Sustainable Chemistry for Climate Action 8 (2026) 100202

Contents lists available at ScienceDirect

Sustainable Chemistry for Climate Action

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

Latest trends and dynamics in green, sustainable and smart biorefineries

Lotta Hohrenk-Danzouma, Adrián Fuente-Ballesteros, Vânia G. Zuin Zeidler *

Institute of Sustainable Chemistry, School of Sustainability, Leuphana University Lüneburg, 21335, Lüneburg, Germany

ARTICLE INFO

Keywords:
Smart biorefinery
Critical bibliometric analysis
Green chemistry
Sustainable chemistry
Machine learning
Artificial intelligence

ABSTRACT

Smart biorefineries, integrating artificial intelligence (AI), machine learning (ML), chemometrics, and digital monitoring technologies, have gained attention as tools to address limitations in large-scale biorefinery implementation. Here, a decade-long (2014–2025) bibliometric analysis was conducted to evaluate their evolution and identify current trends. The results indicate an increase in research activity since approximately 2019, with AI and ML applications becoming progressively more frequent. Smart methods are applied across eight main purposes, including process modelling, optimisation, parameter selection, sample characterization, decision making, climate or spatial monitoring, and automation. Artificial neural networks are the most frequently applied tools, together with established chemometric approaches such as principal component analysis and response surface methodology, while methods such as random forest and support vector machines have gained relevance in recent years. The analysed applications mainly target lignocellulosic biomass, agricultural residues, and microalgae, focusing on the production of platform chemicals and biofuels. The reviewed studies show that smart methods can support process optimisation and predictive control and may contribute to sustainability-oriented strategies. However, most applications are concentrated in process optimisation and modelling, with limited implementation in automation and system-level integration, and only a minority of studies incorporate quantitative sustainability assessment tools such as life cycle analysis or techno-economic evaluation. Key challenges remain related to data availability and standardisation, as well as limited consideration of model robustness, interpretability, and environmental trade-offs.

1. Introduction

The world faces numerous interwoven sustainability issues such as environmental pollution, climate change, waste disposal, natural resource depletion and food insecurity. Circular bioeconomy can play an important role in addressing several of these interconnected crises by integrating renewable resources with circular economy principles to sustainably produce goods, energy and materials while minimising or ideally eliminating waste and environmental impacts [1,2]. With improved resource efficiency, the dependence on fossil resources can be reduced and greenhouse gas emissions can be lowered. As is well known, the establishment of green chemistry, and also sustainable chemistry laid a fundament for the further development of the biorefinery concept [3]. Biorefineries are the major facilitating strategies to transform biomass from numerous sources such as waste and by-products from agro-industry, forestry, aquaculture and fishery into biofuels, bioactive chemicals, and biomaterials [4,5]. The concept of biorefineries became popular in the scientific literature in the late 1990s, focusing on the integrated processing of renewable resources for the production of a range of bio-based products and energy [6,7]. Since then, the biorefinery approach has evolved and expanded, becoming a central theme in sustainable development and renewable energy research.

The *Biorefinery Action Plan* by the European Commission 2005 further popularised and defined biorefinery, within the European Union's renewable energy and sustainability strategies [8]. Different trends have shaped biorefinery research over time. Initially, integrated biorefinery systems showcased closed-loop systems where all biomass components were valorised, minimising waste or hybrid systems,

Abbreviations: ANN, artificial neural networks; AI, artificial intelligence; COSMO-RS, conductor-like screening model for real solvents; DES, deep eutectic solvents; DNN, deep neural network; DoE, design of experiment; DOM, desirability optimization method; DT, decision tree; GA, genetic algorithm; GIS, geographic information systems; GPU, graphics processing unit; IoT, internet of things; LCA, life cycle analysis; MD, molecular dynamics; ML, machine learning; PCA, principal component analysis; PLS, partial least squares; QC, quantum chemistry; RF, random forest; RSM, response surface methodology.

* Corresponding author.
E-mail addresses: lotta.hohrenk-danzouma@hamburgwasser.de (L. Hohrenk-Danzouma), adrian.fuente_ballesteros@leuphana.de (A. Fuente-Ballesteros), vania.zuin@leuphana.de (V.G. Zuin Zeidler).

https://doi.org/10.1016/j.scca.2026.100202
Received 10 December 2025; Received in revised form 9 April 2026; Accepted 11 April 2026
Available online 12 April 2026

combining biochemical and thermochemical processes to optimise resource use and product diversity. More recently, waste valorisation concentrating more on converting agricultural and industrial waste into valuable products has been discussed [4] and the latest development is the concept of smart biorefineries that integrate digital technologies, such as internet of things (IoT) and artificial intelligence (AI) to optimise processes, enhance energy efficiency, and monitor sustainability metrics in real-time [9].

Biorefineries are often complex systems including chemical, physical and biological interactions that require optimising different parameters, with varying conditions (e.g., biomass properties). Selecting optimal operational conditions, as well as monitoring process and product quality can become demanding and it is in many cases still a hindrance to a large-scale application [10]. A crucial factor for the success of biorefinery concepts compared to conventional fossil fuel-based processes is cost competitiveness [9]. With this in mind, biorefineries should be built on the basis of a production and business model that makes efficient use of available resources, as non-edible waste or unavoidable food ones, for instance, and an optimal logistics network for the highest possible production yield, lower costs and a lower environmental impact.

Chemometric tools have been successfully used to optimise different analytical methods. For example, with different design of experiments (DoE) approaches, parameters can be optimised with a limited number of experiments to reduce the time-consuming and expensive experimental work. However, recent rapid developments in the field of AI and machine learning (ML) have opened up new possibilities for data processing and raised hopes of finding comprehensive solutions to existing difficulties, thereby driving the market penetration of biorefinery concepts. AI tools can analyse large volumes of data in a very short time, identify patterns and provide accurate results that could help to increase the productivity of the bio-based process. In the initial stages of developing new biorefinery production models, such tools can aid in finding optimal solutions of parameter settings estimated through the interrelationship and combined evaluation of all relevant variables. In the operational phase, they can also be useful for forecasting production results that are difficult to predict, helping to implement logistics and automation effectively, and anticipating errors and anomalies that favour immediate corrective action [9]. According to Culaba et al. 2022 [11] to create a smart biorefinery system, AI models that partially or completely control operations need to be embedded. Integrating industrial-grade microcomputers, multiple low-powered sensors, scanners or cameras into the system can provide real-time data acquisition, which are needed for decision-making systems. Current tendencies of a smart biorefinery system include a dense technology-cored system but through AI, specifically bio-inspired and evolutionary algorithms, which can be further optimized. In addition, computer vision combined with deep learning enables rapid, non-destructive (minimal-sample-preparation) identification of feedstock signatures, which are key inputs for decision-support systems.

In this context, the main aim of this work is to address recent advances related to biomass valorisation into renewable energy, chemicals and additional products through chemo-catalysis, bio-catalysis and digital strategies within smart biorefinery concepts. Although green and sustainable chemistry and, more recently, AI and ML have attracted increasing attention in biorefinery research, the existing literature remains highly fragmented. Most studies focus on individual case applications or specific algorithms, while a systematic and critical overview linking AI/ML methods to their concrete roles along the biorefinery workflow is still lacking. It remains unclear how different smart approaches are distributed across process modelling, optimisation, evaluation, decision support or sustainability assessment, and how these applications relate to feedstocks and products. To address this gap, a decade-long (2014–2025) bibliometric analysis was conducted to examine the evolution of biorefinery research trends and to verify the rise of smart biorefinery concepts as a distinct research direction. The use of AI and ML tools is analysed, and their application purposes are categorised, providing a structured overview of methods and use cases. This approach enables a critical comparison of current practices and supports the identification of limitations and future research priorities in smart and sustainable biorefineries.

2. Methodology for bibliometric analysis

2.1. Data collection

Scopus, Web of Science, Science Direct and Google Scholar were used as databases to search for references. Table 1 shows an overview of data collection and filtering procedure. Data set A includes all publications including the term *biorefiner** (e.g., root term for biorefinery, biorefineries) to employ analysis on the general trends in the research field. For data set B, the search was narrowed down with the search terms shown in Table 1 to analyse smart biorefinery and green and sustainable concepts. For both data sets only English publications within a ten-year period (2014–2025) were included. Review articles were not considered as these already display summaries of applied methods that may lead to double counting in later bibliometric analysis.

In a next step, all abstracts and, when necessary, full texts were manually screened by the authors following predefined exclusion criteria; 1) out of scope: for example, some publications did not apply any ML methods, but only mentioned them in some way or used the term smart in another context; 2) reviews/opinion paper: even though review articles were excluded previously in the search, several reviews or general opinion paper remained in the data set that had to be excluded manually. In this way, it was assured that only studies with practical implementation of ML and biorefinery approaches were considered for data evaluation. 29 publications were excluded from the list and 44 publications remained for further literature analysis. However, reviews or general discussion papers were considered for general discussion of the topic.

2.2. Systematic literature search

To analyse research trends over the past 10 years and observe the upcoming tendency of topics, the occurrence of selected keywords in the title, authors keywords or abstracts of publications was systematically searched. Each publication was counted only once, also if the keyword was mentioned several times in the title, keywords and abstract. For time trend analysis, the publication years were extracted and the mean,

Table 1
Summary of search strategy and resulting publication sets.

Data set	Search terms	Filters applied	Publications (N)	Purpose
A	Biorefiner*	Years 2014–2025; English language; Articles only (no reviews)	22,337	Bibliometric analysis
B	Biorefiner* AND (“machine learning” OR “AI” OR “artificial intelligence” OR “smart” OR “big data” OR “chemometr*”) AND (green* OR sustain*)	Years 2014–2025; English language; Articles only (no reviews)	74	Bibliometric analysis
C	Subset selected from Data Set B after abstract/full-text screening	Exclusion criteria: out of scope; reviews removed	44	Categorisation and literature analysis

median, minimum, maximum, first- and third quartiles (Q1/Q3) were calculated to create boxplots. The boxplot representation gives the possibility to visualise the timewise development of the usage of certain search terms in the literature, which is assumed to correlate to the application of the corresponding algorithm, methodology, biomass resource or biorefinery product in the studies. The min and max show when a term first occurred in the literature and if it is still used. The position of the box, limited by the first and third quartile, gives an impression of the time span in which a term was mainly used and the mean and median show at what time a certain term was most popular. The frequency of occurrence of the terms is indicated by colouring. The selection of keywords was based on the initial analysis with the R bibliometrix package [12] and the trend topic analysis algorithm. The search terms that were most often mentioned on general biorefinery trends were used for data set A. For data set B, different terms associated to different chemometric, and AI methods were selected.

2.3. Literature analysis and categorization

To analyse the application of different AI or ML methods in detail, all 44 publications of data set C were categorized. First, all methods, that were used were listed and number of applications were counted. Different categories on the purpose of the application of the methods were identified, and all publications were assigned to these: process modelling, evaluation and optimization, parameter selection, sample characterization, decision making, climate/spatial monitoring and automation. In a next step, further categories for used resources (Macroalgae; Animal Fat; Food/Municipal Waste; Waste Oil; Microalgae; Agricultural Residues; Biomass; Lignocellulose) and products (Platform Chemical; Biofuel; Biodiesel, Biochar, Bioactive compounds; Energy; Materials) were identified and assigned to the publications. Then, aspects of sustainability of the proposed biorefinery concepts and the role of smart methods to enhance this aspect were analysed. Additionally, a geographical trends of country collaboration in the research field was performed using R bibliometrix package [12]. Data set A and B were uploaded separately to the shiny application of R package and evaluated with country collaboration function.

3. Results

3.1. Bibliometric trend analysis

To analyse general trends in biorefinery research, all the studies found in the databases with the search term *biorefiner** totalling 22,337 publications, were selected. Different terms in the title, abstract or keywords were studied as described above. The boxplots in Fig. 1 illustrate how often (colour-bar) and when (box-plots) a certain term was mentioned in the scientific literature.

The term *Sustainability* was mentioned with a high frequency, indicating its importance in the field. However, the general sustainability of a biorefinery concept can be derived from different aspects, as the use of renewable resources or green extraction methods. To have a closer look at the role of green and sustainable chemistry methods, different pairing words were analysed. The pairing of terms *Green* with *Chemistry* or *Extraction* shows a higher frequency compared to those in combination with *Sustainable.* However, the term *Sustainable Extraction* seems to have a trending occurrence frequency (median 2023), similar to the term *Ecofriendly*, while the latter is used much more often.

The term *Integrated Biorefinery* shows a similar time distribution as *Biorefinery*, indicating that it has been an important aspect since the beginning. Analysing the terms *Circular Bioeconomy* and *Circular Economy,* it can be noted that *Circular Economy* is used with a higher frequency, but *Circular Bioeconomy* shows a trending tendency. The high frequency of the terms *Valorisation* and *Value-added* shows that these are key aspects in the research field. *Waste Valorisation* and *Zero Waste* show a trending appearance in the literature with a median in 2023. However, *Big Data, Smart and Artificial Intelligence* are not yet used with a high frequency yet; *Big Data* appeared first in 2019 and shows an occurrence like AI, indicating that this is very recent development and not an established approach in the literature.

To gain a deeper insight into the trends on the aspect of the “smartness” of biorefinery approaches, various keywords associated with this field and frequently applied methods were subjected to a time trend analysis. The smart biorefinery literature shows, as expected, a much more dynamic development compared to the general data set on

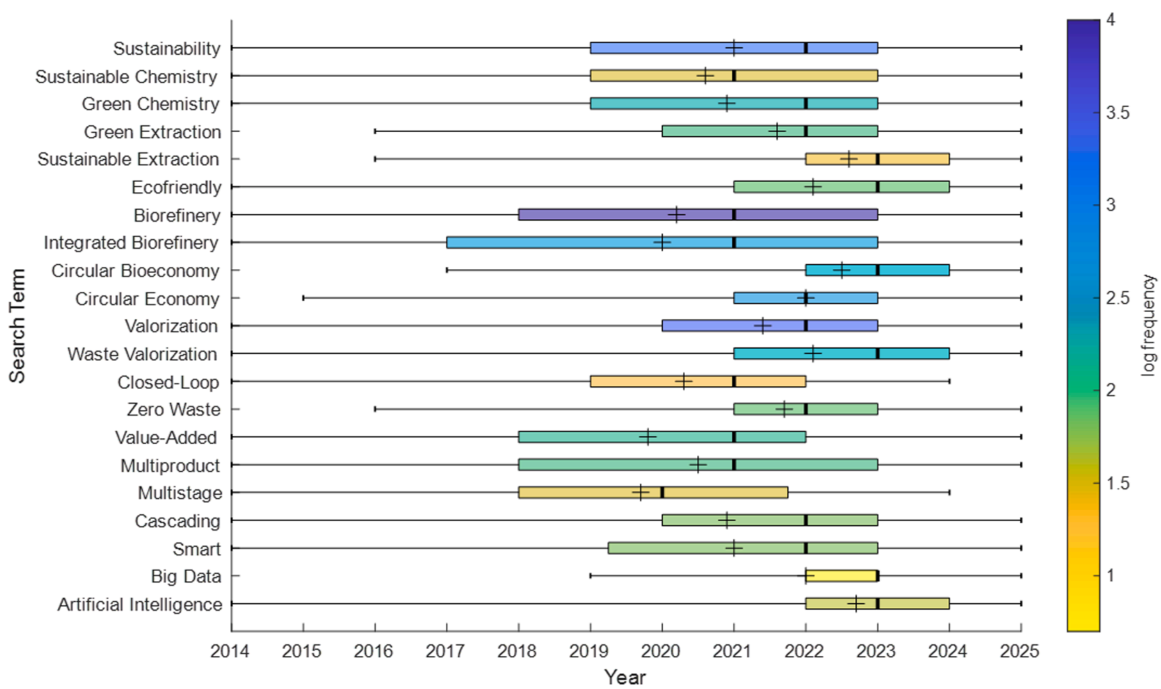

Fig. 1. General trends of biorefinery research. The boxplot shows the timewise development of occurrence in the scientific literature and the colour bar indicates occurrence frequency. Minor variations in the most recent year may reflect database indexing timelines.

biorefineries (see Fig. 1). Currently, with the increasing digitisation of industry and the rapid development of data science, new approaches using AI or ML have come to the fore. This is shown in Fig. 2 as trending terms regarding time distribution and frequency are AI and ML. In addition, these terms have only appeared in the literature more recently e.g., ML or Big Data in 2019. While terms as *Chemometrics* and more classical methods such as *Principal Component Analysis* (*PCA)* have been mentioned since 2014, this is not the case for ML and AI; in fact ML is a subset of AI, aiming to train a machine to solve a problem by itself using data obtained from various sources, including data collected over time and statistical analysis of that information. As shown in Fig. 2 more recently different ML methods such as *Random Forest* (2021) or *Support Vector Machines* (2022) appeared.

Classical chemometric methods have always played an important role in the development of biorefinery concepts to analyse complex multivariate data sets. For example, different regression or clustering approaches have been applied for feedstock characterisation or for process monitoring to identify patterns or outliers in reaction conditions [13–15]. This is also visible in the time trend analysis in Fig. 2 as terms like *Multivariate Analysis* (2015), *Cluster Analysis* (2016), *Classification* (2017), have appeared over the years in the scientific literature.

3.2. Smart methods used

The distinction between ML and chemometrics is neither sharp nor used consistently in the literature, as ML tools can also be used for classification, regression and clustering tasks. However, in comparison to classical chemometric methods, ML is more flexible and powerful, capable of learning complex, non-linear relationships from data, then adapting and improving its models from experience. Therefore, it is applied for more sophisticated tasks in biorefinery research such as predicting yields, optimising conversion processes, and identifying novel pathways in more complex biological systems [9]. Fig. 3 presents an overview of different algorithms that are commonly used, which can be categorised as either supervised or unsupervised tools. Supervised methods use labelled data to build models that predict known outcomes, while unsupervised methods analyse unlabelled data to uncover patterns or groupings without predefined categories.

PCA is a dimensionality reduction technique that transforms a data set into a set of orthogonal components, capturing the maximum variance to simplify analysis while preserving essential patterns [16]. DoE techniques can be used to model the relationship between several independent variables and one or more response variables to find optimal conditions for a process. Response surface methodology (RSM) is often used to optimise reaction conditions or extraction processes by identifying the best combinations of factors, such as e.g. temperature, pH, pressure, etc. to maximise product yields [17,18].

Artificial neural networks (ANNs) are computing systems inspired by the biological neural networks of human brains. They comprise interconnected nodes or neurons that are structured into layers (input, hidden and output), in which each node has a set of outputs and inputs from the former to the next layer. For biorefinery process optimisation, the number of neurons in the input layer is directly proportional to the number of process parameters utilised during training to make predictions about the external environment. The output layer can be represented by the neurons measuring the treatment efficacy parameters or, e.g. the biomass production/growth [19].

The advantages of ANNs are that they are non-linear, unconstrained, very adaptable and fault tolerant. Their disadvantage of being a “black box” model is their lack of ability to provide an explanation for their reasoning and sometimes produce unexpected results [19]. Support vector machines (SVMs) are a type of supervised learning algorithm used for classification and regression tasks. The algorithm uses a nonlinear mapping function to project data into a high-dimensional feature space. It then constructs a decision boundary, known as a hyperplane, which separates the space into different classes, allowing for the classification of unknown data points. The algorithm identifies support vectors, data points that lie closest to the optimal hyperplane, which help maximise the margin between the hyperplane and the nearest samples in the high-dimensional space. These support vectors are key in defining the optimal decision boundary [10]. Decision tree (DT) and random forest (RF) algorithms are supervised learning methods that are widely used in both classification and regression tasks. The DT processes the data based on a concept tree, which summarises the given information of the training data set in a tree structure. A RF model is made up of a collection of DTs. The results of these DTs for a given data point are then combined, using majority voting for classification and averaging for regression tasks, to produce a prediction to achieve a performance that is better than the performance of each individual DT alone [20]. Table 2 gives an overview of different methods that were used in the analysed publications (data set C) and indicates if it is a supervised or unsupervised method and the main or general

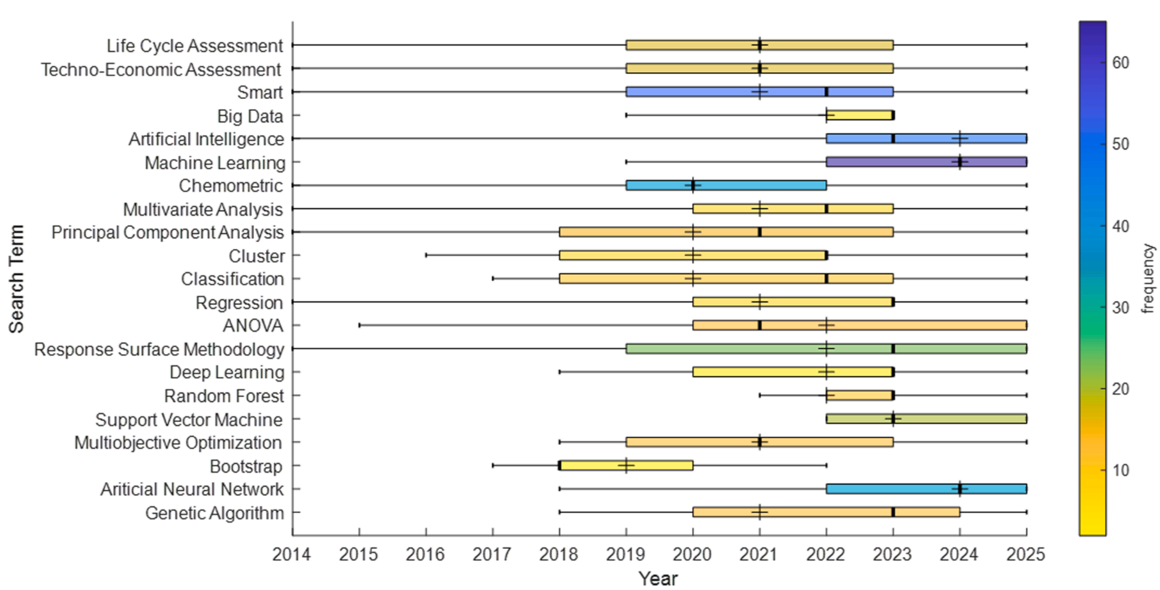

Fig. 2. Trends of smart biorefinery research. The boxplot shows the timewise development of occurrence in the scientific literature and the colour bar indicates occurrence frequency.

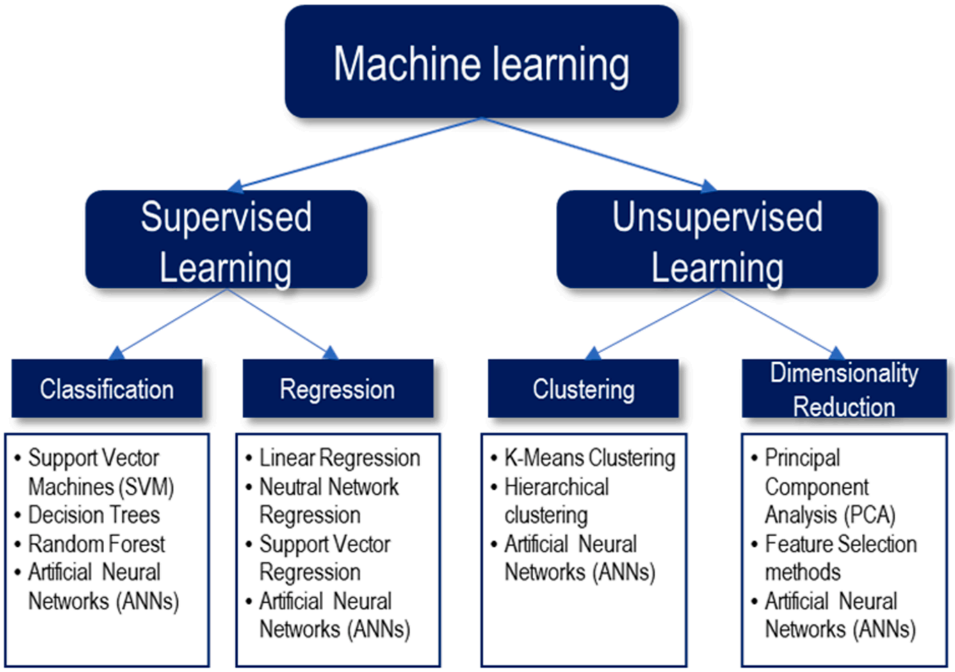

Fig. 3. Overview of the ML algorithm based on Tauler et al. (2009) [16]. Note that reinforcement learning is an additional category of the ML algorithm that was not covered here.

function. As can be seen in Table 2 ANNs were among the most frequently applied approaches. In addition, several well-established chemometric techniques, such as PCA and RSM, were also commonly used.

Beyond the overview of algorithms, the critically analysed literature reveals that individual algorithms can be applied to multiple purposes, while different analytical methods can also address similar biorefinery challenges, reflecting the methodological flexibility of smart biorefinery systems. ANN-based approaches represent the most frequently applied tools, particularly in process optimisation and predictive modelling, which can be explained by their ability to capture complex and non-linear relationships characteristic of biorefinery systems where multiple interacting variables influence process efficiency, yield, and product distribution. Despite their strong predictive performance, only a limited number of studies address aspects related to model interpretability, uncertainty evaluation, or transferability to industrial-scale applications, which remain important challenges for practical implementation. At the same time, classical chemometric and experimental design approaches such as PCA and RSM continue to play a fundamental role, demonstrating that smart biorefineries typically rely on hybrid analytical frameworks that combine established statistical tools with advanced ML techniques. These approaches are frequently used for preliminary variable selection, dimensional reduction, or optimisation prior to applying more computationally intensive predictive models. Furthermore, supervised learning algorithms such as RF, SVM, and gradient boosting show increasing relevance, especially in feedstock classification, parameter selection, and yield prediction, reflecting a gradual transition towards data-driven decision-support strategies.

3.3. Purpose and areas of application of smart methods in biorefineries

A variety of different methods have been used for very different purposes: some to select optimal processing parameter others to characterize samples, monitor or even fully automate a biorefinery operation. To explore the various possible applications of smart methods, the publications were assigned to different categories for the “purpose” as well as for the “resources” used and resulting “products”. Table 3 shows the categories for all 44 publications (data set C) as well as the used methods, as noted some publications have been assigned to several categories.

3.3.1. Purpose

The main purposes of the smart methods described in recent literature were process optimization, followed by parameter selection, sample characterization, decision making, process modelling, climate or spatial modelling, process evaluation and automation (see Fig. 4). In the following different examples, the application of ML will be given to show its potential.

Process optimization. The operation of biorefineries often must deal with the changing nature of biomass properties, biomass feedstock supply, ambient temperature, product demands or other conditions. A new optimal resource allocation scheme was developed for a multiproduct lignocellulosic biorefinery to instantly detecting any fluctuations in the biorefinery to avoid oversupply or undersupply issues. The proposed resource allocation system based on deep neural network (DNN) uses the parameters that undergo fluctuations as input nodes of the DNN, while the output nodes are the flowrate allocation of biomass to different chemical and energy conversion pathways [51]. Another study utilized Bayesian Optimization for tuning a biorefinery to produce lignin-carbohydrate complexes, and achieved by iteratively gathering data and examine the effects of key processing conditions and finding optima trade-offs through Pareto front analysis [45]. The production of biodiesel from animal fat is optimized using a multi-objective mathematical model [48].

Parameter selection. The optimal parameter for homogenizer-assisted extraction of bioactive compounds from mango waste were determined by RSM [34]. An ANN combined with genetic algorithm (GA) optimization tool was employed for predicting optimal process conditions for enhancing the biomass of the green microalgae [31].

Sample characterization. For sample characterization, a study applied PCA and partial least squares (PLS) to connect physicochemical characteristics of mangoes with their chemical composition, especially

Table 2
Overview of AI / data analysis methods in the reviewed literature.

Method	Acronym	Type	Function	Fr. *
Adaptive Neuro-Fuzzy Inference System	ANFIS	Supervised	Classification/ Regression	3
Analysis of Variance	ANOVA	Supervised	Statistical Testing	1
Artificial Neural Network	ANN	Supervised	Classification/ Regression	11
Bayesian Optimization	—	Unsupervised	Optimization	2
Box-Behnken	—	Supervised	Experimental Design	3
Classification Trees	—	Supervised	Classification	1
Cluster Analysis	—	Unsupervised	Clustering	3
Conductor-like Screening Model for Real Solvents	COSMO-RS	Unsupervised	Solvent Screening	1
Decision Tree	DT	Supervised	Classification/ Regression	3
Deep Learning	DL	Supervised	Classification/ Regression	1
Deep Neural Network	DNN	Supervised	Classification/ Regression	2
Desirability Optimization Method	DOM	Unsupervised	Optimization	1
Generative Adversarial Networks	GAN	Unsupervised	Data Generation	1
Genetic Algorithm	GA	Unsupervised	Optimization	5
Geographic Information Systems	GIS	—	—	1
Gradient Boost	—	Supervised	Classification/ Regression	2
Kernel Density Estimation	—	Unsupervised	Density Estimation	1
K-Means / Multidimensional Scaling	—	Unsupervised	Clustering / Dimensional Reduction	2
Least-Squares Boosting	LSBoost	Supervised	Classification/ Regression	1
Least-Squares Support Vector Machine	LSSVM	Supervised	Classification/ Regression	2
Life-Cycle Assessment	LCA	—	Environmental Impact	2
Linear Regression	LR	Supervised	Regression	2
Logistic Regression	—	Supervised	Classification/ Regression	2
Monte Carlo Simulation	—	Unsupervised	Simulation / Uncertainty	2
Multilayer Perceptron	MLP	Supervised	Classification/ Regression	4
Multi-Objective Optimization	MOO	Unsupervised	Optimization	3
Neutrosophic Fuzzy Optimization	—	Unsupervised	Optimization	1
Partial Least-Squares Regression	PLS	Supervised	Regression	3
Principal Component Analysis	PCA	Unsupervised	Dimension Reduction	7
Random Forest	RF	Supervised	Classification/ Regression	4
Response Surface Methodology	RSM	Supervised	Optimization	5
Sensors/IoT	—	Unsupervised	Data Acquisition	1
Sparse Nonlinear Optimizer	SNOPT	Unsupervised	Optimization	1
Support Vector Machines	SVM	Supervised	Classification	4
Support Vector Regression	SVR	Supervised	Regression	3
Techno-Economic Analysis	TEA	—	Cost-Benefit Analysis	3

* Frequency (Number of times a method was used).

bioactive compounds, in different cultivars [60]. Another work used an X-ray fluorescence-based technique and PLS to discriminate sugarcane juices in different clarification stages to estimate their mineral content, an impurity that diminishes the quality of the sample [56].

Decision making. A five-step approach supporting the decision making on biomass, processes and technologies for the bioconversion of lignocellulosic biomass was proposed employing big data and sustainability assessment [28]. Another study established a novel integrated approach for the solution of biomass supply chain network design using AI, geographic information systems, multi-criteria decision making and mathematical modelling [24]. For the decision on appropriate deep eutectic solvents (DES) for lignin fractionation conductor-like screening model for real solvents (COSMO-RS) model, quantum chemistry (QC) calculation, and molecular dynamics (MD) simulations were applied [29].

Process modelling. For modelling the pyrolysis reaction of *Saccharum bengalense* biomass a ANN regression model was applied [41]. Another study also used ANN and found an excellent performance for predicting effects of O/C and H/C ratios on the chemical exergy and standard entropy of technical lignin [43]. A study on Ozonolysis as pre-treatment for lignocellulosic biomass used gradient-based sparse nonlinear optimizer (SNOPT) method with sequential quadratic programming (SQP) algorithm to estimate kinetic reaction parameters are computationally [42].

Climate or spatial modelling. With an advanced forecasting algorithm, the climate conditions for a case study region were predicted to optimize a microalgae biorefinery. By that, the most feasible microalgae species were identified, and optimal harvest periods of cultivated microalgae were determined. Monte Carlo simulation was conducted for uncertainty analysis of the selected process [23]. Geographic information systems (GIS) have been used to find potential biorefinery sites [25] or biogas facility locations [24]. These studies highlight the growing relevance of climate and spatial modelling as supporting tools for smart biorefinery planning, particularly for feedstocks that are highly dependent on environmental and geographical conditions. Climate-driven modelling approaches allow the evaluation of seasonal variability, resource availability, and production stability, which are critical factors for ensuring continuous and efficient biomass supply. Similarly, spatial analysis using GIS enables the integration of environmental, logistical, and economic parameters to support strategic decision-making in facility location and biomass distribution. Despite these advantages, the analysed literature suggests that climate and spatial modelling applications remain less explored compared to optimisation-focused ML applications. The integration of predictive climate models with real-time environmental monitoring, remote sensing technologies, and supply chain optimisation models therefore represents an important opportunity for advancing smart and resilient biorefinery systems, particularly under the challenges imposed by climate change and resource variability.

Process evaluation. For the evaluation of biorefinery processes several studies use life cycle analysis (LCA) in combination with other ML methods. For example, to estimate LCA metrics classification trees and ANN have been applied to produce input-output relationships through predictor variables that refer to the molecular structure of bio-chemical or bio-fuel products of interest, the feedstocks used, and the process technologies characteristics [37]. Another study on the production of bioplastics uses LCA together with Monte Carlo simulation, RF and ANNs to analyse feedstock composition variations and uncertainties [39]. One study used a hybridization of data science techniques and environmental analysis to evaluate different pre-treatment processes of biomass (corn stalk and rice straw). They combine LCA with dimension reduction applying the Multidimensional Scaling and clustering

Table 3
Methods and assigned categories (Purpose, Resources and Product) of literature analysis.

#	Purpose	Smart Method	Feedstock	Product/Output	Ref.
1	Automation, process optimization	Sensors	Biomass	Chemicals	[21]
2	Automation, decision making, process optimization	Sensors	Waste	Chemicals, bioactive compounds	[22]
3	Climate/spatial modelling	Monte Carlo Simulation	Microalgae	Biofuel	[23]
4	Climate/spatial modelling, decision making	SVR, GIS	Waste, agricultural residues	Energy	[24]
5	Climate/spatial modelling, decision making	SVR, GIS, ANFIS	Agricultural residues, waste oil, animal fat	Biofuel	[25]
6	Climate/spatial modelling, process optimization	ANN, RSM, DOM	Waste	Chemicals	[26]
7	Decision making	MOO, MADM	Biomass	Ethanol	[27]
8	Decision making	Data queries (SQL requests) online analytical processing (OLAP)	Lignocellulose	Energy, chemicals, materials	[28]
9	Decision making	COSMO-RS	Lignocellulose	Biofuel, chemicals, materials	[29]
10	Parameter selection	LR, SVM, ANN, ANFIS, LSSVM, PCA, GA	Waste oil	Biodiesel	[30]
11	Parameter selection	ANN, GA	Microalgae	Biofuel	[31]
12	Parameter selection	Bayesian Optimization	Lignocellulose	Chemicals	[32]
13	Parameter selection	PLS	Macroalgae	Chemicals	[33]
14	Parameter selection	RSM	Agricultural residues	Bioactive compounds	[34]
15	Parameter selection, decision making	LSBoost, SVR	Microalgae	Biofuel	[35]
16	Process evaluation	LCA, k-means, multidimensional scaling	Lignocellulose, agricultural residues	Biofuel	[36]
17	Process evaluation	LCA, ANN, classification trees	Biomass	Biofuel, chemicals	[37]
18	Process evaluation	LCA, TEA	Lignocellulose	Chemicals	[38]
19	Process evaluation	Monte Carlo simulation, RF, ANN	Biomass	Materials	[39]
20	Process evaluation	TEA	Lignocellulose	Chemicals, biofuel	[40]
21	Process modelling	ANN, MLP, PCA	Biomass	Chemicals	[41]
22	Process modelling	SNOPT	Lignocellulose	Chemicals	[42]
23	Process modelling	ANN	Lignocellulose	Chemicals	[43]
24	Process modelling	LSSVM	Biomass	Energy	[44]
25	Process optimization	Bayesian Optimization	Lignocellulose	Chemicals	[45]
26	Process optimization	RSM, ANN, GA	Microalgae	Bioactive compounds	[46]
27	Process optimization	ANN	Microalgae	Biodiesel	[47]
28	Process optimization	MOO	Animal fat	Biodiesel	[48]
29	Process optimization	RSM, Box-Behnken, ANN, GA, ANFIS	Waste oil	Biodiesel	[49]
30	Process optimization	DL, Gradient Boost, DT, ANN, RF, SVM	Lignocellulose	Biofuel	[50]
31	Process optimization, parameter selection	DNN	Lignocellulose	Energy, chemicals	[51]
32	Process optimization, sample characterization	SVM, RF, Gradient Boost	Biomass	Biochar, chemicals	[52]
33	Process optimization	Sensors	Microalgae	Biomass	[53]
34	Process optimization	TEA	Lignocellulose	Chemicals	[54]
35	Process optimization	MOO, GA	Biomass	Energy	[55]
36	Sample characterization	PLS	Biomass	Ethanol	[56]
37	Sample characterization	LR	Macroalgae, biomass	Energy, chemicals	[57]
38	Sample characterization	PLS	Microalgae	Chemicals	[58]
39	Sample characterization	PCA, ANOVA	Agricultural residues	Bioactive compounds	[59]
40	Sample characterization	PCA, PLS, cluster analysis	Agricultural residues	Bioactive compounds	[60]
41	Sample characterization	PCA	Agricultural residues	Bioactive compounds	[61]
42	Sample characterization	Generative Adversarial Networks, Kernel Density Estimation	Lignocellulose	Chemicals, biochar	[62]
43	Sample characterization	RF, PCA, DNN	Microalgae	Chemicals	[63]
44	Sample characterization	PCA, minimum spanning tree (MST)	Biomass	Energy	[64]

technique using k-mean [36]. This approach enabled the detection of relationships between environmental impact categories and processing routes by transforming complex LCIA matrices into lower-dimensional representations. Consequently, it allowed a structured comparison of process sustainability and identification of environmentally critical pre-treatment technologies.

Automation. Just one study presented an AI system for the entirely automated fed-batch fermentation of 1,3-propanediol, including a sensor, predictor, controller, and automation system. They were able to successfully integrate AI to establish an automatic, robust system with enhanced 1,3-propanediol concentration. They could increases glycerol utilization efficiency and decreases the cost of the medium and were able to eliminate the dependence on expensive online instruments and staffing [21].

3.3.2. *Biorefinery applications*

In addition to the smart methods and their purpose of application, the type of biorefinery approaches were analysed. To this end, the nature of biomass used and the products generated were categorised and are shown in Fig. 5. Lignocellulosic biomass, followed by other biomasses (e.g. sugarcane) were the most cited (see Fig. 5A). The third category were agricultural residues (as grape seeds, mango processing waste), and microalgae. Further, waste oil, animal fat and food or municipal solid waste were considered in some studies as well as seaweed. The resulting end products of biorefinery are presented in Fig. 5B Mainly platform chemicals as aromatic compounds, acetone, alcohols, esters etc. were produced as well as biofuel in the form of biodiesel or bioethanol. In addition, value-added or bioactive compounds were retrieved and materials such as bioplastics. Energy was produced in the form of e.g., chemical exergy or electricity.

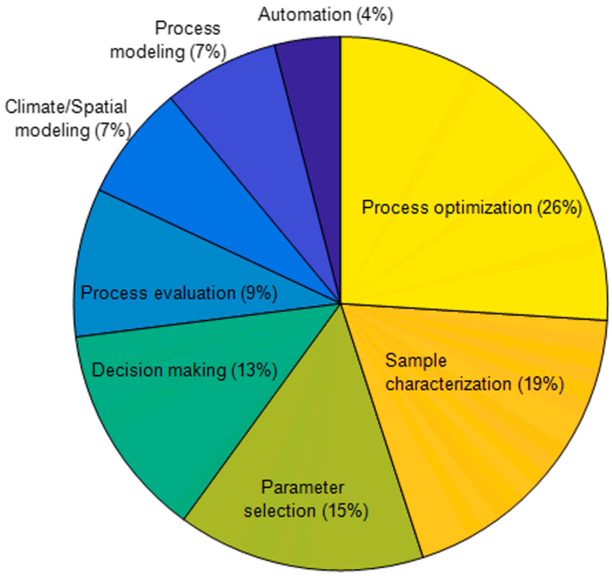

Fig. 4. Overview of different purposes for the application of smart data analysis methods (%).

3.3.3. Sustainability evaluation

Biorefineries have sustainability at their core as they transform renewable resources into high-value materials and/or renewable energy. However, there are certain developments in biorefinery research with the aim of increasing the sustainability of the concepts focusing on different aspects. Examples are using greener and more sustainable transformation process and technologies, using waste materials as feedstock to avoid competing food production or using cascading process or aiming at zero-waste concepts. Smart methods can also be used to enhance sustainability; thus, it was evaluated if and how the studied publications used smart methods to enhance sustainability by different criteria, results are summarized in Table 4.

Green and sustainable extraction and conversion processes. Different advanced extraction methods can be used to reduce time, solvents, energy consumption and improve the product quality and extraction yield compared to conventional methods. In addition, alternative solvents such as bio-based solvents, ionic liquids, and DES, which are often biodegradable and non-toxic can be used. In total nine publications fall into this category using microwave-assisted extraction [18], AquaSolv omni [32,45] microwave pyrolysis [52] homogenizer-assisted extraction [34] green catalyst made from *Brotia costula* shells [49] or using DES [29,54]. However, also many conventional conversion processes are applied in the analysed literature, showing that this aspect needs more attention in the future.

Waste utilization. The types of feedstocks can be used to establish a hierarchy of biorefineries, classified as first to third generations. First-generation biorefineries are based on raw materials, such as biomass from crops (e.g., corn, sugarcane, wheat, etc.), but are criticized for potentially competing with food production for land use. Second-generation biorefineries are based on raw material residues from crops, agroindustry or other waste materials such as municipal waste, animal fat, and waste oil. The third-generation biorefineries are based on raw materials related to microalgae and macroalgae [65]. Sixteen of the analysed papers explicitly mention the use of waste materials or residues. In addition, eight publications rely on microalgae as feedstock. Other publications do not mention a specific type of biomass and are thus also adaptable for waste utilization.

Multiproduct. Biorefineries where a single product is obtained from one type of biomass are also referred to as Stage I biorefineries. In a Stage II biorefinery, from a single raw material, many procedures are employed to obtain many significant products. Moreover, in a Stage III biorefinery, different raw materials are used [66]. A cascading biorefinery process involves sequentially converting biomass into multiple products in a stepwise manner, maximizing resource efficiency. Some studies focus on the product of biofuels (bioethanol, biohydrogen and biobutanol) and high-value products or platform chemicals [40] or value-added products or energy through different conversion pathways [51]. One study uses ML to facilitate a multiproduct lignocellulosic biorefinery that converts various types of biomass into value-added products and energy through different conversion pathways. A resource allocation system based on DNN is used detecting any fluctuations in the biorefinery to avoid oversupply or undersupply issues. The optimization results of the DNN yields an average optimality of 97.7 % and reduces the response time by 99.5 % as compared to the conventional nonlinear solver [51].

However, most studies focus on one type of products as biochar, bio-

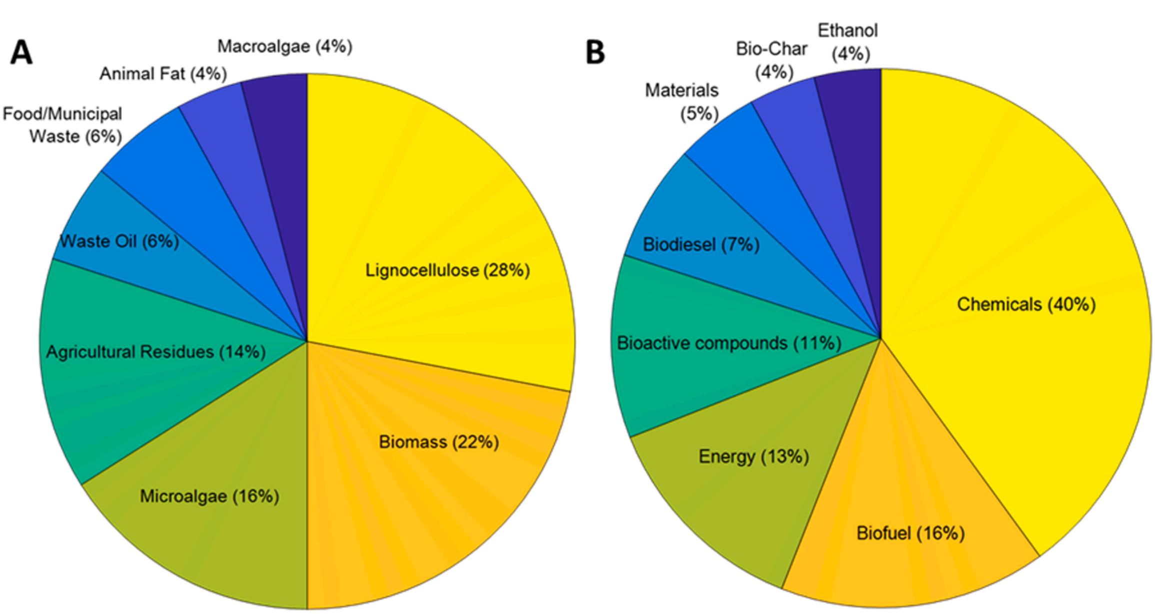

Fig. 5. Overview of proportions of different feedstocks (A) and products (B) of biorefinery concepts.

Table 4

Evaluation of the inclusion of various aspects relating to sustainability. A (green): green conversion methods; B (yellow): Waste utilization; C (orange): Multiproduct processes; D (blue): increased efficiency; E (purple): sustainability evaluation. Numbering of publications (#) as in Table 3. The assignment of publications to sustainability categories was performed through manual screening based on predefined criteria evaluating the presence of specific sustainability-related strategies.

#	A	B	C	D	E	Ref.	Year
1						22	2023
2		■	■			23	2023
3				■		24	2022
4		■				25	2023
5					■	26	2024
6		■			■	27	2023
7					■	28	2018
8		■	■			29	2019
9	■		■			30	2022
10		■		■		31	2022
11				■		33	2018
12	■	■				34	2022
13				■		35	2016
14	■	■				36	2020
15				■		32	2023
16		■			■	37	2022
17			■		■	38	2022
18					■	39	2025
19					■	40	2024
20		■	■			41	2025
21						42	2023
22	■					43	2021
23				■		44	2022
24				■		45	2024
25	■					46	2024
26	■			■		47	2024
27				■		48	2024
28		■		■		49	2024
29	■	■				50	2024
30		■				51	2023
31			■	■		52	2021
32	■					53	2022
33						54	2025
34	■	■		■		55	2025
35			■		■	56	2025
36				■		57	2022
37			■			58	2019
38						59	2020
39		■				60	2019
40		■				61	2024
41		■	■			62	2018
42					■	63	2022
43						64	2025
44						65	2025

oil, and syngas [52] or several bioactive compounds [34,46,61]. This shows that truly closed- loop systems that optimize biomass utilization and minimize environmental impact and waste generation are still challenging. Nevertheless, some publications see the potential of AI and ML to advance this field further [67].

Efficiency. Several studies use ML and AI tools to optimize a conversion process for a more efficient process and gaining higher yields. Fifteen papers report a higher efficiency for the optimized process that were possible by smart methods. For example, with an AI system, consisting of a sensor, predictor, controller, and automation system, for entirely automatic fed-batch fermentation of 1,3-propanediol a concentration increase of 75.7 % was possible [21]. Another study employs new microalgae biorefinery system with an integration of the hydrothermal liquefaction of wet microalgae and hydrodeoxygenation of biocrude oil to produce green products of green diesel, naphtha, and heavy oil. LSBoost model outperformed other ML-assisted modelling approaches and the performance of the microalgae biorefinery system increased with a green diesel yield up to 33 wt % and the energy recovery estimated at 75 % higher with optimized process [35].

Evaluation. Nine studies use AI or ML to facilitate or enhance sustainability analyses. For example, a data-science based framework with capabilities to estimate LCA metrics of bio-based and biorefinery processes in early design phases was developed [37]. Another study uses desirability optimization method (DOM) to improve the in-depth mechanistic understanding of environmental systems and identify the most benign configurations [26]. Another approach integrates sustainability, resilience, and responsiveness within the (lean, agile, responsive, green) LARG framework through the development of a ML based decision -making system [25]. The hybridization of data science techniques and environmental analysis was proposed to improve LCA with several unsupervised ML methods for dimension reduction and clustering [36].

As can be seen in Table 4 several studies do not fall into neither of the categories that were evaluated here. Most of them used smart tools in combination with analytical methods for sample characterization. On a long term this of course is also aiming at gaining higher efficiencies increasing the sustainability of the approach. However, in these cases no report of higher yields etc. were mentioned.

The above analysis demonstrates that AI and ML methods contribute not only to the large-scale implementation of biorefinery concepts, but also to sustainability enhancement through efficiency gains, waste valorisation, multiproduct integration, and sustainability assessment frameworks. The integration of these diverse methods at various stages, from concept development to automation, is therefore essential for the success of the circular economy.

4. Critical discussion

The results of this study provide an overview of how AI, ML, and chemometric approaches are currently applied in smart biorefinery systems. However, beyond the descriptive categorisation of methods and applications, several critical aspects must be considered to properly assess the limitations, and practical relevance of these approaches. A first key observation is that the increasing adoption of AI/ML methods in biorefinery research does not necessarily translate into robust or transferable solutions. Many of the analysed studies rely on relatively small, case-specific datasets, often generated under controlled laboratory conditions. While these datasets are suitable for proof-of-concept modelling, they may not capture the variability and complexity of real-world biorefinery systems. As a result, model robustness remains a major limitation, particularly when models are applied beyond the specific conditions under which they were trained. Issues such as overfitting, lack of external validation, and insufficient reporting of uncertainty are frequently observed and limit the reliability of predictive outcomes.

Closely related to this is the challenge of model interpretability. Although advanced methods such as ANNs, DL, and ensemble models often show superior predictive performance, they are typically characterised by a “black-box” nature. This limits the ability to extract mechanistic understanding or chemical insight from the models, which is essential in complex systems such as biorefineries. In contrast, classical chemometric approaches, while potentially less powerful in capturing non-linear relationships, offer greater transparency and interpretability. This highlights a fundamental trade-off in current research between predictive performance and explainability, which is not yet systematically addressed in most studies. Another limitation concerns the scalability of AI/ML approaches. While many studies improved process optimisation or prediction accuracy at laboratory or pilot scale, there is limited evidence of successful implementation at industrial scale. Real-world biorefinery systems involve fluctuating feedstock compositions, operational variability, economic constraints, and regulatory requirements, all of which are rarely incorporated into current modelling approaches. This gap between laboratory-scale modelling and industrial application suggests that the current literature may overestimate the immediate applicability of smart methods in large-scale biorefinery systems.

Our analysis also reveals that different AI/ML approaches are often applied interchangeably to similar problems, with limited justification for method selection. While algorithms such as ANN, SVM, RF, and RSM are frequently used, comparative assessments between methods are scarce. Consequently, it remains unclear whether the choice of algorithm is driven by methodological suitability, data characteristics, or simply by availability and familiarity. More systematic benchmarking studies using shared datasets would be necessary to establish clearer guidelines for method selection and to improve reproducibility across studies.

The connection between smart biorefineries and sustainability, although frequently highlighted, also requires a more critical evaluation. Many studies assume that improved efficiency or optimisation leads to enhanced sustainability; however, this relationship is not always explicitly quantified. Only a limited number of studies incorporate comprehensive sustainability assessment tools such as LCA or TEA in combination with AI/ML methods. As a result, the environmental benefits of smart biorefineries remain largely inferred rather than demonstrated.

In particular, the potential contribution of AI-driven optimisation to climate action should be evaluated in terms of measurable indicators such as greenhouse gas emissions, energy consumption, and resource efficiency. While optimisation strategies may reduce energy demand or improve yields, the additional computational requirements associated with AI/ML models also entail energy consumption and associated carbon emissions. Therefore, the net environmental benefit of smart biorefinery approaches should be assessed through integrated frameworks that consider both process-level improvements and the environmental footprint of digital technologies. Also, in the context of the circular bioeconomy, AI and ML offer promising tools for improving resource allocation, waste valorisation, and system integration. However, the literature suggests that their application is still focused on isolated process improvements rather than on fully integrated circular systems. The transition towards truly circular and smart biorefineries will require the integration of data across different stages of the value chain, including feedstock supply, processing, product distribution, and end-of-life considerations. This level of system integration remains largely unexplored in current studies.

5. Challenges and perspectives of smart biorefineries

Currently, the availability and quality of data is often a major challenge for the next stage of smart biorefineries, especially for ML applications. However, the technological advances of sensors or sensitive analytical methods should increase the available data in the future. In addition, a homogeneous data set often leads to better prediction performance, therefore the implementing a data protocol for each biorefinery system is a crucial factor for robust data training and subsequent applications. Also, most ML methods are prone to overfitting or underfitting and susceptible to missing values. Consistent and well-trained model evaluation is essential to avoid misinterpretation. Arias et al. 2023 emphasises the need for a methodological guideline for AI tools including the assessment of their accuracy [9]. In addition, the combination of different algorithms in so-called “integrated models” has shown to outperform standalone models in terms of prediction performance, overfitting risk, and robustness, which could be an interesting direction for future research [19].

Algorithms with complex structures can achieve better performance with complex data sets, but they are also more difficult to interpret, which can be a further hindrance in the implementation of such an algorithm. Developing new strategies for ensemble learning models and more robust evaluation approaches will be relevant to meet the increasing needs of different scientific tasks [10]. Besides all the potentials of ML and AI tools to improve or even facilitate new more sustainable biorefineries, they also have an environmental impact that has to be considered. Training and operating AI and ML tools requires high computing power, and thus energy consumption leading to high carbon footprints. In particular, energy consumption for running and cooling server hardware in data centres is huge and the worldwide the energy consumption is predicted to continue to grow over the next decade [68]. Furthermore, resources used in manufacture all computational components, from chips to data centre buildings, must be considered. To illustrate the scale of the carbon footprint of AI tools, the estimated carbon footprint of training a single large language model (around 300, 000 kg CO_2 emission) was compared to that of 125 round-trip flights between New York and Beijing [69]. However, possible mitigation strategies exist, for instance: i) selecting efficient ML model architectures and advancing ML quality, such as sparse models versus dense models; ii) using processors optimised for ML training such as tensor processing unit (TPUs) or recent graphics processing units (GPUs); iii) computing in the cloud rather than on-premise can improve datacentre energy efficiency and iv) sustainable energy generation can be used to operate data centres and, by using cloud computing ML, practitioners can pick the location with the cleanest energy [70]. As a future direction, the combination of fast computation and reliable data sources may offer opportunities to create digital twins, and thus better process, tracking control, leading to more innovation systems towards circular bio-economy [10]. With increasing digitalisation of industry, for example, adopting IoT, cloud computing and automation, new vulnerabilities regarding cyber security are introduced. Therefore, there is a need for systems that promote the protection of IT infrastructure [9].

Moreover, the design and implementation of smart biorefineries can be affected by the fomentation of such approaches considering the geopolitical and resource conditions in the Globe. Fig. 6 illustrates the country collaboration world map for different data sets; (A) detaches the number of publications with the general biorefinery search term is reflected in many collaborations and an interwoven network worldwide. By refining the literature search and decreasing the number of publications, the total number of collaborations also decreases, but the relative contribution of different countries or regions of the world can be discussed. Data set B focusing on smart biorefineries, shows a high density in the Asian region, where a strong collaboration of China, India, Korea and Norway can be observed, and much fewer contributions from Europe, South America or Oceania are visible until now.

6. Conclusions

According to the most recent literature, it is possible to observe a growing research interest and emerging potential for establishing integrated and smart biorefineries as low-impact, potentially carbon-neutral or net-zero emission-oriented systems. The aim of this paper was to identify the literature trends and dynamics related to smart biorefinery approaches generating bio-based products, such as biofuels, energy and other chemicals and materials. In recent years, there has been a noticeable increase in studies focusing on smart tools, indicating a progressive shift towards data-driven approaches in biorefinery research. Various AI and ML algorithms have shown promising potential to support the establishment of smart biorefineries at larger scales, although this remains to be further validated. Such field of very dynamic development can seem to be a bit challenging as there are so many options, also because some algorithms can be applied for different purposes on one side, but different methods can be useful for the same purpose on the other side.

To provide a better overview and understanding of applications and to break down the diversity of possibilities, 8 different categories were identified: process modelling, evaluation and optimization, parameter selection, sample characterization, decision making, climate/spatial monitoring and automation, and several examples are given for each category. Further, the type of biorefinery in terms of biomass and resulting products were also considered. An overviewing table of all used smart methods in the analysed literature is provided with indication for what purpose the methods are applied to give the reader an overview of this diverse field and guidance in the search for a suitable method. In addition, it was analysed how smart tools may contribute to enhancing the sustainability of biorefinery concepts. The results of this work indicate that current applications are predominantly concentrated in process optimisation and modelling, while areas such as automation and system-level integration remain underdeveloped. Also, possible hindrances such as data quality, robustness and interpretability have been discussed here. It was observed that a lack of knowledge and /or experience of scientists with AI tools may represent an additional factor to be overcome in the future. Nevertheless, an emerging trend for smart biorefineries was observed, also suggesting the growing production of scientific knowledge. Interestingly, it was possible to note that studies focused on smart biorefineries are more frequent in the Asian region, especially in China, India, and Korea, which may be indicative of the great potential given to biomass valorisation in this region.

Biorefineries are generally associated with sustainability goals, as they transform non-edible renewable resources into high-value materials, but it is important to enable a transformation process that is as green and sustainable as possible. To develop and operate biorefinery processes efficiently and ensure resource-saving in more sustainable ways, the current trends and dynamics in the field suggest that smart methods could be integrated, so that digital manufacturing can have a greater effect on the effectiveness, economics, and quality of valuable chemical products and processes. By combining several smart-integrated greener and more sustainable processes, stronger synergies can be created, in which these processes not only complement but also optimize each other, enabling the initial non-edible biomass could be fully utilized. As observed, a myriad of different combinations of biomass, process conditions, and conversion possibilities are conceivable when achieving integrated smart biorefinery models, which cannot be verified through practical experimentation alone. For this reason, the consolidation of the term and development of adaptable and scalable smart biorefineries may represent a promising approach to reduce time, production costs, minimize environmental impacts such as climate change, promote more effective material and energy circularity, and potentially contribute to socio-sustainability globally. However, these contributions are still rarely supported by quantitative sustainability assessments, indicating that the transition from promising concepts to fully validated and scalable systems remains an open challenge. We should consider that while AI and ML are expanding in biorefinery research, their effective contribution to sustainable and industrially relevant systems will depend on the development of robust data frameworks, critical evaluation approaches, and integrated system-level applications.

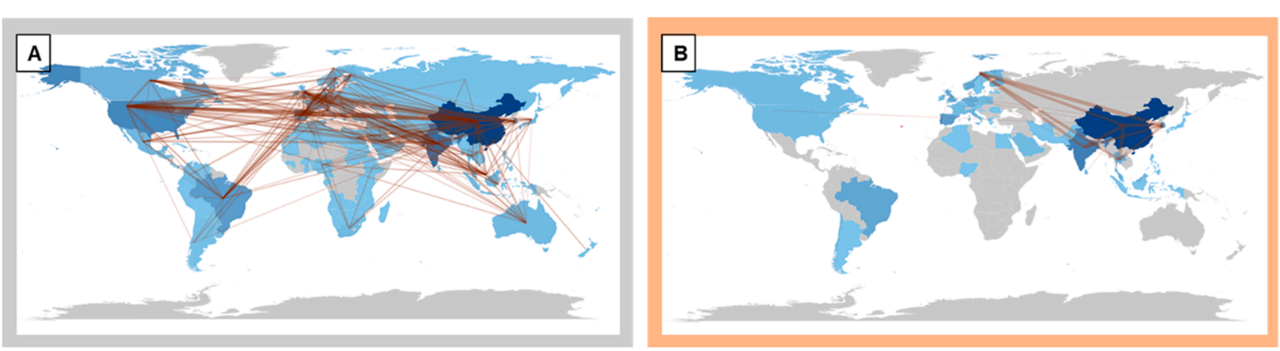

Fig. 6. Geographical maps showing country collaboration in the scientific community. 1: in grey data set A, 2: in orange data set B.

Notes

The authors declare no competing financial interest.

CRediT authorship contribution statement

Lotta Hohrenk-Danzouma: Writing – review & editing, Writing – original draft, Visualization, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Adrián Fuente-Ballesteros:** Writing – review & editing, Visualization. **Vânia G. Zuin Zeidler:** Writing – review & editing, Writing – original draft, Visualization, Supervision, Data curation, Conceptualization.

Declaration of competing interest

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

Acknowledgements

Authors thank the Sustainable Chemistry of Renewable Organic Resources Professorship at Leuphana University of Lüneburg (Germany) for its support.

References

[1] A. Muscat, E.M. de Olde, R. Ripoll-Bosch, H.H.E. van Zanten, T.A.P. Metze, C.J.A. M. Termeer, M.K. van Ittersum, I.J.M. de Boer, Principles, drivers and opportunities of a circular bioeconomy, Nat. food 2 (2021) 561–566.
[2] V.G. Zuin, Circularity in green chemical products, processes and services: innovative routes based on integrated eco-design and solution systems, Curr. Opin. Green Sustain. Chem. 2 (2016) 40–44.
[3] V.G. Zuin Zeidler, Defining sustainable chemistry-an opportune exercise? Science 382 (2023) eadk7430.
[4] H.Y. Leong, C.-K. Chang, K.S. Khoo, K.W. Chew, S.R. Chia, J.W. Lim, J.-S. Chang, P. L. Show, Waste biorefinery towards a sustainable circular bioeconomy: a solution to global issues, Biotechnol. Biofuels 14 (2021) 87.
[5] A.T. Ubando, C.B. Felix, W.-H. Chen, Biorefineries in circular bioeconomy: a comprehensive review, Bioresour. Technol. 299 (2020) 122585.
[6] L.R. Lynd, C.E. Wyman, T.U. Gerngross, Biocommodity Engineering, Biotechnol. Prog. 15 (1999) 777–793.
[7] M. Andersen, P. Kiel, Integrated utilisation of green biomass in the green biorefinery, Ind. Crops Prod. 11 (2000) 129–137.
[8] *Communication from the commision: Biomass action plan. {SEC(2005) 1573}*, 2005.
[9] A. Arias, G. Feijoo, M.T. Moreira, How could Artificial Intelligence be used to increase the potential of biorefineries in the near future? A review, Environ. Technol. Innov. 32 (2023) 103277.
[10] T.-H. Tsui, M.C.M. van Loosdrecht, Y. Dai, Y.W. Tong, Machine learning and circular bioeconomy: building new resource efficiency from diverse waste streams, Bioresour. Technol. 369 (2023) 128445.
[11] A.B. Culaba, A.P. Mayol, J.L.G. San Juan, C.L. Vinoya, R.S. Concepcion, A. A. Bandala, R.R.P. Vicerra, A.T. Ubando, W.-H. Chen, J.-S. Chang, Smart sustainable biorefineries for lignocellulosic biomass, Bioresour. Technol. 344 (2022) 126215.
[12] M. Aria, C. Cuccurullo, bibliometrix : an R-tool for comprehensive science mapping analysis, J. Inf. 11 (2017) 959–975.
[13] A. Soomro, S. Chen, Z. Sun, S. Ma, W. Xiang, Chemometric modelling on element compositions and product distributions of cellulose and lignin, Biomass Conv. Bioref 11 (2021) 2233–2246.
[14] K. Bashir, S. Jan, M. Habib, D.C. Saxena, A. Tarafdar, R. Sindhu, V. Kumar, K. Jan, Utilization of agro-industrial wastes (banana peel): development, characterization, and classification of biodegradable composite pots using PCA approach, Biomass Conv. Bioref (2024), https://doi.org/10.1007/s13399-024-06014-1.
[15] T.P. Houghton, D.M. Stevens, P.A. Pryfogle, C.T. Wright, C.W. Radtke, The effect of drying temperature on the composition of biomass, Appl. Biochem. Biotechnol. 153 (2009) 4–10.
[16] R. Tauler, Comprehensive Chemometrics. Chemical and Biochemical Data Analysis, Elsevier, Amsterdam, 2009.
[17] S. Rohilla, A. Jaiswal, B. Singh, C.L. Mahanta, Ultrasound-assisted extraction of phytochemicals from eryngium foetidum leaves using response surface methodology, Biomass Conv. Bioref (2023), https://doi.org/10.1007/s13399-023-04784-8.
[18] M.L. Segatto, K. Zanotti, V.G. Zuin, Microwave-assisted extraction and matrix solid-phase dispersion as green analytical chemistry sample preparation techniques for the valorisation of mango processing waste, Curr. Res. Chem. Biol. 1 (2021) 100007.
[19] R.K. Oruganti, A.P. Biji, T. Lanuyanger, P.L. Show, M. Sriariyanun, V.K. K. Upadhyayula, V. Gadhamshetty, D. Bhattacharyya, Artificial intelligence and machine learning tools for high-performance microalgal wastewater treatment and algal biorefinery: a critical review, Sci. Total Environ. 876 (2023) 162797.
[20] F. Maleki, K. Ovens, K. Najafian, B. Forghani, C. Reinhold, R. Forghani, Overview of machine learning part 1: fundamentals and classic approaches, Neuroimaging Clin. N. Am. 30 (2020) e17–e32.
[21] J. Huang, C.-D. Li, H. Zhao, M. Yu, A. Zhang, B. Fang, Artificial intelligence system for enhanced automated 1,3-propanediol green biosynthesis, Green Chem 25 (2023) 9175–9186.
[22] Z. Said, P. Sharma, Q. Thi Bich Nhuong, B.J. Bora, E. Lichtfouse, H.M. Khalid, R. Luque, X.P. Nguyen, A.T. Hoang, Intelligent approaches for sustainable management and valorisation of food waste, Bioresour. Technol. 377 (2023) 128952.
[23] J.Y. Lim, S.Y. Teng, B.S. How, K. Nam, S. Heo, V. Máša, P. Stehlík, C.K. Yoo, From microalgae to bioenergy: identifying optimally integrated biorefinery pathways and harvest scheduling under uncertainties in predicted climate, Renew. Sustain. Energy Rev. 168 (2022) 112865.
[24] A.A. Senocak, H. Guner Goren, Three-phase artificial intelligence-geographic information systems-based biomass network design approach: a case study in Denizli, Appl. Energy 343 (2023) 121214.
[25] M.S. Habib, S.-J. Hwang, Developing sustainable, resilient, and responsive biofuel production and distribution management system: a neutrosophic fuzzy optimization approach based on artificial intelligence and geographic information systems, Appl. Energy 372 (2024) 123683.
[26] D. Fozer, P. Nimmegeers, A.J. Toth, P.S. Varbanov, J.J. Klemeš, P. Mizsey, M. Z. Hauschild, M. Owsianiak, Hybrid prediction-driven high-throughput sustainability screening for advancing waste-to-dimethyl ether valorization, Environ. Sci. Technol. 57 (2023) 13449–13462.
[27] J. Wheeler, M.A. Páez, G. Guillén-Gosálbez, F.D. Mele, Combining multi-attribute decision-making methods with multi-objective optimization in the design of biomass supply chains, Comput. Chem. Eng. 113 (2018) 11–31.
[28] J.-P. Belaud, N. Prioux, C. Vialle, C. Sablayrolles, Big data for agri-food 4.0: application to sustainability management for by-products supply chain, Comput. Ind. 111 (2019) 41–50.
[29] M. Zhou, O.A. Fakayode, A.E. Ahmed Yagoub, Q. Ji, C. Zhou, Lignin fractionation from lignocellulosic biomass using deep eutectic solvents and its valorization, Renew. Sustain. Energy Rev. 156 (2022) 111986.
[30] S. Brahma, B. Nath, B. Basumatary, B. Das, P. Saikia, K. Patir, S. Basumatary, Biodiesel production from mixed oils: a sustainable approach towards industrial biofuel production, Chem. Eng. J. Adv. 10 (2022) 100284.
[31] M. Nayak, G. Dhanarajan, R. Dineshkumar, R. Sen, Artificial intelligence driven process optimization for cleaner production of biomass with co-valorization of wastewater and flue gas in an algal biorefinery, J. Clean Prod. 201 (2018) 1092–1100.
[32] J. Löfgren, D. Tarasov, T. Koitto, P. Rinke, M. Balakshin, M. Todorović, Machine learning optimization of lignin properties in green biorefineries, ACS Sustain. Chem. Eng. 10 (2022) 9469–9479.
[33] R. Jiang, Y. Linzon, E. Vitkin, Z. Yakhini, A. Chudnovsky, A. Golberg, Thermochemical hydrolysis of macroalgae ulva for biorefinery: Taguchi robust design method, Sci. Rep. 6 (2016) 27761.
[34] V.G. Zuin, M.L. Segatto, K. Zanotti, Towards a green and sustainable fruit waste valorisation model in Brazil: optimisation of homogenizer-assisted extraction of bioactive compounds from mango waste using a response surface methodology, Pure Appl. Chem. 92 (2020) 617–629.
[35] W. Wu, C.-M. Huang, Y.-H. Tsai, Design and validation of a microalgae biorefinery using machine learning-assisted modeling of hydrothermal liquefaction, Algal. Res. 74 (2023) 103230.
[36] N. Prioux, R. Ouaret, G. Hetreux, J.-P. Belaud, Environmental assessment coupled with machine learning for circular economy, Clean Techn. Environ. Policy (2022), https://doi.org/10.1007/s10098-022-02275-4.
[37] P. Karka, S. Papadokonstantakis, A. Kokossis, Digitizing sustainable process development: from ex-post to ex-ante LCA using machine-learning to evaluate bio-based process technologies ahead of detailed design, Chem. Eng. Sci. 250 (2022) 117339.
[38] B. Ahn, S. Jo, J. Na, J. Jay Liu, Y.-J. Kim, W. Won, Maximizing biomass utilization: an integrated strategy for coproducing multiple chemicals, J. Energy Chem. 100 (2025) 180–191.
[39] J. Li, Y. Wang, C. Xu, S. Liu, J. Dai, K. Lan, Bioplastic derived from corn stover: life cycle assessment and artificial intelligence-based analysis of uncertainty and variability, Sci. Total Environ. 946 (2024) 174349.
[40] R. Patel, T.S. Rajaraman, P.H. Rana, N.J. Ambegaonkar, S. Patel, A review on techno-economic analysis of lignocellulosic biorefinery producing biofuels and high-value products, Results Chem. 13 (2025) 102052.
[41] A. Asghar, C.-G. Liu, I. Ali, A.Z. Khan, H. Zhu, N. Wang, M. Nawaz, T.A. Tabish, M. A. Mehmood, R.T. Rasool, Bioenergy potential of saccharum bengalense through pyrolysis, reaction kinetics, TG-FTIR-GCMS analysis of pyrolysis products, and validation of the pyrolysis data through machine learning, Chem. Eng. J. 465 (2023) 142930.
[42] A. Shamjuddin, N.S. Ab Rasid, M.M. Michele Raissa, M.A. Abu Zarin, W.N.N. Wan Omar, A. Syahrom, M.A.F. Mohd, Szali Januddi, N.A. Saidina Amin, Kinetic and dynamic analysis of ozonolysis pre-treatment of empty fruit bunch in a well-mixed reactor for sugar production, Energy Convers. Manag. 244 (2021) 114526.
[43] Y. Huang, H. Wang, X. Zhang, Q. Zhang, C. Wang, L. Ma, Accurate prediction of chemical exergy of technical lignins for exergy-based assessment on sustainable utilization processes, Energy 243 (2022) 123041.

[44] L. Wang, G. Hou, Z. Tian, A. Baghban, Biomass chemical exergy estimation using novel intelligent tool: application to biorefinery sustainability, Energy Sources A: Recovery Util. Environ. Eff. 46 (2024) 52–66.

[45] D. Diment, J. Löfgren, M. Alopaeus, M. Stosiek, M. Cho, C. Xu, M. Hummel, D. Rigo, P. Rinke, M. Balakshin, Enhancing lignin-carbohydrate complexes production and properties with machine learning, ChemSusChem (2024) e202401711.

[46] U. Udaypal, R.K. Goswami, P. Verma, Strategies for improvement of bioactive compounds production using microalgal consortia: an emerging concept for current and future perspective, Algal. Res. 82 (2024) 103664.

[47] R.R. Kumar, D. Sarkar, R. Sen, Simultaneously maximizing microalgal biomass and lipid productivities by machine learning driven modeling, global sensitivity analysis and multi-objective optimization for sustainable biodiesel production, Appl. Energy 358 (2024) 122597.

[48] S.K. Singh, A. Chauhan, B. Sarkar, Resilience of sustainability for a smart production system to produce biodiesel from waste animal fat, J. Clean Prod. 452 (2024) 142047.

[49] W.R. Singh, H.N. Singh, Maximizing waste cooking oil biodiesel production employing novel Brotia costula derived catalyst through statistical and machine learning optimization techniques, Mater. Today Commun. 41 (2024) 110838.

[50] V. Sharma, M.-L. Tsai, C.-W. Chen, P.-P. Sun, P. Nargotra, C.-D. Dong, Advances in machine learning technology for sustainable biofuel production systems in lignocellulosic biorefineries, Sci. Total Environ. 886 (2023) 163972.

[51] W.M. Chan, D.V.-K. Le, Z. Chen, J. Tan, I.M.L. Chew, Resource allocation in multiple energy-integrated biorefinery using neuroevolution and mathematical optimization, Process Integr Optim. Sustain. 5 (2021) 383–416.

[52] Y. Yang, H. Shahbeik, A. Shafizadeh, N. Masoudnia, S. Rafiee, Y. Zhang, J. Pan, M. Tabatabaei, M. Aghbashlo, Biomass microwave pyrolysis characterization by machine learning for sustainable rural biorefineries, Renew. Energy 201 (2022) 70–86.

[53] M.R. Di Cicco, S. Altieri, A. Spagnuolo, C. Ciniglia, C. Germinario, S. Bove, A. Masiello, C. Vetromile, I. Galante, C. Lubritto, Sustainable cultivation of galdieria phlegrea in an IoT-integrated twin-layer photobioreactor: system design, growth dynamics, and isotopic perspective, Appl. Sci. 15 (2025) 5220.

[54] R. Yang, N. Zhao, J. Li, M. Liu, X. Li, Boosting corn stover bioconversion to fatty acids with Gaussian Process optimization, Ind. Crops Prod. 227 (2025) 120767.

[55] M. Morid, M.H. Khoshgoftar Manesh, An innovative optimal integrated solar-lignocellulosic biomass polygeneration system with biorefinery and solid oxide electrolyzer cell, Energy Convers. Manag. 327 (2025) 119557.

[56] M.A. Sperança, P.A.M. Nascimento, A.C. Olivieri, F.M.V. Pereira, XRF -based analytical methods for supporting sustainability in sugarcane biorefineries, Biofuels Bioprod. Bioref. 16 (2022) 758–765.

[57] Y. Peleg, S. Shefer, L. Anavy, A. Chudnovsky, A. Israel, A. Golberg, Z. Yakhini, Sparse NIR optimization method (SNIRO) to quantify analyte composition with visible (VIS)/near infrared (NIR) spectroscopy (350 nm-2500 nm), Anal. Chim. Acta 1051 (2019) 32–40.

[58] M. Sá, C.G. Bertinetto, N. Ferrer-Ledo, J.J. Jansen, R. Wijffels, J.G. Crespo, M. Barbosa, C.F. Galinha, Fluorescence spectroscopy and chemometrics for simultaneous monitoring of cell concentration, chlorophyll and fatty acids in Nannochloropsis oceanica, Sci. Rep. 10 (2020) 7688.

[59] M. Lucarini, A. Durazzo, J. Kiefer, A. Santini, G. Lombardi-Boccia, E.B. Souto, A. Romani, A. Lampe, S. Ferrari Nicoli, P. Gabrielli, N. Bevilacqua, M. Campo, M. Morassut, F. Cecchini, Grape seeds: chromatographic profile of fatty acids and phenolic compounds and qualitative analysis by FTIR-ATR spectroscopy, Foods (2019) 9, https://doi.org/10.3390/foods9010010.

[60] P. Chaiwan, P. Rachtanapun, Y. Phimolsiripol, W. Ruksiriwanich, C. Li, L. Luo, D. Shen, H.-H. Chung, D. George, T. Tangpao, S.R. Sommano, P. Sunanta, Physicochemical marker for determination of value-adding component in over-ripe Thai mango peels, Horticulturae 10 (2024) 1036.

[61] M. Lucarini, A. Durazzo, A. Romani, M. Campo, G. Lombardi-Boccia, F. Cecchini, Bio-based compounds from grape seeds: a biorefinery approach, Molecules 23 (2018), https://doi.org/10.3390/molecules23081888.

[62] A. Ganguly, R.C. Brown, M.M. Wright, Investigating the impacts of feedstock variability on a carbon-negative autothermal pyrolysis system using machine learning, Front. Clim. 4 (2022), https://doi.org/10.3389/fclim.2022.842650.

[63] T.C. Somasundaram, T.S. Mock, D.L. Callahan, D.S. Francis, Nutrient based classification of phyllospora comosa biomasses using machine learning algorithms: towards sustainable valorisation, Food Res. Int. 201 (2025) 115554.

[64] É.G.S. Souza, F.M.V. Pereira, Sustainable water management for steam generation in sugarcane biorefineries: applying PCA and MST clustering in sample analysis, J. Braz. Chem. Soc. (2025), https://doi.org/10.21577/0103-5053.20250006.

[65] J.C. Solarte-Toro, C.A. Cardona Alzate, Sustainability of biorefineries: challenges and perspectives, Energies 16 (2023) 3786.

[66] J. Clark, F. Deswarte, in: J. Clark, F. Deswarte (Eds.), Introduction to Chemicals from Biomass, Wiley, 2015, pp. 1–29.

[67] N. Marzban, M. Psarianos, C. Herrmann, L. Schulz-Nielsen, A. Olszewska-Widdrat, A. Arefi, R. Pecenka, P. Grundmann, O.K. Schlüter, T. Hoffmann, V.S. Rotter, Z. Nikoloski, B. Sturm, Smart integrated biorefineries in bioeconomy: a concept toward zero-waste, emission reduction, and self-sufficient energy production, Biofuel Res. J. 12 (2025) 2319–2349.

[68] A. Andrae, T. Edler, On global electricity usage of communication technology: trends to 2030, Challenges 6 (2015) 117–157.

[69] P. Dhar, The carbon impact of artificial intelligence, Nat. Mach. Intell. 2 (2020) 423–425.

[70] D. Patterson, J. Gonzalez, U. Hölzle, Q. Le, C. Liang, L.-M. Munguia, D. Rothchild, D. So, M. Texier and J. Dean, *The Carbon Footprint of Machine Learning Training Will Plateau, Then Shrink*, 2022.