrave, D.*, Machine learning approach to understanding patterns of engagement with internet-delivered mental health interventions, JAMA Netw. Open **3** (2020) e2010791. doi: https://doi.org/10.1001/jamanetworkopen.2020.10791
183. *Dineva, K., Atanasova, T.*, Systematic look at machine learning algorithms – advantages, disadvantages and practical applications, International Multidisciplinary Scientific GeoConference, Sofia, Bulgaria, **2.1** (2020) 317-324.
184. *Niazian, M., Niedbała, G.,* Machine learning for plant breeding and biotechnology, Agriculture **10** (2020) 436. doi: https://doi.org/10.3390/agriculture10100436