

Review

AI-Driven Biohydrogen Production from Agri-Food Residues: Process Integration, Modeling Architectures, and Industrial Perspectives

Mahdi Rashvand ^{1,*} , Mohammad Tavakolian ² , Hessam Jahangiri ² , Attilio Matera ³, Marilisa Giavalisco ³ , Giuseppe Altieri ³ , Giovanni Carlo Di Renzo ³ , Teresa Zotta ³  and Francesco Genovese ^{3,*} 

- ¹ Centre for Business and Industry Transformation, Nottingham Business School, Nottingham Trent University, Nottingham NG1 4FQ, UK
- ² Advanced Food Innovation Centre, Sheffield Hallam University, Sheffield S1 1WB, UK; m.tavakolian@shu.ac.uk (M.T.); h.jahangiri@shu.ac.uk (H.J.)
- ³ DAFE, Department of Agricultural, Forestry, Food and Environmental Sciences, University of Basilicata, 85100 Potenza, Italy; attilio.matera@unibas.it (A.M.); marilisa.giavalisco@unibas.it (M.G.); giuseppe.altieri@unibas.it (G.A.); giovanni.direnzo@unibas.it (G.C.D.R.); teresa.zotta@unibas.it (T.Z.)
- * Correspondence: mahdi.rashvand@ntu.ac.uk (M.R.); francesco.genovese@unibas.it (F.G.)

Abstract

Agri-food residues offer renewable feedstocks for hydrogen production, but compositional variability, conversion constraints, and inconsistent data reporting hinder reliable modelling and scale-up. This review examines artificial intelligence (AI) across dark fermentation, photo-fermentation, microbial electrolysis, gasification, and pyrolysis–reforming. Its distinctive contribution is a residue-centred comparison linking process constraints to AI tasks, validation requirements, and industrial decisions, rather than considering prediction accuracy alone. Supervised, unsupervised, optimisation-assisted, and hybrid approaches are assessed using feedstock relevance, data provenance, physical consistency, and transferability. Selected studies provide quantitative evidence of improved prediction, process optimisation, and computational efficiency; these outcomes are distinguished from experimentally verified hydrogen gains. Despite promising results, laboratory validation and simulation performance do not establish cross-feedstock robustness or industrial readiness. Mechanistic integration can help represent acid accumulation, nitrogen-dependent inhibition, hydrogen partial pressure, and thermochemical energy constraints, provided parameters and uncertainty are validated. Future progress requires harmonised datasets, independent validation, and integration with techno-economic and life-cycle assessment. Digital twins remain a development opportunity whose operational benefits require prospective demonstration.

Keywords: agri-food residues; artificial intelligence; biohydrogen; dark fermentation; gasification; hybrid modelling



Academic Editor: Giovanni Esposito

Received: 16 September 2026

Revised: 12 October 2026

Accepted: 14 October 2026

Published: 18 October 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

The agri-food industry is among the major producers of biomass residues globally, including huge amounts of crop residues, food waste, agro-industrial by-products, animal manure and high-strength wastewaters. These waste streams present growing environmental challenges to human health and resource containment, yet at the same time they constitute a substantial but underdeveloped renewable source of clean energy. Reviews indicate that agri-food residues have a good biochemical nature and are suitable for hydrogen-based bioconversion, mainly with the carbohydrates high content biodegradability [1,2]. A

systems perspective suggests that their use should be optimized through advanced process integration and system-level collaboration to address variability and supply chain issues.

Agri-food waste biohydrogen conversion can be performed through various biochemical routes as well as thermochemical methods. Process reviews describe how dark fermentation is still the most developed biochemical pathway for agri-food substrates, despite being extremely susceptible to pH, substrate complex composition, and product inhibition. Residue-specific studies also illustrate that food waste and agricultural residues display significant compositional diversity, directly affecting microbial metabolism and hydrogen production. Overall, these investigations showed that agri-food biohydrogen systems are complex, multiscale, and highly nonlinear.

Despite the crucial role that these conventional model-based predictions, including stoichiometric, kinetic, thermodynamic, and empirical regression models, have played towards undertaking direction in biohydrogen studies to date, it has been found that they also have inherent limitations when used at the agri-food systems level. Hosseinzadeh et al. revealed that empirical and linear models cannot describe the non-linear interactions between volatile fatty acids, trace elements, and operational parameters in dark fermentations [3]. These limiting factors are especially prominent when extrapolating from well-defined laboratory experiment conditions to heterogeneous, real-world Agri-food feedstocks subjected to significant seasonal variation and operational uncertainty.

In this regard, the use of artificial intelligence (AI) and machine learning (ML) has emerged as a promising solution for modeling, prediction, and optimization in agri-food bio-hydrogen systems. Hosseinzadeh et al. reported that ensemble ML models predicted hydrogen production from wastewater fermentation within the evaluated dataset [3]. Similarly, Zhao et al. noted that interpretable ML models were evaluated for SCWG prediction [4]. Beyond prediction, food-waste research [5] shows how hybrid AI-optimization frameworks can be used for the discovery of the optimal operating windows while considering H₂ yield, stability, and process efficiency. In addition, Mougari et al. reported a hybrid BO-ANN approach, based on the fact that integrating kinetic descriptors notably improves model performance in agri-food substrate comparisons [6].

The literature increasingly addresses the increasing use of machine learning in bio-hydrogen and biofuel production, emphasizing algorithmic improvements, predictive capacity, and optimization abilities [3,4,6]. However, in most of the previously published reviews on process synthesis for biorefineries, agri-food residues as such are treated as general biomass feedstocks; discussion on how to integrate processes with each other, hybrid AI-process modeling architectures, and scale-up issues is very limited. Besides, a vast majority of AI-based research is conducted on secondary datasets, and their validation procedures are not standardized and are restricted to laboratory-scale systems, which causes fears concerning robustness, transferability, and industrial applicability.

The distinctive contribution of this review is to connect residue-dependent conversion constraints, AI functions, and evidence needed for industrial decisions across biological and thermochemical pathways. The comparison asks which process variables AI should represent, what quantitative benefit is demonstrated, and what validation supports transfer to a new feedstock or scale.

Process studies were retained when they identified a residue- or waste-derived substrate and reported conditions or hydrogen outcomes relevant to the review. AI studies were considered when they linked process variables to hydrogen-related prediction, diagnosis, optimisation, or system design. Evidence from broader wastewater or co-gasification systems was retained only with its domain limitation stated. General AI applications unrelated to hydrogen conversion were not used to demonstrate biohydrogen improvements.

This distinction avoids conflating defined-substrate experiments, operational residue conversion, and methodological analogues.

The comparison considers feedstock provenance, experimental versus simulated data, model task, validation design, quantitative outcome, and limits to transferability. Supervised learning is assessed as labelled-input prediction; unsupervised learning as exploration or diagnosis; and hybrid modelling as integration with kinetics, thermodynamics, experiments, or mathematical programming. Numbers are reproduced only when their meaning can be identified in the accessible primary publication record; unspecified sample counts, split details, units, or baselines are left unassigned rather than inferred. Abstract-level numerical information is explicitly identified in Table 1. No pooled ranking is attempted across different target definitions, scales, or validation schemes. The resulting structure connects hydrogen-production constraints in Section 2 to AI functions in Section 3, hybrid architectures in Section 4, and operational evidence requirements in Sections 5 and 6.

2. Biohydrogen Production Systems in the Agri-Food Sector

2.1. Biological Pathways for Biohydrogen Production

Biological hydrogen production offers promising opportunities for converting agri-food residues into renewable energy through microbial and light-driven processes. The principal pathways include dark fermentation, photo-fermentation, bio-photolysis, and microbial electrolysis cells, each involving distinct biochemical mechanisms, operating conditions, and microbial communities. Their performance is strongly influenced by feedstock composition, substrate availability, microbial activity, and environmental parameters such as pH, temperature, and nutrient concentration. Although these pathways provide opportunities for sustainable waste valorisation, limitations associated with conversion efficiency, process stability, and scalability remain significant challenges. Understanding their underlying mechanisms and operational constraints is therefore essential for identifying suitable process configurations and developing AI-assisted prediction, monitoring, and optimisation strategies.

2.1.1. Dark Fermentation

Dark fermentation (DF) is the most widely studied and exploited pathway for bioH₂ production. DF is an acidogenic process that converts carbohydrate-rich organic matter into H₂ through the metabolism of obligate and/or facultative anaerobic microorganisms. In both microbial groups, hexoses are converted into pyruvate mainly via the glycolytic pathway, whereas the pentose phosphate or Entner-Doudoroff routes are used to a lesser extent, depending on the microbial species and type of substrates. In obligate anaerobes (e.g., clostridial species), pyruvate is then oxidized to acetyl-CoA and CO₂ by the pyruvate:ferredoxin oxidoreductase (PFOR) enzyme, with production of reduced ferredoxin (Fdrd); the latter can transfer electrons to protons to generate H₂ via a Fd-dependent hydrogenase. In Clostridia and thermophilic anaerobes, additional H₂ may be driven by coupled activity of NADH:ferredoxin oxidoreductase (NFOR) and Fd-hydrogenase.

In facultative anaerobes (e.g., *Escherichia coli*, *Enterobacteria*), instead, pyruvate can be converted into acetyl-CoA and formate by pyruvate-formate lyase (PFL); formate is then cleaved into H₂ and CO₂ by the formate-hydrogen lyase complex (FHL).

In clostridial-type DF, the two main routes for H₂ production are the acetate and butyrate pathways. When glucose is primarily converted into acetate, the maximum theoretical yield is 4 mol H₂/mol glucose; while when butyrate is the main product, the maximum yield is 2 mol H₂/mol glucose because part of the reducing equivalents is retained in the butyrate carbon skeleton, rather than being available for H₂ production.

Strict anaerobes belonging to the species *Clostridium butyricum*, *C. tyrobutyricum* (acidogenic metabolism with acetate and butyrate as main products), *C. acetobutylicum*, *C. beijerinckii*, *Cl. pasteurianus* (able to shift toward solventogenic pathway with production of acetone, ethanol and butanol) commonly drive high-yield DF [7]. Clostridia are spore-forming bacteria able to use a broad range of substrates, including several agri-food wastes (e.g., starch and lignocellulosic hydrolysates, dairy effluents), making them highly versatile for the implementation of industrial-scale processes.

Several thermophilic and hyperthermophilic bacteria, such as *Thermotoga maritima*, *Thermotoga neapolitana*, and *Caldicellulosiruptor saccharolyticus*, can metabolize complex carbohydrates under high temperatures, making them promising candidates for the conversion of agricultural residues and lignocellulosic hydrolysates. However, thermophilic operation entails higher energy requirements and tighter process control [8].

Facultative anaerobic bacteria (e.g., *Citrobacter freundii*, *Enterobacter cloacae*, *Escherichia coli*, *Klebsiella pneumoniae*), although having a lower theoretical H_2 yield than clostridia, are operationally attractive because of their robustness, growth rate, and tolerance to variable environmental conditions.

Although DF is currently the main fermentative pathway for H_2 production, the process still faces several constraints that must be overcome to achieve high-yield operations and stable industrial-scale implementation. The main limitation is the incomplete conversion of substrate into H_2 , with production of undesirable metabolites. The synthesis of reduced end-products such as lactate, ethanol, butanol, or propionate may impair H_2 yield by redirecting metabolic fluxes and reducing equivalents (e.g., NADH consumption for NAD^+ regeneration) away from hydrogenase-mediated pathways. Moreover, volatile fatty acids (VFAs) may lead to acidification and the inhibition of H_2 -producing bacteria when pH is not controlled. Clostridial metabolism may be negatively affected by low pH, leading some strains to shift to solventogenesis with reduced H_2 productivity. Moreover, in processes involving microbial consortia, some microorganisms may inhibit or compete with H_2 production, as mixed cultures may include H_2 -consuming members, such as methanogens, homoacetogens, sulfate reducers, propionic acid producers, and lactic acid bacteria [9]. The latter group, however, has been recently revalued as beneficial, since the lactate produced in early fermentation stages can be reassimilated by H_2 -producing bacteria and converted into pyruvate, acetyl-CoA, acetate, butyrate, and H_2 (lactate-driven DF) [10].

Besides microbial dynamics, however, other operational parameters (such as organic loading, substrate concentration, hydraulic retention time, hydrogen partial pressure, temperature, bioreactor design) are critical and require careful control to achieve a high-yield process [1,11].

The above constraints are particularly relevant for H_2 production involving agri-food wastes as carbon-rich substrates (which often require complex microbial consortia); Therefore, the use of DF as the initial stage of a cascading biorefinery (in combination with other fermentation processes), rather than as a standalone technology, is of increasing interest.

Agricultural-residue studies illustrate why substrate identity must remain explicit. *Thermotoga neapolitana* converted Korean rice straw, with reported yield increasing from 2.27 to 2.70 mmol H_2 /g straw after ammonia–acid pretreatment [12]. *Thermoanaerobacterium thermosaccharolyticum* M18 was investigated for direct conversion of wheat straw, reporting 3.53 mmol H_2 /g substrate [13]. Jackfruit peel and non-edible fractions required heating, size reduction, and acid hydrolysis before fermentation with *Enterobacter aerogenes* [14]. These studies concern residue-derived substrates, but their different organisms, pretreatments, and denominators prevent a simple yield ranking.

Agro-industrial liquids provide more accessible carbon while introducing their own chemical variability. Rice-mill wastewater has been fermented with selected *Bacillus*

strains [15]; palm-oil-mill effluent (POME) has been studied with immobilised *Clostridium butyricum* and mixed-culture continuous reactors [16–18]. Cassava wastewater studies demonstrate the role of hydraulic retention time and loading in continuous hydrogen production [19], while dairy manure required acidification pretreatment [20]. Such results identify candidate AI predictors: substrate composition and pretreatment, microbial inoculum, temperature, pH, organic loading rate (OLR), hydraulic retention time (HRT), and gas-removal conditions. Feedstock labels alone cannot capture differences between raw effluent, hydrolysate, and pre-fermented material.

Recent food-waste hydrolysate research combines process experimentation with ANN-genetic algorithm (GA) optimisation [5], providing a direct link between residue processing and AI-assisted hydrogen production. Genetic engineering offers another route to changing the feasible process envelope: hydrogenase overexpression in *Clostridium pasteurianum*, *fhlA* overexpression in *Enterobacter cloacae* using cotton-stalk hydrolysate, and modifications to *Escherichia coli* carbon metabolism have been examined [21–23]. These interventions should be treated as distinct biological systems in model development. A predictor trained on one strain or consortium cannot be assumed to represent engineered strains, new pretreatments, or a shift between lactate-driven and conventional acidogenic operation.

2.1.2. Photo-Fermentation and Bio-Photolysis

Photo-fermentation (PF) is a light-driven process in which photosynthetic bacteria can convert soluble organic compounds into H_2 and CO_2 under anaerobic conditions.

Purple non-sulfur bacteria (PNSB), such as *Cereibacter sphaeroides* (formerly *Rhodobacter sphaeroides*), *Rhodobacter capsulatus*, *Rhodopseudomonas palustris*, *Rhodospirillum rubrum*, and the marine species *Rhodovulum sulfidophilum*, are principally associated with PF-driven H_2 production [24].

PF may be implemented either as a single-stage process or as part of an integrated two-stage dark–photo fermentation system. In a single-stage approach, PNSB directly convert soluble substrates (simple sugars, low-molecular-weight organic acids) into H_2 ; however, this process is not suitable for agri-food wastes (especially lignocellulosic residues) because of limited efficiency of PNSB in hydrolyzing complex biomass. In a two-stage configuration, PF is employed as a downstream process to improve total H_2 recovery and reduce the burden of acidogenic effluents resulting from the DF phase. This approach is particularly useful for waste degradation, where undesirable metabolites remain largely unconverted after the first fermentative step [25]. The sequential DF-PF process, then, may be exploited as a strategy for complete waste reduction and valorization. However, in an integrated system, buffering, pH adjustment, and effluent dilution should be considered to avoid acid inhibition and to maintain the activity of photosynthetic bacteria.

Nitrogenase is the key enzyme in the PF process; its activity is strongly affected by ammonium concentration, light intensity, O_2 exposure, and organic acids content. Nitrogen availability is a critical parameter when agro-food residues are used, as many substrates contain proteins, amino acids, or ammonium released during hydrolysis and fermentation. For this reason, the PF step often requires careful control of C/N ratio, nitrogen source, ammonium concentration, and waste pretreatment strategies.

Other PF constraints include high light requirements, the need for large surface areas, low volumetric productivity, high photobioreactor costs, and relatively complex scalability. Light intensity, wavelength, photoperiod, penetration depth, reactor geometry, and cell density may affect the amount of light available to the photosynthetic bacteria and, therefore, the scaling-up of the PF process requires an accurate photobioreactor design.

Other microorganisms, such as microalgae and cyanobacteria, may produce H_2 through photobiological pathways rather than the conventional PF process. In green

microalgae (e.g., *Chlamydomonas reinhardtii*, *Chlorella variabilis*), electrons derived from water oxidation or carbohydrate catabolism are transferred to [FeFe]-hydrogenases to generate H₂; in non-diazotrophic cyanobacteria (e.g., *Synechocystis* sp.), H₂ evolution is mediated by a bidirectional [NiFe]-hydrogenase, whereas nitrogenase is the main H₂-producing enzyme in diazotrophic species such as *Anabaena variabilis* and *Nostoc punctiforme* [26]. O₂-sensitivity of H₂-producing enzymes is the main constraint for both microalgae and cyanobacteria; however, diazotrophic strains can partially mitigate this limitation through spatial or temporal separation of O₂ evolution and H₂ production.

The use of agro-industrial waste for H₂ production via microalgal and clostridial metabolism also requires integration with DF, due to the constraints already reported for the PF system. This approach, however, is still poorly investigated [27].

Residue-specific PF studies include sugar-beet molasses and winery wastewater [28,29]. Sequential DF–PF has been examined using cheese whey [30] and corn-straw hydrolysate [31]. In a recent second-cheese-whey study, lactic fermentation preceded PF with *Rhodospseudomonas palustris* under selected-wavelength illumination [32]. Together, these examples show that upstream treatment, nitrogen availability, and illumination must be described alongside hydrogen yield. The corn-straw work provides pilot-scale fermentation evidence; it is not evidence that an AI controller or digital twin caused the reported performance.

PF imposes constraints on illuminated surface area, light penetration, photoperiod, organic-acid concentration, oxygen exposure, and nitrogen availability. Ammonium may suppress nitrogenase-mediated hydrogen production even when biomass growth remains favourable [24]. Candidate AI features therefore include dissolved nitrogen species, optical properties, light intensity at the culture rather than lamp rating alone, reactor geometry, and cell density. Dynamic models should distinguish microbial growth from hydrogen evolution. A recommended light regime must also be assessed against electricity demand and the cost of maintaining a photobioreactor.

2.1.3. Microbial Electrolysis Cells

Microbial electrolysis cells (MECs) are an emerging technology in which electroactive bacteria oxidize organic matter, releasing CO₂, electrons, and protons. Electrons are transferred to the anode and subsequently flow to the cathode through an external circuit; protons migrate through the electrolyte or an ion-exchange separator and combine with the electrons to generate H₂ under a low applied voltage [33]. Members belonging to the *Geobacter* species are the most studied electroactive bacteria, with *G. sulfurreducens* frequently used as a model organism, because of its ability to form a conductive biofilm, oxidize acetate, and transfer electrons through conductive pili and outer-membrane cytochromes. *Shewanella oneidensis* is also studied as an electroactive microorganism, due to its genetic versatility and flavin-mediated electron transfer mechanisms, although H₂ yield is lower than the *Geobacter*-driven process.

The ability to combine organic matter removal with energy recovery makes MECs particularly attractive for treating agri-food wastes, wastewater, and fermentation effluents. Integration of DF–MEC systems may overcome the limits of the DF stand-alone process by using VFAs and organic acids as electron donors at the anode, thereby reducing residual COD and recovering more energy from the same substrate.

MEC performance, however, mainly depends on applied voltage, electrode material, reactor configuration, membrane fouling, substrate composition, and electroactive microbial consortia [34]. Efficient DF–MEC integration, therefore, requires coordinated control of DF-derived VFAs, organic loading, pH, hydraulic retention time, and microbial community

structure to maintain stable anodic activity, enhance residual COD removal, and maximize cumulative H_2 recovery from heterogeneous agri-food waste streams.

Agri-food applications include two-stage DF–MEC treatment of cassava-starch processing wastewater and POME [35,36], as well as fruit-juice wastewater [37]. These configurations use residual soluble organics from fermentation to recover additional hydrogen. A food-waste AD–MEC system also demonstrates the relevance of pressure and reactor configuration [38]. The recent review of DF–bioelectrochemical coupling distinguishes sequential and integrated configurations [33]. These are different modelling tasks: an upstream predictor must estimate effluent composition, while a downstream model must relate that composition and electrochemical conditions to hydrogen recovery.

AI can support prediction of hydrogen recovery, current, effluent quality, or fouling, but optimisation must include electrical energy consumption and electrode replacement. Maximising gas production alone may select a high-voltage condition with poor net energy performance. Model inputs should include measured VFAs, conductivity, pH, loading, applied voltage or electrode potential, and indicators of anode activity. Transfer across electrode materials or separator types requires new validation; a correlation learned for one MEC configuration is not an intrinsic property of the feedstock.

2.2. Biohydrogen in Sustainable Agri-Food Systems and Waste Valorisation

Agri-food wastes are increasingly investigated as low-cost renewable substrates for the biological production of H_2 , as their use combines waste stabilization with energy recovery and supports circular bioeconomy strategies. However, conversion efficiency depends not only on the microbial process employed but also on the type of waste and its physicochemical properties. The choice of feedstock, therefore, is a critical step, as it directly affects H_2 yield, functionalities of microbial consortia, process stability, pretreatment intensity, energy balance, and overall economic feasibility.

One of the main categories used for H_2 generation consists of lignocellulosic agricultural materials (e.g., cereal crop residues, sugarcane, cotton stalks, cassava residues), which are non-edible, widely available, and do not compete directly with food production.

Agricultural wastes are mainly constituted by cellulose, hemicellulose, and lignin, although their concentration and ratio may differ depending on the wastes. Cellulose and hemicellulose may be hydrolyzed into fermentable sugars (e.g., glucose, xylose, galactose, arabinose), whereas lignin, having a greater resistance to microbial degradation, requires physical, chemical, biological, or combined pre-treatments before use as a substrate for H_2 production [2]. Severe pretreatments, however, may improve sugar release but can also generate inhibitory compounds, increase chemical consumption, and reduce the yield and energy balance of the overall process.

Food waste is more biodegradable because it contains readily fermentable carbohydrates, proteins, lipids, and organic acids. Its high moisture content makes it particularly suitable for anaerobic processes, including DF. Carbohydrate-rich residues are generally more favorable for H_2 production than fat-rich substrates because carbohydrates are rapidly hydrolyzed and directed toward pyruvate, acetyl-CoA, reduced ferredoxin, NADH, and hydrogenase-mediated H_2 generation. Lactate-driven DF is relevant for dairy residues, as well as fruit and vegetable wastes, where lactic acid bacteria can dominate early fermentation stages.

Other agro-industrial effluents (e.g., wastewater from distilleries and breweries, starch processing plants, sugar refineries, olive oil mills, dairies, and food processing facilities) are particularly suitable for DF and MECs, as H_2 production can be combined with wastewater treatment and the reduction of organic load.

Animal manures may be potentially useful for H₂ production, as they contain organic matter, nutrients, and indigenous microbial consortia; however, their high nitrogen content may inhibit the key enzymes involved in H₂ generation.

The main biological processes, as well as the microbial strains or consortia used for H₂ production from agri-food waste, are listed in Table 1.

Table 1. Residue-based biological hydrogen production: selected conditions, reported yield, and production rate.

Pathway/Inoculum	Residue or Waste-Derived Substrate	Selected Conditions	Reported H ₂ Yield or Cumulative Output	Reported Production Rate	Reference
DF; <i>Bacillus</i> strains	Rice-mill wastewater	Batch; 37 °C; pH 5.5–6.5; strain-specific wastewater fraction	0.87–1.63 mol/mol glucose	8.51–9.28 mL L ^{−1} h ^{−1}	[15]
DF; <i>T. neapolitana</i>	Korean rice straw; raw vs. ammonia–acid pretreated	Batch; 75 °C; pH 7.5	2.27 vs. 2.70 mmol/g straw	31.77 vs. 112.38 mL L ^{−1} d ^{−1}	[12]
DF; <i>T. thermosaccharolyticum</i> M18	Wheat straw	Batch; 60 °C; pH 7.0	3.53 mmol/g substrate	0.11 mmol L ^{−1} h ^{−1}	[13]
DF; <i>E. aerogenes</i>	Jackfruit peel/non-edible fractions; acid hydrolysate	Batch; 35 °C; pH 6.5	5.02 mol/kg COD; cumulative 3.51 L/L	NR	[14]
DF; immobilised <i>C. butyricum</i>	Raw POME	Batch; 37 °C; pH 5.5	5.35 L/L POME	481 mL L ^{−1} POME h ^{−1}	[16]
	Raw POME	UASFF; 43 °C; HRT 4 h; pH 5.0	1021.74 mL/g COD	5.29 L/d (total output)	[17]
	Nitric-acid-pretreated POME	UASB; 55 °C; HRT 6 h; pH 5.5	2.45 mol/mol total carbohydrate	11.75 L L ^{−1} POME d ^{−1}	[18]
	Food-industry residues/yeast/manure mixture	CSTR; 55 °C; HRT 2 d; pH 7.0	16.5 mL/g VS	NR	[39]
	Cassava-processing wastewater	AFBR; 28 °C; HRT 2 h; pH 5.58	1.91 mol/mol glucose	1.92 L L ^{−1} h ^{−1}	[19]
	Acid-pretreated dairy manure	Stirred reactor; 36 °C; pH 5.0	31.5 mL/g TVS	NR	[20]
	Cheese-whey wastewater	Batch; mesophilic vs. thermophilic	9.2 vs. 8.1 mmol/g COD	5.1 vs. 1.1 mL g ^{−1} VSS h ^{−1}	[40]
	Sugarcane vinasse	Batch; 55 °C; pH 5.0–5.5	95.82 ± 0.45 NmL/g VS	22.26 NmL g ^{−1} vs. d ^{−1} (fitted maximum)	[41]
Lactate-driven DF; specialised consortium	Cheese whey/fruit and vegetable waste	Batch; 35 °C; pH 5.5	1.75 mol/mol hexose	NR	[42]
Lactate-driven DF; native community	Cafeteria food waste	Lactate production followed by batch H ₂ stage; 37 °C	596.3 mL/L	NR	[43]
Lactate-driven DF; specialised consortium	Raw vs. pre-fermented acid cheese whey	CSTR; 35 °C; pH 5.8; HRT 24 h	76.76 vs. 54.6 mL/g COD	8.05 vs. 7.8 L L ^{−1} d ^{−1} (reported maxima/plateau)	[44]

Table 1. Cont.

Pathway/Inoculum	Residue or Waste-Derived Substrate	Selected Conditions	Reported H ₂ Yield or Cumulative Output	Reported Production Rate	Reference
Lactate-driven DF; enriched consortium	Tequila vinasse	Short-HRT operation; 35 °C; pH 5.8	82 NmL/g VS	6.79 L L ⁻¹ d ⁻¹	[45]
PF; <i>R. palustris</i>	Sugar-beet molasses	30 °C; pH 7.5; illumination 114 W/m ²	19.0 mol/mol sucrose	0.55 mmol L ⁻¹ h ⁻¹	[28]
PF; mixed culture	Winery wastewater	Batch; 25 °C; pH 7.0; LED illumination	468 mL/L	NR	[29]
DF–PF; mixed cultures	Corn-straw enzymatic hydrolysate	Pilot continuous baffled reactors	Stage-specific yields not combined here	DF 15.04; PF 8.26 m ³ m ⁻³ d ⁻¹	[31]
DF–PF; <i>E. aerogenes</i> / <i>R. sphaeroides</i>	Raw cheese whey	Sequential batch; 30 °C	Stage outputs: 3270 and 3800 mL	NR	[30]
DF–MEC; thermophilic cultures	Cassava-starch wastewater	DF HRT 24 h; MEC HRT 48 h; 55 °C; 0.6 V	Reported combined recovery 465 mL/g COD	DF 2.163; MEC 0.223 L L ⁻¹ d ⁻¹	[35]
	POME	DF HRT 2 d; MEC HRT 6 d; 55 °C; 0.7 V	Reported combined recovery 236 mL/g COD	DF 2.43; MEC 5.38 L L ⁻¹ d ⁻¹	[36]
DF–MEC; mixed culture	Fruit-juice wastewater	Sequential batch; 37 °C; controlled anode potential	Reported combined recovery 1608.6 mL/g COD	NR	[37]
	Simulated food-waste mixture	30 °C; reduced pressure; controlled anode potential	511.02 mL/g VS	3.48 L L ⁻¹ d ⁻¹	[38]
LF–PF; lactic bacteria/ <i>R. palustris</i>	Second cheese whey	Sequential; selected-wavelength LED illumination	PF 0.46 L/L substrate	1.42 mL L ⁻¹ h ⁻¹	[32]

NR: not retained in this condensed synthesis; it does not imply that the original study failed to report the variable. Values preserve source-specific denominators and gas-volume conventions. *N* denotes a source-reported normalised gas volume; reference temperature and pressure must be checked before conversion. Ranges combine conditions within one study, not a cross-study meta-analysis. Stage-specific volumetric rates are not added because reactor-volume denominators may differ. Combined recoveries are identified as source-reported values. COD supplied versus consumed, working volume, pretreatment inputs, and uncertainty must be reconciled before any dataset is pooled.

Biological H₂ production from agri-food residues relies primarily on the use of DF as a single-stage process, although several studies employ two-stage configurations (e.g., DF coupled with PF or MECs). Downstream PF or MEC stages can further convert residual metabolites and increase overall H₂ recovery, but they also require additional reactors, process control, and energy input.

Readily biodegradable liquid by-products, such as cheese whey, wastewater, and food-processing effluents, are the most commonly investigated feedstocks. Solid lignocellulosic residues, particularly rice, wheat, and corn straw, are less frequently studied and generally require pretreatment to improve carbohydrate accessibility. Feedstock composition and seasonal variability remain major constraints to process scale-up because they affect biodegradability, nutrient balance, inhibitor concentrations, and microbial metabolism.

Overall, Table 1 shows that process performance depends on interactions of several factors (e.g., substrate, microorganisms, fermentation process, operating conditions) rather

than on any single input considered independently. AI-driven modeling should therefore adopt a multilevel framework capable of capturing the nonlinear and cross-level interactions among the different process inputs.

Table 1, therefore, provides an overview of variables to be considered when developing predictive AI-driven models for optimizing H₂ production processes. Specifically, production pathway, process type, single strains or microbial consortium, feedstock, pretreatment, reactor configuration, operative parameters (e.g., temperature, pH, hydraulic retention time, organic loading rate, agitation, gas-flushing conditions, applied voltage, illumination) may be considered as candidate predictors, while data on H₂ production (e.g., H₂ yield, production rate, substrate conversion) may represent potential target variables.

At the same time, the heterogeneity reported in Table 1 highlights the need for standardized data collection and AI architectures capable of accounting for the combined effect of raw materials, microorganisms, and operating conditions. This approach is crucial for improving the robustness and transferability of models, as well as their applicability to the scale-up processes.

Table 1 is a compact process-evidence map, not a newly assembled training dataset. It summarises the residue-based studies discussed above and separates hydrogen yield from production rate. Its purpose is to identify the variables and reporting differences an AI dataset would need to preserve. The entries were not pooled or used to train a model in this review. Details such as agitation and flushing protocols remain available in the cited studies; the table retains the conditions most relevant to comparability and pathway selection.

Three commonalities emerge. First, pretreatment or upstream fermentation changes the actual substrate delivered to the hydrogen-producing system; raw material and hydrolysate must therefore be separate features. Second, the biological pathway and reactor configuration determine whether sugars, lactate, or residual VFAs are the relevant electron donors. Third, reported performance includes unlike quantities: cumulative output, substrate-normalised yield, volumetric productivity, and specific rates. A high value in one column does not establish superior energy efficiency or performance in another.

Table 1 includes the use of natural strains and/or microbial consortia for the production of H₂ from agri-food waste.

However, genetic and metabolic engineering strategies have also been implemented to enhance fermentative H₂ production by increasing hydrogenase or formate hydrogenlyase activity, improving substrate assimilation, eliminating competing pathways, and rebalancing intracellular reducing equivalents [46]. In *Clostridium pasteurianum* DSM525-H4, for example, the homologous overexpression of the H₂-evolving [FeFe]-hydrogenase gene (*hydA*), yielded 1.10 mol H₂ mol^{−1} glycerol, approximately 70% above the parental strain [21]. In *Enterobacter cloacae* WL1318, the homologous overexpression of the formate hydrogenlyase transcriptional activator gene (*fhlA*) increased H₂ production from cotton-stalk hydrolysate by 188% after 24 h and improved the substrate-specific H₂ yield by 37% [22]. In *Escherichia coli* WDH-GFA, the sequential deletion of the formate hydrogenlyase repressor (*hycA*), glucose-specific phosphotransferase-system transporter (*ptsG*), fumarate reductase membrane-anchor subunit (*frdD*), and fermentative D-lactate dehydrogenase (*ldhA*) improved the glucose–xylose co-utilization and H₂ production from mixed sugars and wheat-straw hydrolysate [23].

Although these reports highlight the potential of metabolic engineering to improve H₂ yield, the genetic stability and strain performance on agri-food residues remain critical challenges for industrial-scale implementation.

2.3. Thermochemical Biohydrogen Pathways for Agricultural Biomass

2.3.1. Gasification of Crop Residues and Livestock-Derived Biomass

Gasification converts biomass through interacting oxidation, devolatilisation, char conversion, and reforming reactions. Its hydrogen output depends on the gasifying agent, temperature, pressure, reactor design, and downstream conditioning. Hydrogen concentration in syngas must be distinguished from molar yield per mass of biomass and from purified hydrogen recovered. Nitrogen dilution can lower hydrogen fraction without an equivalent reduction in the underlying reaction yield. Consequently, AI studies should define the target and gas basis before comparing reactor configurations [4,47].

Agricultural residues introduce additional interactions through moisture, lignin, ash, and inorganic constituents. A change in feedstock blend can modify both reaction chemistry and heat demand; high-ash material can also constrain bed operation. AI models that use only temperature and a categorical biomass name may misattribute composition effects to reactor conditions. A useful feature set includes ultimate and proximate analyses, moisture, particle properties, and reactor metadata, with elemental and energy balances checked alongside prediction error [47].

A recent co-gasification study combines poultry litter with sewage sludge and couples upstream conversion to downstream hydrogen purification [48]. Poultry litter is relevant to agricultural valorisation, but sewage sludge broadens the feedstock domain beyond agri-food residues. This case is therefore used as explicit evidence of process-level AI integration, not as a universal demonstration for agricultural biomass. Its numerical contributions are evaluated in Section 4.3. Overall, gasification provides a clear AI task: predict hydrogen recovery within physically feasible operating and heat-supply limits, rather than optimise syngas concentration in isolation.

Air supplies oxidant economically but introduces nitrogen; oxygen reduces dilution while adding oxygen-production demand. Steam participates in char–steam and hydrocarbon-reforming reactions, and its influence depends on available heat and residence time. Increasing temperature can improve endothermic conversion and tar destruction but does not guarantee a monotonic increase in hydrogen fraction because water–gas-shift equilibrium, residence time, and feedstock chemistry also change. Pressure can alter methane formation, gas–solid transport, and integration with purification. Its effect cannot be summarised as a universal positive or negative coefficient. Table 2 links these factors to the features and constraints required in AI models [47].

Table 2. Gasification factors, hydrogen-related consequences, and implications for AI-driven optimisation.

Factor/Mode	Hydrogen-Related Consequence	Necessary AI Inputs	Evidence Basis
Air, oxygen, steam, or mixtures	Nitrogen dilution, oxidation heat, reforming, and shift reactions affect concentration and recovery differently.	Agent composition; equivalence ratio; steam/biomass ratio; wet/dry and N ₂ -free gas basis; auxiliary energy.	[4,47]
Temperature and heat supply	Endothermic conversion and tar reforming can increase; hydrogen fraction also depends on shift equilibrium and residence time.	Actual temperature profile; heat losses; energy balance; catalyst state; conversion limits.	[47]
Pressure	Changes chemical equilibrium, residence behaviour, and compression requirements; direction is system dependent.	Absolute pressure; partial pressures; gas density; equilibrium and transport descriptors.	[47]

Table 2. Cont.

Factor/Mode	Hydrogen-Related Consequence	Necessary AI Inputs	Evidence Basis
Autothermal operation	Partial oxidation supplies process heat but consumes feedstock/gas and changes syngas composition.	Oxidant feed; equivalence ratio; heat balance; carbon conversion; oxygen-production cost if applicable.	[47]
Externally heated/allothermal operation	Separates reaction heat from oxidant addition; external heat affects net hydrogen efficiency.	Heat-source temperature and efficiency; available duty; steam generation; heat-transfer limits.	[47]
Dual fluidised bed	Separates combustion and gasification; circulating solids provide heat and may influence catalysis.	Solids circulation; inter-reactor heat transfer; gas leakage; steam feed; bed inventory.	[47]
Fixed vs. fluidised beds and residence time	Mixing, particle conversion, tar carryover, and ash behaviour affect hydrogen-containing gas.	Bed type; particle size; gas/solid residence times; ash chemistry; entrainment and agglomeration limits.	[47]
Feedstock blend and downstream purification	Moisture, ash, organic composition, shift conversion, and separation determine recovered hydrogen.	Blend ratio; proximate/ultimate analysis; moisture; purification recovery; whole-process energy demand.	[4,48]

Directional effects are mechanistic tendencies, not pooled numerical effect sizes. Quantitative ranges depend on feedstock, heat supply, reactor, and measurement basis; no universal temperature, pressure, or mode ranking is inferred. The table is an original synthesis. Autothermal and allothermal describe heat supply, whereas fixed, fluidised, and dual-bed arrangements describe reactor configuration; these categories may overlap.

Future advances in biomass gasification require integration of catalysts, biomass characterisation, monitoring and data-driven optimisation. Hydrogen-oriented AI targets should distinguish gas fraction, yield per feedstock mass and purified recovery, and account for heat supply, gas cleaning and downstream separation. Harmonised experiments and demonstration under realistic operating conditions are needed to validate these models [47,48].

2.3.2. Pyrolysis and Reforming of Agri-Food Solid Wastes

Pyrolysis and catalytic reforming represent complementary thermochemical processes that have attracted increasing attention for the production of renewable hydrogen from agri-food solid wastes. Unlike gasification, which directly converts biomass into syngas under limited oxidizing conditions, pyrolysis is conducted in the absence of oxygen, producing three main fractions: a condensable liquid (bio-oil), a solid carbonaceous compound (biochar) and a non-condensable gas. It is characterised as an endothermic reaction typically occurring at temperatures between 300 and 1000 °C [49]. Pyrolysis can be categorised into fast, intermediate, and slow processes. Table 3 provides an overview of different pyrolysis types [50]:

Table 3. Overview of different pyrolysis types.

Pyrolysis	Condition	Product Yield
Slow	Solid residence time (hours/days), ~290–400 °C	Bio-oil: 10 wt% Water: 20 wt% Gas: 35 wt% Biochar: 35 wt%
Intermediate	Solid residence time (minutes), ~400 °C	Bio-oil: 25 wt% Water: 25 wt% Gas: 25 wt% Biochar: 25 wt%
Fast	Solid residence time (seconds), ~500 °C	Bio-oil: 60 wt% Water: 15 wt% Gas: 13 wt% Biochar: 12 wt%

Residence time is a crucial factor in defining the characteristics of each pyrolysis process. Torrefaction and slow pyrolysis require solid residence times ranging from hours to days. Intermediate pyrolysis operates over minutes, whereas fast pyrolysis is distinguished by an extremely brief solid/vapour residence time of approximately 2 s [49,51].

In slow pyrolysis, the low heating rate, relatively low temperatures (usually below 300 °C), and extended residence time collectively favour the biochar as the primary product [52].

Intermediate pyrolysis functions at moderate heating rates of 100–500 °C/min and temperatures of approximately 400 °C, with vapour residence times in seconds and solid residence times ranging from several minutes. It typically yields balanced ratios of gas, solid, and liquid products. Moreover, depending on process parameters and feedstock characteristics, it can produce energy vectors with improved physicochemical properties [53].

Intermediate pyrolysis differs from other pyrolysis methods in its ability to process finely milled feedstocks obtained through chopping, shredding, or grinding and bulkier materials like briquettes, pellets, and chips. Additionally, it can convert various feedstocks, such as sewage sludge, algae, grass, food and market waste, agricultural residues, industrial by-products, and digestate, into valuable products [50,54,55].

Intermediate pyrolysis enhances the thermal cracking of biomass components and facilitates more precise control over chemical reactions, leading to bio-oil with superior physicochemical properties. Additionally, fast pyrolysis faces difficulties processing feedstocks with high moisture content or larger particle sizes [56,57].

Fast pyrolysis is a continuous process defined by a very short vapour residence time and an elevated heating rate, reaching up to 1000 °C/s. Under these conditions, a high yield of liquid product is produced, approximately 75 wt% on a dry basis. Fast pyrolysis involves the rapid degradation of organic feedstock, forming char and pyrolysis gases [58,59].

In general, pyrolysis produces hydrogen-rich syngas at elevated temperatures across all conventional pyrolysis types by enhancing secondary cracking and devolatilization [60].

Hydrogen can also be achieved through the catalytic steam reforming of pyrolysis vapors or bio-oils rather than during the pyrolysis step itself [49,59,61]. This two-stage configuration offers greater flexibility in process control, allowing the independent optimization of thermal decomposition and hydrogen generation while facilitating the recovery of valuable co-products such as biochar and specialty chemicals.

Agri-food residues such as olive pomace, grape seed, fruit stones, chestnut shells, spent coffee grounds, tomato pomace and cereal processing residues have demonstrated considerable potential for pyrolysis because of their high volatile matter content and widespread availability within the food-processing industry [49,59,61]. Nevertheless, the marked heterogeneity of these feedstocks results in significant variations in bio-oil composition, affecting both reforming efficiency and hydrogen productivity.

The high oxygen content of bio-oils, typically ranging between 35 and 50 wt.%, remains one of the major limitations to efficient hydrogen production. Bio-oils are complex mixtures containing phenols, aldehydes, ketones, organic acids, furans, and anhydro sugars, whose thermal instability and tendency to polymerize promote catalyst deactivation through coke formation [50,62]. Consequently, catalytic steam reforming has become the preferred strategy for upgrading pyrolysis vapors into hydrogen-rich gas. Steam reforming is generally conducted between 700 and 900 °C, where oxygenated compounds are converted into H₂, CO, and CO₂ through a combination of steam reforming and water–gas shift reactions. Under optimized conditions, hydrogen concentrations exceeding 65 vol.% have been reported, although performance strongly depends on catalyst formulation, steam-to-carbon ratio, and feedstock composition.

An emerging trend is the integration of pyrolysis with advanced reforming technologies capable of simultaneously improving hydrogen yield and reducing greenhouse gas emissions. The Thermo-Catalytic Reformer (TCR) unit is a novel pyrolysis reactor that produces hydrogen-rich syngas and high-quality bio-oil. The development of TCR originated with the Pyroformer, an innovative intermediate pyrolysis unit. The Pyroformer consists of a horizontal auger-screw reactor equipped with two coaxially arranged rotating screws. The inner screw transfers a mixture of fresh feedstock and recycled biochar through the reactor. The outer screw returns biochar to the inlet zone, significantly improving heat-transfer efficiency within the feedstock. The biochar layer formed inside the reactor moderates the heating rate, protecting the feedstock from excessively high temperatures and thereby preventing undesirable rapid pyrolysis reactions. Subsequently, the Pyroformer was integrated with a vertical catalytic reformer or packed-bed reactor to form the TCR system in a single reactor configuration [55,61,63].

As illustrated in Figure 1, the TCR process comprises two essential stages:

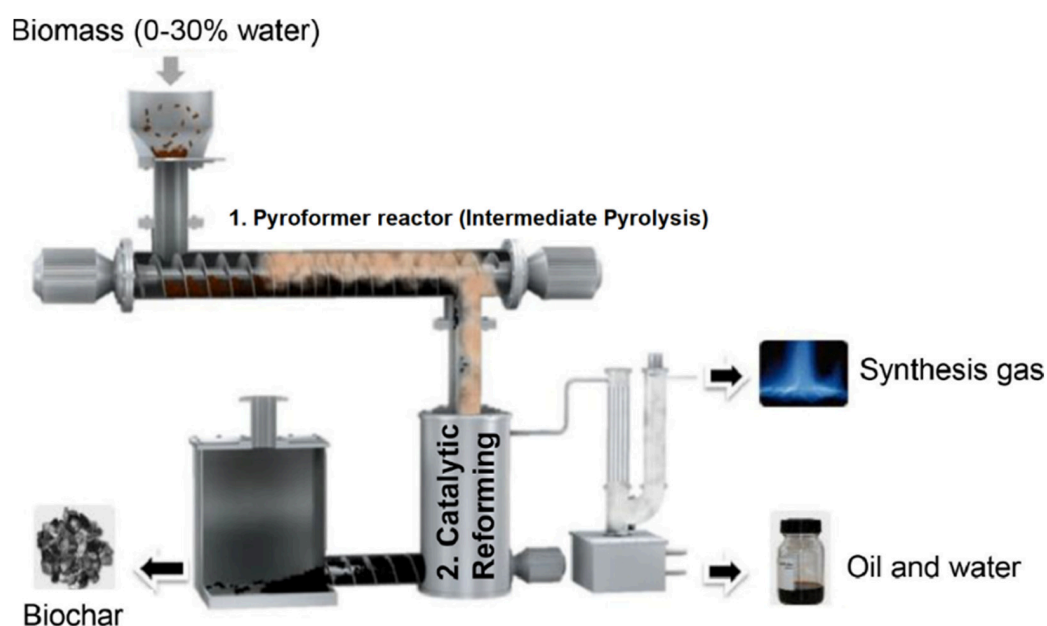


Figure 1. Thermo Catalytic Reforming (TCR) reactor.

1. Intermediate pyrolysis occurs at mild temperatures ranging from 400 to 500 °C, with solid residence times of approximately 5 to 10 min. During this stage, thermal decomposition of the feedstock takes place in the absence of oxygen.
2. Catalytic reforming follows at elevated temperatures between 500 and 700 °C, where vapour-phase catalytic cracking enhances the production of syngas and organic vapours with improved physicochemical properties. Biochar within the reformer serves as the catalytic agent during this process.

Food and market waste (FMW) is one of the most abundant unrecycled waste streams, posing waste management challenges and negative environmental impacts. In a TCR study, FMW was utilized as a feedstock, resulting in the production of 7 wt % organic bio-oil, 53 wt % syngas, and 22 wt % char. The syngas fraction consisted mostly of methane and hydrogen (23 vol% and 20 vol%, respectively) [55,61].

Chemical looping reforming, sorption-enhanced reforming, and catalytic pyrolysis are increasingly considered promising alternatives because they combine hydrogen production with in situ CO₂ capture or direct tar conversion, thereby simplifying downstream gas purification [49,59]. Likewise, microwave-assisted pyrolysis has attracted growing interest

because volumetric heating reduces temperature gradients within the biomass particles, potentially increasing gas quality while decreasing energy consumption. Although these technologies remain largely at the pilot scale, they demonstrate the ongoing transition from conventional reactor optimization towards integrated process intensification.

Despite substantial progress, pyrolysis–reforming systems remain less technologically mature than biomass gasification. Their commercial implementation is constrained by the complexity of multi-stage reactor configurations, the limited long-term stability of reforming catalysts, and the variability of bio-oil composition generated from heterogeneous agri-food residues [49,59,61]. Moreover, most experimental investigations have focused on maximizing hydrogen yield under controlled laboratory conditions, whereas comparatively fewer studies have evaluated process scalability, energy integration and economic feasibility. Future research should therefore prioritize catalyst durability, standardized bio-oil characterization and the integration of techno-economic and life-cycle assessments to identify process configurations capable of delivering both high hydrogen productivity and competitive production costs. Such integrated approaches will be essential for transforming pyrolysis–reforming technologies into viable industrial solutions for the sustainable valorization of agri-food solid wastes.

2.4. Process Complexity, Feedstock Variability, and Data Generation in Agri-Food Biohydrogen Systems

Despite the significant technological advances achieved in biomass gasification, pyrolysis, and catalytic reforming, the large-scale deployment of biohydrogen production systems remains constrained by the intrinsic complexity of thermochemical conversion processes. Unlike fossil-derived feedstocks, agri-food biomass exhibits substantial variability in physical, chemical and thermochemical properties owing to differences in crop species, cultivation practices, harvesting conditions, geographical origin and storage methods. Moisture content, particle size distribution, lignocellulosic composition, ash content and inorganic elements directly influence biomass conversion kinetics, heat and mass transfer, gas composition and catalyst performance, making process optimization considerably more challenging than for conventional fossil resources.

Feedstock heterogeneity is increasingly recognized as one of the principal sources of uncertainty in thermochemical hydrogen production. Agricultural residues such as wheat straw, rice straw, maize stover and olive pruning differ substantially in cellulose, hemicellulose and lignin contents, while livestock-derived biomass is generally characterized by higher moisture, ash and nitrogen contents. These compositional differences affect devolatilization, char formation, and secondary reforming reactions, ultimately leading to considerable variability in hydrogen yield and syngas quality [47,48]. Furthermore, alkali and alkaline-earth metals naturally present in biomass may promote ash melting, bed agglomeration, and catalyst deactivation, particularly in fluidized-bed gasifiers and catalytic reformers. Consequently, biomass characterization should no longer be considered merely a preliminary analytical step but rather an integral component of process design and optimization.

Process complexity is further amplified by the strong interactions among operating variables. Reactor temperature, heating rate, residence time, steam-to-biomass ratio, equivalence ratio, catalyst composition, and pressure simultaneously influence reaction pathways and frequently exhibit nonlinear interactions. Optimizing one operating parameter without considering its interaction with the others often results in suboptimal process performance. For this reason, recent investigations have increasingly adopted multivariate statistical approaches, design-of-experiments methodologies, and response surface analysis to identify optimal operating windows while reducing the experimental burden. These approaches provide a more realistic representation of industrial operating conditions than conventional

one-factor-at-a-time experiments and facilitate the identification of process interactions that are otherwise difficult to detect.

In parallel with advances in experimental methodologies, digitalization is emerging as a key enabler for improving the efficiency and robustness of thermochemical biohydrogen systems. Modern pilot-scale gasifiers and pyrolysis units are increasingly equipped with distributed sensors capable of continuously monitoring temperature profiles, pressure, gas composition, biomass feeding rates, and catalyst performance. These monitoring systems generate large volumes of operational data that can be integrated into predictive models for process optimization and control. Artificial intelligence (AI) techniques, including artificial neural networks, support vector machines and ensemble learning algorithms, have demonstrated considerable potential for predicting hydrogen production, gas composition and cold gas efficiency based on biomass properties and operating conditions. Compared with traditional regression models, machine-learning approaches are generally better suited to describing the nonlinear behaviour of thermochemical conversion processes and can support real-time operational decision-making.

Among emerging digital technologies, the concept of the digital twin has attracted growing attention within the bioenergy sector. A digital twin combines mechanistic process models with real-time operational data to generate a dynamic virtual representation of a physical plant. Such systems could enable continuous process monitoring, predictive maintenance, fault detection, and optimization under changing feedstock conditions. Although applications of digital twins to biomass gasification and biohydrogen production remain relatively limited, reported modelling frameworks suggest considerable opportunities for improving process reliability, reducing downtime and supporting scale-up from laboratory to industrial facilities [64]. The integration of digital twins with machine learning and advanced sensor networks may therefore represent an important step towards intelligent and autonomous biohydrogen production plants.

The increasing complexity of thermochemical conversion systems also highlights the importance of integrating techno-economic analysis (TEA) and life-cycle assessment (LCA) into process development. While laboratory studies frequently focus on maximizing hydrogen yield, industrial deployment ultimately depends on economic viability, energy efficiency, and environmental performance. Recent studies emphasize that feedstock logistics, biomass pre-treatment, catalyst replacement and gas cleaning account for a substantial fraction of production costs, whereas greenhouse gas emissions are strongly influenced by biomass transportation, process energy demand and co-product utilization [47,48]. Consequently, TEA and LCA should no longer be regarded as post-process evaluation tools but as integral components of technology development, supporting the identification of economically and environmentally optimal process configurations.

Overall, future progress in agri-food biohydrogen production will depend not only on improvements in gasification, pyrolysis or reforming technologies but also on the ability to manage feedstock variability through standardized biomass characterization, advanced process monitoring and data-driven optimization. The development of harmonized experimental protocols, publicly available databases, and hybrid modelling frameworks integrating first-principles models with machine-learning algorithms could improve the reproducibility and comparability of research results. Furthermore, combining digital process control with TEA and LCA will facilitate the identification of scalable and economically sustainable solutions capable of supporting the transition from laboratory-scale investigations to commercial biohydrogen production within the circular bioeconomy [47,48]. Thus, process-specific data and constraints define the AI tasks needed to improve hydrogen production.

3. ML Roles in Agri-Food Biohydrogen Systems

The useful AI task depends on whether the goal is to predict hydrogen, diagnose process instability, select operating conditions, or compare plant configurations. Model families are evaluated here in relation to those hydrogen-production decisions. A highly accurate fit to one experiment is different evidence from a validated operating recommendation or improved plant output.

3.1. Supervised, Unsupervised, and Deep Learning Paradigms

Supervised learning techniques continue to be the most commonly used AI approach in agri-food-based biohydrogen studies, mainly because of their excellent predictive capabilities and lower data needs. Several machine learning models such as artificial neural networks (ANNs), support vector regression (SVR), random forests (RF), gradient boosting machines (GBM) and Gaussian process regression (GPR) have been utilized to model hydrogen yield, production rate and metabolite profiles from food waste, agricultural residues and agroindustrial wastewaters. For instance, wastewater studies compare ensemble and regression models for nonlinear relationships among operating conditions, metabolites, and H₂ production [3]. Optimisation studies [65] and comparative prediction studies [3] suggest that the choice of model is feedstock- and data-specific.

Unsupervised learning methods, although less commonly used, are useful for investigating data and diagnosing problems in an ecosystem of agri-food biohydrogen. Clustering approaches such as k-means, hierarchical clustering and self-organizing maps have been employed to identify operating regions, substrate similarity and metabolic trends without labels. Zhang et al. [66] showed that a clustering of lignocellulosic hydrolysate compounds based on molecular descriptors can disclose, in an appropriate chemical and biochemical manner, groupings influencing the fermentation process. An important challenge that arises in unsupervised learning is the problematic interpretation of clusters and how they can be directly transformed into actionable process insights, especially when dealing with small or extremely diverse data sets.

Deep learning methods, such as multilayer perceptrons and deep neural networks (DNNs) or recurrent neural networks, are applied more often in complex biohydrogen systems with large datasets. Deep architectures are good at learning hierarchical features and being able to capture complex nonlinearity, an opportunity that requires sufficiently diverse training data. However, deep learning models are still data-demanding and computationally costly, making them generally not so practical in agri-food biohydrogen systems where experimental data are limited due to being expensive and even conflicting among studies. They face grand challenges such as overfitting, lack of interpretability, and difficulty in cross-scale generalization.

The supervised, unsupervised, and deep-learning approaches discussed above address different modelling tasks. Figure 2 connects these tasks to the process relationships that produce nonlinear hydrogen responses. Hydrogen yield depends on interactions among feedstock composition, operating conditions, microbial communities and metabolite accumulation. Seasonal variability changes substrate biodegradability, while pH, hydraulic retention time, and temperature affect microbial activity. Volatile fatty acids and alcohols can introduce inhibitory feedback. These relationships motivate multivariate and hybrid models, but improved prediction and transferability must be demonstrated rather than assumed [1,3,9].

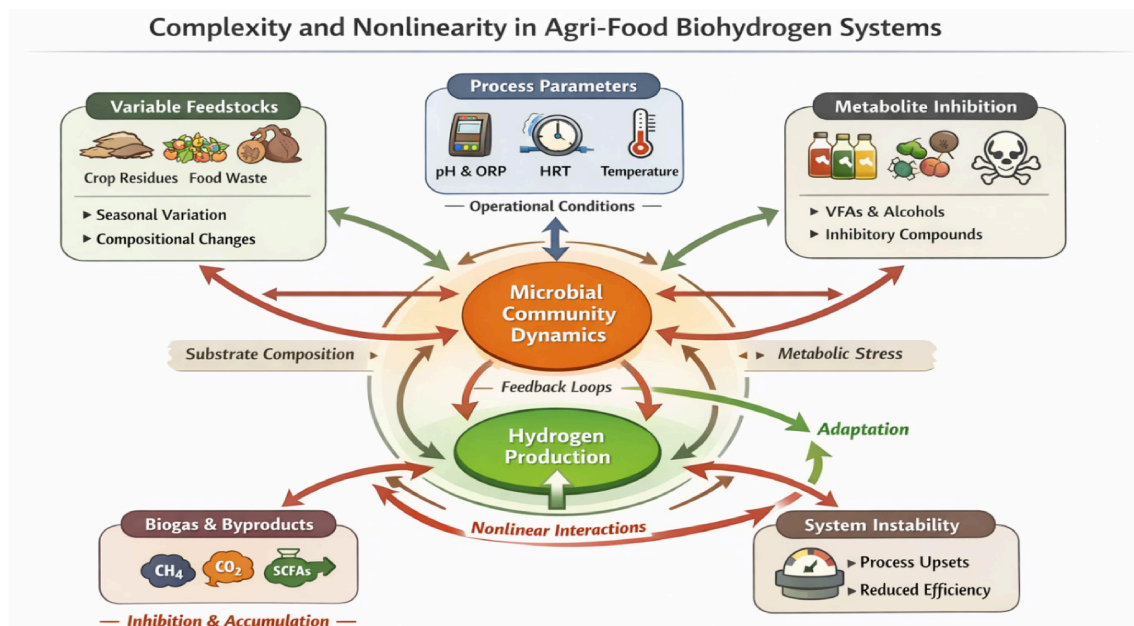


Figure 2. Complexity and nonlinearity in agri-food biohydrogen systems.

3.2. Evolutionary, Swarm Intelligence, and Optimization Approaches

Evolutionary algorithms and swarm intelligence are very important tools for agri-food biohydrogen systems to deal with complex multi-objective optimization problems that may not be easily solved by gradient-based methods. Moreover, genetic algorithms (GA), particle swarm optimization (PSO), grey wolf optimization, and ant colony optimization have frequently been combined in conjunction with ML models for the possible operating parameters, including pH, temperature, substrate loading, inoculum supplementation ratio, and HRT. Anand et al. [5] revealed that ANN–GA models can have the potential advantage beyond conventional response surface methodology for improving hydrogen production at nonlinear/mechanically coupled operating ranges. However, the coffee-mucilage comparison reported a higher R^2 for fuzzy logic (>0.8485) than ANN (>0.7866), illustrating that ANN superiority is not universal [67].

The above-mentioned properties make swarm intelligence algorithms very appealing in the case of agri-food systems, as they have been proved to be able to face local optima and noise. Recent explainable optimisation research compares metaheuristic-assisted models for hydrogen production [68]. Nonetheless, such approaches are usually computationally involved and sensitive to parameter tuning, while the stochastic nature implies different optimization results under different runs and is a burden for reproducibility as well as industrial application [65].

Bayesian optimization (BO) has gained popularity as a viable option for data-efficient optimization of biohydrogen systems, especially in situations where experimental data are scarce. The combination of probabilistic surrogate models and acquisition functions enables BO to perform systematic exploration–exploitation trade-offs. Mougari et al. reported high BO–ANN predictive performance using kinetic descriptors and five-fold validation. Though Bayesian optimization sounds promising, it is largely under-explored in agri-food biohydrogen research, and the major challenge is to define the right priors and deal with high-dimensional spaces (decision space, as well as integrating uncertainty into real-time control [6].

3.3. AI-Driven Prediction and Performance Evaluation

AI-based prediction is the most mature application of machine learning to agri-food biohydrogen systems. Various supervised learning models such as artificial neural networks, random forests, support vector regression and ensemble methods have been used to predict hydrogen yield, production rate, gas composition and substrate conversion efficiency for different agri-food residues [3,4,69]. These models allow quick predictions of system performance in different operating conditions and feedstock compositions, helping to prioritise further experiments. They are capable of modeling non-linear effects, reflecting the influence between performance variables such as pH, organic loading rate, volatile fatty acid accumulation, and trace metal availability that is hard to model with conventional techniques.

Prediction aside, AI methodology has been more and more applied to process diagnosis and performance assessment. Classification models, anomaly detection algorithms and feature-importance approaches have been employed to detect operational instability, inhibitory conditions and dominant process variables. Recent explainable optimisation research [68] illustrates how explainable AI tools, namely SHAP and permutation-based sensitivity analysis, are important for understanding model behavior and enhancing trust in AI-assisted diagnostics. However, diagnostic solutions are constrained by the small amounts of labeled fault data; poorly reported experimental conditions and the prevalence of steady-state datasets mean that AI models cannot capture transient dynamics and early-warning signals.

3.4. Process Optimization

AI-driven process optimization is a promising tool for navigating such complex multi-objective decision spaces as are common to agri-food biohydrogen systems. Machine learning (particularly surrogate models) has been increasingly employed for optimization studies to find optimal operating conditions, trade off the H₂ yield and stability of the process, and mitigate computational effort in large-scale system analysis. Shi et al. [48] demonstrated that surrogate modeling using an ML-based approach can remarkably increase computational speed for techno-economic and environmental optimization of bioenergy systems by substituting expensive computer simulations. This capacity is particularly beneficial for agri-food applications, where the variability of feedstock and operational uncertainty requires flexible and adaptable optimization strategies.

At the process and system scale, AI has some capabilities to provide operational assistance, quick decision-making in real time, and to digitize biohydrogen agri-food facilities. New ideas like AI-supported decision support systems and digital twins offer scenario-based analysis, predictive maintenance, or adaptive control during varying feedstock supply and in the presence of process upsets. However, this opportunity has not yet been realised in the industry due to the issues of data integration across various production sections, sensor reliability, model transferability, and cybersecurity. For future implementation, next-generation agri-food biohydrogen plants will need transparent, uncertainty-aware, and physics-informed AI toolboxes to assess robust operation over the life span of a project as well as regulatory acceptance and extended industrial scalability [64].

Table 4 replaces broad algorithm lists and estimated dataset-size ranges with study-level evidence. Reported scores are retained with their task and validation limitations. Numerical comparisons across rows are inappropriate when targets, normalisation, or sampling differ. Studies based on accessible abstracts or primary institutional records are identified so that missing full-text details are not silently treated as known. The table also distinguishes direct agri-food evidence from related wastewater applications.

Table 4. AI evidence relevant to hydrogen prediction and optimisation: quantitative outcomes and limits of interpretation.

AI Approach	Process Domain	Data	Reported Quantitative Contribution	limitation	Reference
GBM, SVM, RF, AdaBoost; prediction	Wastewater DF; agricultural provenance not established here	Primary abstract; grid search and permutation importance; sample count/split not assigned	R^2 : 0.893, 0.885, 0.902, 0.889; MSE: 0.015, 0.015, 0.016, 0.015, respectively	Prediction evidence; MSE scale follows source; no measured yield gain inferred	[3]
ANN vs. fuzzy model; prediction	Coffee mucilage with fruit/vegetable wastes	Source-reported experimental/model fit; independent transfer not established	$R^2 > 0.7866$ (ANN); >0.8485 (fuzzy)	Counterexample to universal neural-network superiority	[67]
ANN–GA vs. RSM–CCD; condition selection	Food-waste hydrolysate DF; C. pasteurianum	Experimental study; source reports outcomes under selected conditions	1108 vs. 1039 mL/L; fitted maximum rates 185.34 vs. 153.74 mL L ^{−1} h ^{−1}	Within-study process comparison; fitted rates are not direct steady-state measurements	[5]
BO–ANN with Gompertz descriptors; prediction	Literature-compiled organic-waste DF	Primary abstract; 5-fold cross-validation; grouping details not assigned	R^2 0.9980; RMSE 0.0117; MAE 0.0062	Source target scale; check whether kinetic features are available before prediction	[6]
SVR, XGBoost and metaheuristics; prediction/optimisation	Biohydrogen fermentation; exact feedstock provenance not resolved in accessible record	Primary institutional abstract; separate train/test scores	SVR test R^2 0.974 vs. XGBoost 0.856; optimum 95.1746 mL/g TS	Optimised value is a model result; external/feedstock validation not inferred	[68]
ANN surrogates + MILP; system optimisation	Poultry litter/ sewage-sludge co-gasification and purification	1400 simulation points; two subprocess surrogates; primary institutional abstract	R^2 0.98/0.99; optimisation 45–100 s vs. simulation-based heuristic 3–6 h	Computational evidence within a simulator; not measured plant speed or yield gain	[48]
Descriptor-based ML/clustering; characterisation	Acid-pretreated cornstalk cultivation medium	50 compounds; primary publisher abstract	Chemical grouping and influential descriptors reported; no yield-gain percentage assigned	Supports feature design; not a closed-loop optimisation demonstration	[66]

Table 4. Cont.

AI Approach	Process Domain	Data	Reported Quantitative Contribution	limitation	Reference
MLP-ANN; prediction	Organic-waste DF	Primary publisher record; details not assigned from inaccessible full text	No numerical score transcribed without unambiguous verification	Recent directly relevant predictive study; quantitative ranking withheld	[69]
Augmentation + HGB; prediction	Wastewater DF; related methodological evidence	Primary publisher abstract; synthetic-data contribution requires independent real-data testing	Test $R^2 \approx 0.95$; RMSE < 0.06; MAE < 0.05	Source-scale errors; augmentation does not substitute for new experiments	[70]

R^2 is the coefficient of determination; MSE, RMSE, and MAE depend on the source target scale and units. Scores labelled only as reported are not reclassified as independent external tests. Values reproduced from primary abstracts or institutional publication records have not been supplemented with invented sample sizes, splits, or uncertainty intervals. The evidence supports different functions; it does not establish an algorithm ranking across pathways. This is an original comparative synthesis, not a reproduced third-party table.

The next section links these AI tasks to hydrogen-production process models and constrained decisions.

4. Hybrid AI-Process Modeling

The following hybrid approaches connect AI predictions to hydrogen-production constraints and engineering decisions.

4.1. AI-Based Models for Agri-Food Experimental and Simulation-Driven Systems

AI-based models are often employed for the approximation of computationally expensive experimental or simulation-based biohydrogen systems; this is especially the case in agri-food phenomena where there exist large parametric spaces and nonlinear effects prevail. Surrogate models trained on the outputs of biochemical experiments, Aspen Plus simulations, as well as CFD-based gasification models can accelerate repeated evaluations within their validated domain [48]. Surrogate models based on ANN and GPR have also been developed to mimic microorganism kinetics, metabolite profiles in a fermentation system with different substrate levels, as well as real operational conditions for fast sensitivity analysis and optimization [6,71].

Although surrogate models are powerful, they encounter significant issues in agri-food applications. The training data are typically obtained from a small process window, or idealised simulations that do not account for real-world variations in feedstock properties and operational disturbances. The surrogate estimation may significantly deteriorate when being extrapolated out of the training domain, especially for heterogeneous agricultural residues. As a result, the development of strong surrogates demands careful experimental design, including uncertainty-sensitive modelling techniques and frequent updating with real operational data to remain adaptable in the face of changing agri-food supplies.

In experimental screening, a predictor estimates hydrogen outcomes over a defined design space and an optimiser proposes candidates for confirmation. This approach is strongest when the confirmation experiments use new batches or conditions and measure the same outcome used in optimisation. ANN-GA selection for food-waste hydrolysate

reported 1108 rather than 1039 mL/L under the compared ANN–GA and RSM–CCD conditions, corresponding to a calculated 6.64% increase. The fitted maximum-rate comparison corresponds to a calculated 20.55% increase [5]. These percentages are calculated here from source values and remain specific to that study; the fitted-rate difference is not a measured steady-state productivity increase.

Figure 3 represents these alternatives with separate data and mechanistic inputs. The validation block includes physical checks, prediction error, uncertainty, and candidate verification. It deliberately avoids prescribing a CNN, LSTM, or other architecture for every dataset. Deep temporal models may be appropriate for monitored trajectories, while simpler models may better match sparse batch experiments. The hydrogen-production benefit depends on the validated task rather than the depth of the network.

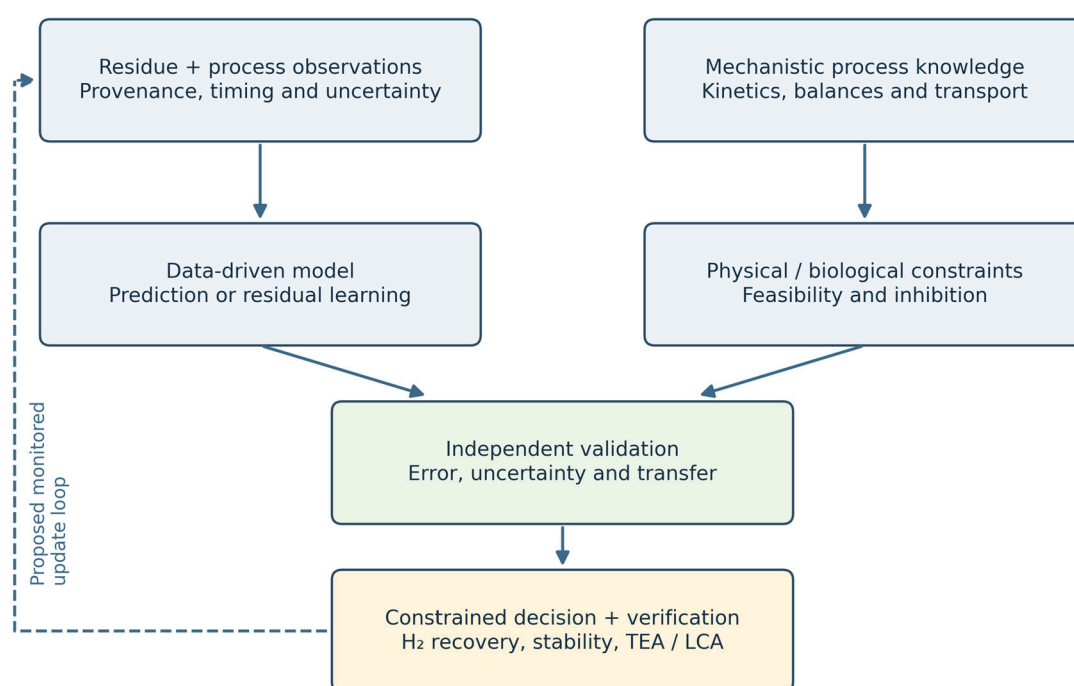


Figure 3. Proposed hybrid AI–process framework for residue-derived hydrogen. Original conceptual architecture: mechanistic and data-driven branches meet in validation and constrained decisions. The dashed feedback represents a future monitored deployment requiring prospective testing, not an implemented digital twin.

4.2. Integration of AI with Kinetic Models

The connection between AI methodologies and kinetic, biochemical, and thermodynamic models is seen as a serious approach to improving predictive accuracy together with interpretability in biohydrogen in agri-food systems. There have been a number of mechanistic models for other processes such as microbial growth, substrate degradation, hydrogenase activity, and metabolite formation that can be useful in capturing domain-specific knowledge but often suffer from parameter uncertainty or are based on oversimplified assumptions. AI has been more prevalent in estimating kinetic parameters, correcting model residuals, and inferring missing functional form within these models [4,6,72].

Kinetic integration may use mechanistic parameters as features, estimate uncertain rates, or learn residual corrections to differential equations. The BO–ANN study uses modified-Gompertz descriptors and five-fold cross-validation, reporting $R^2 = 0.9980$ [6]. That is evidence for its reported predictive task, not proof of prospective control across new feedstocks. Because fitted maximum production, lag time, or rate parameters may already encode the target trajectory, their availability must be audited. If they require the full

experiment, the model should be described as retrospective reconstruction or parameter-assisted prediction; prospective use requires independently measured descriptors or an early-data estimation procedure.

In thermochemical pathways, hybrid AI–thermodynamic models can combine gas-composition prediction with process constraints. Interpretable machine learning has been applied to supercritical water gasification [4], while ANN surrogates have been coupled to process optimisation [48]. These applications should be distinguished from models that explicitly enforce thermodynamic laws. Physical consistency requires checks of mass and energy balances, feasible compositions and relevant equilibrium or rate limits. Yet, challenges still remain in merging data-driven flexibility with mechanistic consistency, especially when the biological submodels are linked in integrated agri-food energy systems.

4.3. Coupling AI with Mathematical Programming

The combination of AI with mathematical programming tools like mixed-integer linear (MILP) and nonlinear programs (MINLP) has provided avenues for optimal system-level design of agri-food biohydrogen production networks. In such a hybrid setting, AI models can be employed as fast-response surrogates of complex unit operations so as to avoid utilisation of detailed simulations within the optimisation loops. This method has been used to enhance H₂ generation, cost, GHG emissions, and resource allocation in agri-food waste-to-energy systems [48].

Mathematical programming connects unit-operation predictions to objectives such as purified hydrogen recovery, cost, and environmental performance. The co-gasification study used 1400 simulated cases, ANN subprocess models with R² of 0.98 and 0.99, and MILP optimisation [48]. Reported solution times were 45–100 s, compared with 3–6 h for simulation-based heuristic optimisation. Sustainability indices changed from 0.79 to 0.80 upstream and 0.84 to 0.91 downstream. These are model-based computational and objective improvements, not measured operating-plant gains.

This example demonstrates how AI can reduce the cost of exploring hydrogen-production configurations. However, optimisation time must be compared using the reported problem formulation, hardware, convergence tolerances, and accounting for surrogate-training effort. The broad source ranges should not be reduced to one universal speed-up factor. Before a solution is applied, it should be checked in the detailed process model, including material/energy feasibility and the uncertainty of feedstock supply. Non-linear surrogate outputs also need an explicit representation if they are incorporated into a linear or mixed-integer formulation.

Despite the potential, AI–optimization combinations create new streams of uncertainty and bias. Search space inaccuracies might propagate to the optimization solution, resulting in suboptimal or infeasible solutions when not adequately restricted. In addition, the incorporation of stochastic agri-food feedstock availability and market variability into deterministic optimization schemes is still a pending issue. The importance of uncertainty-aware optimization and multi-scenario analysis to guarantee robustness in decision-making under real agri-food systems is starting to be highlighted.

4.4. Digital Twins and Virtual Agri-Food Biohydrogen Plants

The digital twins are an emerging hybrid modeling paradigm that combines AI, physical models, and currently monitored data streams to develop the virtual replication of real-life biohydrogen systems. In the agri-food domain, digital twins have been suggested as a means for tracking fermentation reactors, gasifiers, and integrated waste-to-energy plants under varying feedstock conditions on the fly [64]. The combination of AI-based

prediction and process understanding with a digital twin could support scenario testing, diagnosis of faults, and predictive maintenance without disturbing plant operation.

However, the application of digital twins in agri-food biohydrogen plants is generally a concept. Some of the major areas of advancement will be sensor positioning in challenging biological environments, data standardization across unit operations, cyber security, and model coordination over long operational periods [64]. A modelling framework alone does not demonstrate pilot or industrial deployment. Operational claims require an identified physical asset, measured update streams and prospective evidence of benefit. Closing this gap relies on conjoint progress in data infrastructure, AI robustness and process instrumentation that is dedicated to the specific characteristics of agri-food bioenergy systems.

The proposed deployment loop in Figure 3 is therefore shown as a future validation stage. Evidence for an operational twin would need to identify the physical reactor, sensors, update interval, deployment duration, independent prediction performance, intervention policy, and comparison with routine control. Prospective operator-supervised testing should precede autonomous operation. Pilot fermentation evidence [31] demonstrates the importance of real reactor physics, but does not by itself demonstrate digital-twin or AI-controlled performance. AI-driven hydrogen production should retain this separation between plausible capability and measured implementation.

4.5. Data Sources, Feature Engineering, and Validation Strategies

Insufficient data supply and quality are major bottlenecks for hybrid AI–process modeling of agri-food biohydrogen systems since the majority of studies are based on limited experimental datasets or secondary literature data lacking standardized reporting, detailed operational metadata, and microbial characterization [3]. The constraints in the dataset are further exacerbated by seasonal and compositional variation of agri-food residues, making data integration difficult and requiring careful preprocessing and normalization. Feature engineering has hence emerged as an essential facilitator of model generalization, and research studies have indicated that kinetic descriptors, molecular features, and stoichiometric ratios provide significantly improved predictive performance compared to raw operating variables alone [4,66]. Nevertheless, validation is not yet pursued strictly enough, since many models are still developed on the basis of random train–test splits, which do not always represent real-world extrapolation issues, and we believe that cross-feedstock validation, uncertainty quantification, and scale-sensitive benchmarking are currently key steps towards translating hybrid AI models from laboratory-based to industrial agri-food biohydrogen plants.

Table 5 summarises hybrid forms by where process understanding enters and what must be validated. This replaces an extensive list of loosely related studies with a comparison of integration mechanisms. The rows distinguish evidence already reported from proposals in this review. Feature engineering is valuable when it provides information available at the intended decision time; it becomes problematic when retrospective outcome information leaks into a supposedly prospective predictor.

AI value in biohydrogen production is therefore supported by a sequence of evidence: trustworthy data, an operationally meaningful target, independent prediction, physical feasibility, and verified decisions. Hybridisation can help organise this sequence, but it does not eliminate the need for new residue-specific and scale-specific observations.

The practical test of hybrid AI is whether it supports a validated improvement in hydrogen recovery, stability, or resource use.

Table 5. Hybrid integration mechanisms and validation requirements for AI-assisted hydrogen production.

Hybrid Form	Where Process Knowledge Enters	Hydrogen-Related Decision	Required Validation	References
ML-assisted experiments	Designed process variables and new confirmation runs	Select operating conditions for residue-fed fermentation	Compare measured outcomes with baseline; ANN–GA food-waste case provides within-study comparison	[5]
Mechanistic simulator + ML surrogate	Simulator supplies training scenarios and physical assumptions	Screen conversion/purification configurations	Separate surrogate error from simulator-to-plant error; co-gasification evidence is simulation based	[48]
Kinetic-descriptor learning	Gompertz parameters provide substrate-response features	Predict fermentation output or trajectories	Check descriptor availability and target leakage; reported five-fold CV is not cross-scale testing	[6]
Mechanistic model + classification	Process model and observed operating classes	Identify biological operating regimes	Validate class labels, diagnostic false alarms, and transfer; do not infer universal fault detection	[72]
Chemical descriptors + ML	Compound properties represent hydrolysate chemistry	Characterise substrate/inhibitor patterns	Chemical association does not prove a yield-changing intervention	[66]
Thermochemical prediction/constrained design	Operating conditions and explicitly defined physical feasibility	Predict feasible gas/hydrogen outputs	Distinguish interpretable prediction from direct enforcement of thermodynamic laws	[4,47]
AI + mathematical programming	Process surrogates enter system objectives and constraints	Choose configuration/cost–performance trade-offs	Recheck selected solution in detailed model; quantify optimisation error and training effort	[48]
Monitored virtual plant/digital twin	Proposed live data, state estimates, and updated process models	Potential diagnostic and operating recommendations	Prospective sensor-connected deployment must be demonstrated; conceptual capability is insufficient	[64]

5. Reliability, Scalability, and Industrial Relevance

For industrial hydrogen production, predictive accuracy is one requirement among several. Reliability also includes physical feasibility, stability under composition changes, quantified uncertainty, and the availability of inputs during operation. The numerical evidence in Table 4 demonstrates predictive or computational contributions under study-specific conditions; it does not establish a general industrial performance guarantee. This section explains why apparently successful AI models may fail when metabolic or thermodynamic conditions change.

The accuracy of AI-derived models for biohydrogen agri-food systems is only the beginning of the reliability test, which must also include the robustness of the solution, its interpretability, and physical coherence. High R^2 within a laboratory dataset does not establish performance outside its training domain. Wastewater-fermentation prediction [3] and interpretable SCWG modelling [4] demonstrate study-specific predictive applications; they do not establish universal resistance to feedstock changes. Uncertain parameters and identifiability in kinetic–AI hybrids could be a source of lost confidence if not examined. Accordingly, dependability of agri-food biohydrogen systems is heavily based on transparency of the hybridisation approach, independent validation and uncertainty assessment rather than algorithmic complexity.

A model can lose accuracy because the process itself changes regime. In DF, VFA accumulation lowers pH and increases the undissociated fraction of some acids, potentially

inhibiting uptake and hydrogen-producing activity. Community shifts toward lactate, ethanol, solventogenesis, or hydrogen-consuming populations change the relationship between substrate conversion and net hydrogen. A predictor trained within a stable acidogenic regime may extrapolate incorrectly across that shift even if its original fit is accurate [1,7,9].

A biology-informed architecture could retain dynamic balances for fermentable substrate, biomass, individual acids, and gas production. AI would estimate uncertain uptake or inhibition terms within those balances, rather than directly predicting an unrestricted hydrogen value. For a monoprotic acid, the approximate undissociated concentration can be computed from total concentration and pH as $C_{HA} = C_{total} / [1 + 10^{(pH - pK_a)}]$, subject to the relevant solution assumptions. Using species-specific acid states is more informative than total VFA alone. Any inhibition function and threshold must be fitted for the organism, medium, and temperature; the proposed architecture does not reverse inhibition physically.

Seasonal nitrogen variability creates different problems across pathways. In PF, changes in ammonium availability can suppress nitrogenase-mediated hydrogen evolution, so a model should represent nitrogen species and C/N context alongside light and carbon availability [24]. In DF or MECs, nutrient deficiency and excessive nitrogen-derived inhibition depend on organism and operating conditions [9,34]. Total nitrogen should therefore not be assigned one universal adverse effect. A model intended to represent ammonia stress should distinguish total ammoniacal nitrogen from free ammonia, using pH- and temperature-dependent speciation and a validated inhibition relationship.

Compensation requires an operational response, not merely a better equation. A constrained decision layer could recommend lower loading, dilution, buffering, blending, or an upstream nitrogen-removal step when predicted inhibition becomes unacceptable. These recommendations must be evaluated against cost, energy, and hydrogen recovery. Independent disturbance experiments are needed to test whether the model correctly predicts onset, recovery time, and intervention benefit. When process measurements cannot distinguish competing mechanisms, uncertainty should widen rather than force a precise recommendation.

Hydrogen partial pressure affects the feasibility of hydrogen-generating reactions and links biochemical behaviour to gas removal. A thermodynamically informed model can check $\Delta G = \Delta G^\circ + RT \ln Q$ using a defined reaction, temperature, and activity/partial-pressure basis. A trajectory should not imply sustained flux through an unfavourable reaction without an explicitly justified coupled mechanism. Such checks constrain predictions, but kinetics, enzyme activity, and mass transfer still determine the realised rate [7,46].

The scalability issue is even worse since the majority of AI applications in biohydrogen have not yet been extended from laboratory or pilot scale. One partial solution is simulation-based surrogate modelling. Shi et al. reported ANN–MILP optimisation times of 45–100 s, compared with 3–6 h for the thermodynamic-model/heuristic approach in their waste-resource hydrogen system [48]. This is a computational comparison, not a measured industrial hydrogen-yield improvement. Surrogates remain bound by the accuracy of the simulator and the relevant training domain. Collectively, these limitations indicate that scale-up requires scale-aware validation and mechanistic models that capture reactor physics and process constraints.

From an industrial point of view, the importance of AI in agri-food biohydrogen systems is ultimately characterized by the extent to which it succeeds in assisting operational and strategic planning decisions under real-time constraints. Shi et al. [48] showed that AI-empowered surrogate models can greatly speed up system-of-systems-level optimisation and provide cost-and-emission trade-off-based decision support for the AFWN in MILP

contexts. At the operative level, Monroy et al. demonstrated that AI-based monitoring can identify off-range conditions and stability losses in photo-fermentation systems; however, these methods were dependent on site-specific data range values and controlled environments [72]. Digital-twin frameworks [64] frame digital twins as an attractive means for real-time decision support and predictive maintenance, whose industrial uptake remains constrained by issues related to data infrastructure, sensor reliability, and cyber security. Collectively, these works suggest that the uptake of AI for agri-food biohydrogen systems in industry will not be driven so much by further methodological advances but by seamless data integration with existing feedback and feed-forward control architectures, model transparency and interoperability with mainstream engineering practice in the context of techno-economic and regulatory frameworks specific to the agri-food sector.

6. Research Gaps and Future Perspectives

A challenge in the context mentioned above across AI-based agri-food biohydrogen applications is not only to develop robust generalization, but also reliable performance to real-world variability. Although strong predictive performance under laboratory conditions is observed in many models, these figures frequently do not take into account the effects of seasonal feedstock variability and compositional heterogeneity present in agri-food residues. In dark fermentative systems, Models trained in narrow operating windows require independent tests on new substrates or changed conditions [3]. For thermochemical pathways, cross-feedstock and scale transfer require separate validation [4,48]. While emerging research is advocating these challenges, there is a lack of systematic uncertainty quantification, cross-feedstock validation, and scale-aware benchmarking [3,4]. So they are inconsistently applied and do not have any standard evidence behind them, hindering trust in industrial application of AI.

An enduring challenge involves the extent to which AI models may incorporate process-level knowledge in a meaningful way. The majority of hybrid AI-process frameworks presented in the literature follow loosely coupled approaches, where the AI is performed as a surrogate or optimization layer. ANN surrogates demonstrate computational benefits within simulated domains [48], while explicit physical consistency still requires validation. In spite of these advancements, the direct inclusion of kinetic constraints, thermodynamic laws, and biological features in learning architectures is still not a widespread practice. Some successful feature engineering approaches based on molecular descriptors and kinetic fingerprints have been used to enhance cross-feedstock learning [6,66]. These are not yet widely deployed, and adaptive or reinforcement learning methods for dynamic agri-food biohydrogen systems are largely unexplored.

Future progress in AI-enabled agri-food biohydrogen systems will require a transition toward decision-centric, scalable, and uncertainty-aware modeling frameworks that align with industrial and policy-relevant objectives. Physics-informed and biology-informed AI architectures offer a promising pathway to enhance extrapolation capability while maintaining interpretability and physical consistency. At the system level, tighter integration of AI with techno-economic analysis, life-cycle assessment, and supply-chain optimization, an approach explored by Shi et al. [48], can support strategic planning and investment decisions in agri-food contexts. In parallel, the development of digital twins and real-time decision support platforms, as proposed in digital-twin frameworks [64], provides a foundation for predictive maintenance and operational resilience, provided challenges related to data infrastructure, sensing reliability, and cybersecurity are addressed. Ultimately, advancing AI from an analytical tool to a reliable enabler of sustainable agri-food biohydrogen systems will depend on coordinated progress in data standardization, hybrid modeling,

and interdisciplinary collaboration across process engineering, data science, and agri-food system management.

7. Conclusions

This review presents a broad-based and critical overview of the AI-enabled biohydrogen production from agri-food residues, covering: biohydrogen pathways; machine learning approaches; hybrid AI–process modeling systems; their accuracy and industrial applicability. Through critical review of biochemical and thermochemical conversion pathways, as well as data-driven and hybrid modeling frameworks, the complexity associated with agri-food biohydrogen systems is shown to stem from feedstock non-linearity, carbon-conversion pathways, and multi-scale synergies. Selected studies report useful results compared to traditional modeling techniques for prediction, optimization, and decision support; but their actual power is becoming more and more evident in smart hybridization with process insights rather than pure predictive capabilities. The results suggest that hybrid AI–process models, surrogate-based optimization, and interpretable learning frameworks can help address the variability and uncertainty that come with agri-food systems.

To effectively realize AI within biohydrogen for agri-food production, it will be data ecosystem-driven, including physically informed modelling and decision-centric operational integration spanning the value chain. Innovations in feature engineering, uncertainty-aware learning and scale-aware validation need to be combined with techno-economic, environmental and supply-chain analyses for industrial relevance and regulatory acceptance. Emergent ideas like digital twins and AI-assisted optimization seem to provide two promising threads in this direction of real-time operational guidance and system resilience, but their success will depend on the quality of data as well as the transparency of models and crosstalk among disciplines. In conclusion, this review highlights the role of AI not as a substitute for process understanding but as a driving force toward scalable, reliable, and sustainable agri-food biohydrogen systems contributing to the roadmaps towards circular bioeconomy and low-carbon energy infrastructures.

Author Contributions: Conceptualization, M.R., M.T., G.A. and F.G.; investigation, M.R., H.J., T.Z., M.G. and F.G.; visualization, M.R., M.T., H.J., A.M., G.C.D.R. and G.A.; writing—original draft preparation, M.R., F.G., G.C.D.R. and T.Z.; writing—review and editing, M.R., T.Z., F.G., A.M. and M.G.; supervision, M.R. and F.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: No new experimental datasets or predictive models were generated in this narrative review. Reported process and model-performance values originate from the publications cited in the text and tables. The conceptual figures do not contain new experimental data.

Acknowledgments: ChatGPT Pro (OpenAI; accessed 6 October 2026) assisted with revision drafting, response preparation, and code for the new conceptual Figure 3. No generative-image tool was used for Figure 2.

Conflicts of Interest: The authors declare no conflict of interest.

Nomenclature and Abbreviations

Term	Meaning
AD	Anaerobic digestion
AFBR	Anaerobic fluidised-bed reactor

AI	Artificial intelligence
ANN/MLP-ANN	Artificial neural network/multilayer-perceptron ANN
BO	Bayesian optimisation
BRO	Battle Royale Optimisation
CCD	Central composite design
C/N	Carbon-to-nitrogen ratio; basis must be specified
COD	Chemical oxygen demand
CSTR	Continuously stirred-tank reactor
CV	Cross-validation
DF/PF/LF	Dark fermentation/photo-fermentation/lactic fermentation
GA	Genetic algorithm
GBM/HGB	Gradient boosting machine/histogram-based gradient boosting
H ₂ /CO ₂ /CH ₄ /N ₂	Hydrogen/carbon dioxide/methane/nitrogen
HRT	Hydraulic retention time
LCA/TEA	Life-cycle assessment/techno-economic analysis
LED	Light-emitting diode
LSTM/CNN	Long short-term memory/convolutional neural network
MAE/MSE/RMSE	Mean absolute error/mean squared error/root mean square error
MEC	Microbial electrolysis cell
MILP	Mixed-integer linear programming
ML	Machine learning
NR	Not retained or not assigned in this synthesis; see table notes
OLR	Organic loading rate
POME	Palm-oil-mill effluent
RF	Random forest
RSM	Response surface methodology
SCWG	Supercritical water gasification
SHAP	SHapley Additive exPlanations
SVM/SVR	Support vector machine/support vector regression
TCR	Thermo-catalytic reforming
TS/VS/TVS/VSS	Total solids/volatile solids/total volatile solids/volatile suspended solids
UASB/UASFF	Up-flow anaerobic sludge blanket/up-flow anaerobic sludge fixed-film reactor
VFA	Volatile fatty acid
XGBoost	Extreme gradient boosting
C _{HA} /C _{total}	Undissociated/total acid concentration on the same basis
pK _a	Acid dissociation constant expressed as $-\log_{10} K_a$
$\Delta G/\Delta G^\circ$	Reaction Gibbs energy/standard reaction Gibbs energy (consistent energy basis)
R/T/Q	Gas constant/absolute temperature/dimensionless reaction quotient
R ²	Coefficient of determination
NmL	Normalised gas-volume unit as reported by a source; reference conditions must be checked
PFOR/NFOR	Pyruvate:ferredoxin/NADH:ferredoxin oxidoreductase
PFL/FHL	Pyruvate formate lyase/formate hydrogen lyase
Fd/Fdred	Ferredoxin/reduced ferredoxin
PNSB	Purple non-sulfur bacteria
PSO/GWO	Particle swarm/grey wolf optimisation
GPR/DNN	Gaussian process regression/deep neural network
CFD	Computational fluid dynamics
MINLP	Mixed-integer nonlinear programming
ORP	Oxidation–reduction potential

PCA/MPCA	Principal component/multiway principal component analysis
AnMBR–SBR	Anaerobic membrane bioreactor–sequencing batch reactor
AFWN	Agri-food waste network

References

1. Mokhtarani, B.; Zanganeh, J.; Moghtaderi, B. A Review on Biohydrogen Production Through Dark Fermentation, Process Parameters and Simulation. *Energies* **2025**, *18*, 1092. [\[CrossRef\]](#)
2. Zhao, L.; Sun, Z.-F.; Zhang, C.-C.; Nan, J.; Ren, N.-Q.; Lee, D.-J.; Chen, C. Advances in Pretreatment of Lignocellulosic Biomass for Bioenergy Production: Challenges and Perspectives. *Bioresour. Technol.* **2022**, *343*, 126123. [\[CrossRef\]](#) [\[PubMed\]](#)
3. 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.* **2022**, *343*, 126111. [\[CrossRef\]](#) [\[PubMed\]](#)
4. Zhao, S.; Li, J.; Chen, C.; Yan, B.; Tao, J.; Chen, G. Interpretable Machine Learning for Predicting and Evaluating Hydrogen Production via Supercritical Water Gasification of Biomass. *J. Clean. Prod.* **2021**, *316*, 128244. [\[CrossRef\]](#)
5. Anand, A.; Mahata, C.; Moholkar, V.S. Biohydrogen Synthesis from Food Waste Hydrolysate: Optimization Using Statistical Design of Experiments (DoE) and Artificial Neural Network (ANN). *Biomass Bioenergy* **2024**, *191*, 107452. [\[CrossRef\]](#)
6. Mougari, N.E.; Gherzi, D.E.; Iachachene, F.; Largeau, J.F.; Arici, M. A Generalized Hybrid Machine Learning Framework for Predicting Biohydrogen Production via Dark Fermentation of Organic Wastes. *Bioprocess Biosyst. Eng.* **2026**, *49*, 351–372. [\[CrossRef\]](#) [\[PubMed\]](#)
7. Mazzoli, R.; Pescarolo, S.; Gilli, G.; Gilardi, G.; Valetti, F. Hydrogen Production Pathways in Clostridia and Their Improvement by Metabolic Engineering. *Biotechnol. Adv.* **2024**, *73*, 108379. [\[CrossRef\]](#) [\[PubMed\]](#)
8. Shao, W.; Wang, Q.; Rupani, P.F.; Krishnan, S.; Ahmad, F.; Rezaia, S.; Rashid, M.A.; Sha, C.; Md Din, M.F. Biohydrogen Production via Thermophilic Fermentation: A Prospective Application of Thermotoga Species. *Energy* **2020**, *197*, 117199. [\[CrossRef\]](#)
9. Cabrol, L.; Marone, A.; Tapia-Venegas, E.; Steyer, J.-P.; Ruiz-Filippi, G.; Trably, E. Microbial Ecology of Fermentative Hydrogen Producing Bioprocesses: Useful Insights for Driving the Ecosystem Function. *FEMS Microbiol. Rev.* **2017**, *41*, 158–181. [\[CrossRef\]](#) [\[PubMed\]](#)
10. Adnan, A.I.; Mohamed, H.; Roslan, E.; Mohamad Aziz, N.A.; Firah, M.; Salehmin, M.N.I.; Zainal, B.S.; Zulkornain, M.F.; Nurhikmah, F.; Nomanbhay, S.; et al. A Review on Biohydrogen Production through Dark Fermentation via the Lactate Pathway. *Bioresour. Technol.* **2026**, *454*, 134726. [\[CrossRef\]](#) [\[PubMed\]](#)
11. Qu, X.; Zeng, H.; Gao, Y.; Mo, T.; Li, Y. Bio-Hydrogen Production by Dark Anaerobic Fermentation of Organic Wastewater. *Front. Chem.* **2022**, *10*, 978907. [\[CrossRef\]](#) [\[PubMed\]](#)
12. Nguyen, T.-A.D.; Kim, K.-R.; Kim, M.S.; Sim, S.J. Thermophilic Hydrogen Fermentation from Korean Rice Straw by Thermotoga Neapolitana. *Int. J. Hydrogen Energy* **2010**, *35*, 13392–13398. [\[CrossRef\]](#)
13. Cao, G.-L.; Zhao, L.; Wang, A.-J.; Wang, Z.-Y.; Ren, N.-Q. Single-Step Bioconversion of Lignocellulose to Hydrogen Using Novel Moderately Thermophilic Bacteria. *Biotechnol. Biofuels* **2014**, *7*, 82. [\[CrossRef\]](#) [\[PubMed\]](#)
14. Chandra, S.; Pandit, S.; Roy, A.; Rab, S.O.; Roy, A.K.; Saeed, M.; Kumar, A.; Sharma, K.; Ranjan, N.; Raj, S. Bioenergy Recovery from Jackfruit Waste via Biohydrogen Production through Dark Fermentation and Power Generation through Stacked Microbial Fuel Cells. *Int. J. Hydrogen Energy* **2025**, *102*, 845–855. [\[CrossRef\]](#)
15. Ramu, S.M.; Dinesh, G.H.; Thulasinathan, B.; Thondi Rajan, A.S.; Ponnuchamy, K.; Pugazhendhi, A.; Alagarsamy, A. Dark Fermentative Biohydrogen Production from Rice Mill Wastewater. *Int. J. Energy Res.* **2021**, *45*, 17233–17243. [\[CrossRef\]](#)
16. Singh, L.; Wahid, Z.A.; Siddiqui, M.F.; Ahmad, A.; Rahim, M.H.A.; Sakinah, M. Biohydrogen Production from Palm Oil Mill Effluent Using Immobilized Clostridium Butyricum EB6 in Polyethylene Glycol. *Process Biochem.* **2013**, *48*, 294–298. [\[CrossRef\]](#)
17. Zainal, B.S.; Akhbari, A.; Zinatizadeh, A.A.; Mohammadi, P.; Danaee, M.; Mohd, N.S.; Ibrahim, S. UASFF Start-up for Biohydrogen and Biomethane Production from Treatment of Palm Oil Mill Effluent. *Int. J. Hydrogen Energy* **2019**, *44*, 20725–20737. [\[CrossRef\]](#)
18. Mahmood, S.S.; Azahar, A.M.; Tan, J.P.; Jahim, J.M.; Abdul, P.M.; Mastar, M.S.; Anuar, N.; Mohammed Yunus, M.F.; Asis, A.J.; Wu, S.-Y. Operation Performance of Up-Flow Anaerobic Sludge Blanket (UASB) Bioreactor for Biohydrogen Production by Self-Granulated Sludge Using Pre-Treated Palm Oil Mill Effluent (POME) as Carbon Source. *Renew. Energy* **2019**, *134*, 1262–1272. [\[CrossRef\]](#)
19. Amorim, N.C.S.; Alves, I.; Martins, J.S.; Amorim, E.L.C. Biohydrogen Production from Cassava Wastewater in an Anaerobic Fluidized Bed Reactor. *Braz. J. Chem. Eng.* **2014**, *31*, 603–612. [\[CrossRef\]](#)
20. Xing, Y.; Li, Z.; Fan, Y.; Hou, H. Biohydrogen Production from Dairy Manures with Acidification Pretreatment by Anaerobic Fermentation. *Environ. Sci. Pollut. Res.* **2010**, *17*, 392–399. [\[CrossRef\]](#) [\[PubMed\]](#)
21. Sarma, S.; Ortega, D.; Minton, N.P.; Dubey, V.K.; Moholkar, V.S. Homologous Overexpression of Hydrogenase and Glycerol Dehydrogenase in Clostridium Pasteurianum to Enhance Hydrogen Production from Crude Glycerol. *Bioresour. Technol.* **2019**, *284*, 168–177. [\[CrossRef\]](#) [\[PubMed\]](#)

22. Zhang, Q.; You, S.; Li, Y.; Qu, X.; Jiang, H. Enhanced Biohydrogen Production from Cotton Stalk Hydrolysate of *Enterobacter Cloacae* WL1318 by Overexpression of the Formate Hydrogen Lyase Activator Gene. *Biotechnol. Biofuels* **2020**, *13*, 94. [[CrossRef](#)] [[PubMed](#)]
23. Balderas-Hernandez, V.E.; Landeros Maldonado, K.P.; Sánchez, A.; Smoliński, A.; De Leon Rodriguez, A. Improvement of Hydrogen Production by Metabolic Engineering of *Escherichia Coli*: Modification on Both the PTS System and Central Carbon Metabolism. *Int. J. Hydrogen Energy* **2020**, *45*, 5687–5696. [[CrossRef](#)]
24. Gupta, S.; Fernandes, A.; Lopes, A.; Grasa, L.; Salafranca, J. Photo-Fermentative Bacteria Used for Hydrogen Production. *Appl. Sci.* **2024**, *14*, 1191. [[CrossRef](#)]
25. Raj, S.; Kanthakere, S.; Ujwal, P. Recent Trends and Technological Advances in Sustainable Microbial Biohydrogen Production: A Review. *Syst. Microbiol. Biomanuf.* **2026**, *6*, 110. [[CrossRef](#)]
26. Redding, K.E.; Appel, J.; Boehm, M.; Schuhmann, W.; Nowaczyk, M.M.; Yacoby, I.; Gutekunst, K. Advances and Challenges in Photosynthetic Hydrogen Production. *Trends Biotechnol.* **2022**, *40*, 1313–1325. [[CrossRef](#)] [[PubMed](#)]
27. Lacroux, J.; Llamas, M.; Dauphin, K.; Avila, R.; Steyer, J.-P.; van Lis, R.; Trably, E. Dark Fermentation and Microalgae Cultivation Coupled Systems: Outlook and Challenges. *Sci. Total Environ.* **2023**, *865*, 161136. [[CrossRef](#)] [[PubMed](#)]
28. Sagir, E.; Ozgur, E.; Gunduz, U.; Eroglu, I.; Yucel, M. Single-Stage Photofermentative Biohydrogen Production from Sugar Beet Molasses by Different Purple Non-Sulfur Bacteria. *Bioprocess Biosyst. Eng.* **2017**, *40*, 1589–1601. [[CrossRef](#)] [[PubMed](#)]
29. Policastro, G.; Luongo, V.; Fabbicino, M. Biohydrogen and Poly- β -Hydroxybutyrate Production by Winery Wastewater Photofermentation: Effect of Substrate Concentration and Nitrogen Source. *J. Environ. Manag.* **2020**, *271*, 111006. [[CrossRef](#)] [[PubMed](#)]
30. Rao, R.; Basak, N. Sequential Dark-Photo Batch Fermentation and Kinetic Modelling for Biohydrogen Production Using Cheese Whey as a Feedstock. *Appl. Biochem. Biotechnol.* **2022**, *194*, 3930–3960. [[CrossRef](#)] [[PubMed](#)]
31. Zhang, Q.; Jiao, Y.; He, C.; Ruan, R.; Hu, J.; Ren, J.; Toniolo, S.; Jiang, D.; Lu, C.; Li, Y.; et al. Biological Fermentation Pilot-Scale Systems and Evaluation for Commercial Viability towards Sustainable Biohydrogen Production. *Nat. Commun.* **2024**, *15*, 4539. [[CrossRef](#)] [[PubMed](#)]
32. Bernabò, L.; Daly, G.; Galli, V.; Guerrini, S.; Viti, C.; Granchi, L.; Adessi, A. Selected-Wavelength Illumination for Enhanced Hydrogen and Poly- β -Hydroxybutyrate Production from Second Cheese Whey by *Rhodospseudomonas Palustris*. *Microorganisms* **2026**, *14*, 32. [[CrossRef](#)] [[PubMed](#)]
33. Fu, Y.; Song, Y.; Chen, H.; Lei, A.; Zhang, Y.; Liu, Z.; Wei, X. Integrated Dark Fermentation and Bioelectrochemical Systems for Waste Valorization: A Review of Microbial Fuel Cell and Microbial Electrolysis Cell Coupling Approaches. *Renew. Sustain. Energy Rev.* **2026**, *237*, 117021. [[CrossRef](#)]
34. Sharma, A.; Hussain Mehdi, S.E.; Pandit, S.; Eun-Oh, S.; Natarajan, V. Factors Affecting Hydrogen Production in Microbial Electrolysis Cell (MEC): A Review. *Int. J. Hydrogen Energy* **2024**, *61*, 1473–1484. [[CrossRef](#)]
35. Khongkhiang, P.; Kongjan, P.; Utarapichat, B.; Reungsang, A.; O-Thong, S. Continuous Hydrogen Production from Cassava Starch Processing Wastewater by Two-Stage Thermophilic Dark Fermentation and Microbial Electrolysis. *Int. J. Hydrogen Energy* **2017**, *42*, 27584–27592. [[CrossRef](#)]
36. Khongkhiang, P.; Jehlee, A.; Kongjan, P.; Reungsang, A.; O-Thong, S. High Efficient Biohydrogen Production from Palm Oil Mill Effluent by Two-Stage Dark Fermentation and Microbial Electrolysis under Thermophilic Condition. *Int. J. Hydrogen Energy* **2019**, *44*, 31841–31852. [[CrossRef](#)]
37. Marone, A.; Ayala-Campos, O.R.; Trably, E.; Carmona-Martínez, A.A.; Moscoviz, R.; Latrille, E.; Steyer, J.-P.; Alcaraz-Gonzalez, V.; Bernet, N. Coupling Dark Fermentation and Microbial Electrolysis to Enhance Bio-Hydrogen Production from Agro-Industrial Wastewaters and by-Products in a Bio-Refinery Framework. *Int. J. Hydrogen Energy* **2017**, *42*, 1609–1621. [[CrossRef](#)]
38. Huang, J.; Feng, H.; Huang, L.; Ying, X.; Shen, D.; Chen, T.; Shen, X.; Zhou, Y.; Xu, Y. Continuous Hydrogen Production from Food Waste by Anaerobic Digestion (AD) Coupled Single-Chamber Microbial Electrolysis Cell (MEC) under Negative Pressure. *Waste Manag.* **2020**, *103*, 61–66. [[CrossRef](#)] [[PubMed](#)]
39. Karlsson, A. Effects of Temperature, Hydraulic Retention Time and Hydrogen Extraction Rate on Hydrogen Production from the Fermentation of Food Industry Residues and Manure. *Int. J. Hydrogen Energy* **2008**, *33*, 953–962. [[CrossRef](#)]
40. Azbar, N.; Dokgöz, F.T.; Keskin, T.; Eltem, R.; Korkmaz, K.S.; Gezgün, Y.; Akbal, Z.; Öncel, S.; Dalay, M.C.; Gönen, Ç.; et al. Comparative Evaluation of Bio-Hydrogen Production From Cheese Whey Wastewater Under Thermophilic and Mesophilic Anaerobic Conditions. *Int. J. Green Energy* **2009**, *6*, 192–200. [[CrossRef](#)]
41. Volpi, M.P.C.; Adarme, O.F.H.; Araújo, M.F.; Bella, T.R.; Procópio, P.F.; de Abreu, L.G.F.; Carazzolle, M.F.; Pereira, G.A.G.; Mockaitis, G. Biohydrogen and Methane Production via Silage-Based Dark Co-Fermentation Using Vinasse and Filter Cake. *Bioresour. Technol. Rep.* **2024**, *27*, 101927. [[CrossRef](#)]
42. García-Depraet, O.; León-Becerril, E. Use of a Highly Specialized Biocatalyst to Produce Lactate or Biohydrogen and Butyrate from Agro-Industrial Resources in a Dual-Phase Dark Fermentation. *Fermentation* **2023**, *9*, 787. [[CrossRef](#)]

43. Villanueva-Galindo, E.; Pérez-Rangel, M.; Moreno-Andrade, I. Biohydrogen Production from Lactic Acid: Use of Food Waste as Substrate and Evaluation of Pretreated Sludge and Native Microbial Community as Inoculum. *Int. J. Hydrogen Energy* **2025**, *108*, 2–11. [\[CrossRef\]](#)
44. Aranda-Jaramillo, B.; García-Depraect, O.; Aguilar-Juárez, O.; Archundia-Amador, M.; León-Becerril, E. Continuous Hydrogen Production from Cheese Whey in a Single-Stage Lactate-Driven Dark Fermentation Reactor. *J. Taiwan Inst. Chem. Eng.* **2025**, *177*, 106307. [\[CrossRef\]](#)
45. García-Depraect, O.; Mena-Navarro, V.; Muñoz, R.; Rene, E.R.; León-Becerril, E. Effect of Nitrogen and Iron Supplementation on the Process Performance and Microbial Community Structure of a Hydrogen-Producing Reactor Continuously Fed with Tequila Vinasse. *Fuel* **2023**, *334*, 126736. [\[CrossRef\]](#)
46. Cao, Y.; Liu, H.; Liu, W.; Guo, J.; Xian, M. Debottlenecking the Biological Hydrogen Production Pathway of Dark Fermentation: Insight into the Impact of Strain Improvement. *Microb. Cell Fact.* **2022**, *21*, 166. [\[CrossRef\]](#) [\[PubMed\]](#)
47. Sikarwar, V.S.; Zhao, M.; Fennell, P.S.; Shah, N.; Anthony, E.J. Progress in Biofuel Production from Gasification. *Prog. Energy Combust. Sci.* **2017**, *61*, 189–248. [\[CrossRef\]](#)
48. Shi, T.; Zhou, J.; Ayub, Y.; Toniolo, S.; Ren, J. Novel Process Optimization Based on Machine Learning: A Study on Biohydrogen Production from Waste Resources. *Biomass Bioenergy* **2024**, *185*, 107222. [\[CrossRef\]](#)
49. Kan, T.; Strezov, V.; Evans, T.J. Lignocellulosic Biomass Pyrolysis: A Review of Product Properties and Effects of Pyrolysis Parameters. *Renew. Sustain. Energy Rev.* **2016**, *57*, 1126–1140. [\[CrossRef\]](#)
50. Jahangiri, H.; Santos, J.; Hornung, A.; Ouadi, M. Thermochemical Conversion of Biomass and Upgrading of Bio-Products to Produce Fuels and Chemicals. In *Catalysis for Clean Energy and Environmental Sustainability: Biomass Conversion and Green Chemistry*; Pant, K.K., Gupta, S.K., Ahmad, E., Eds.; Springer International Publishing: Cham, Switzerland, 2021; Volume 1, pp. 1–47.
51. Li, Y.; Gupta, R.; Zhang, Q.; You, S. Review of Biochar Production via Crop Residue Pyrolysis: Development and Perspectives. *Bioresour. Technol.* **2023**, *369*, 128423. [\[CrossRef\]](#) [\[PubMed\]](#)
52. Homagain, K.; Shahi, C.; Luckai, N.; Sharma, M. Biochar-Based Bioenergy and Its Environmental Impact in Northwestern Ontario Canada: A Review. *J. For. Res.* **2014**, *25*, 737–748. [\[CrossRef\]](#)
53. Kazawadi, D.; Ntalikwa, J.; Kombe, G. A Review of Intermediate Pyrolysis as a Technology of Biomass Conversion for Coproduction of Biooil and Adsorption Biochar. *J. Renew. Energy* **2021**, *2021*, 5533780. [\[CrossRef\]](#)
54. Elsis, S.F.; Omar, M.N.; Azam, M.M.; Eissa, A.H.A.; Gomaa, E.M. Effect of Pyrolysis Process on the Properties of Briquettes Produced from Different Particle Size Peanut Shells and Grape Pruning Residues. *Biomass Bioenergy* **2025**, *193*, 107532. [\[CrossRef\]](#)
55. Ouadi, M.; Fivga, A.; Jahangiri, H.; Saghir, M.; Hornung, A. A Review of the Valorization of Paper Industry Wastes by Thermochemical Conversion. *Ind. Eng. Chem. Res.* **2019**, *58*, 15914–15929. [\[CrossRef\]](#)
56. Santos, J.; Ouadi, M.; Jahangiri, H.; Hornung, A. Valorisation of Lignocellulosic Biomass Investigating Different Pyrolysis Temperatures. *J. Energy Inst.* **2020**, *93*, 1960–1969. [\[CrossRef\]](#)
57. Santos, J.; Jahangiri, H.; Bashir, M.A.; Hornung, A.; Ouadi, M. The Upgrading of Bio-Oil from the Intermediate Pyrolysis of Waste Biomass Using Steel Slag as a Catalyst. *ACS Sustain. Chem. Eng.* **2020**, *8*, 18420–18432. [\[CrossRef\]](#)
58. Bridgwater, A.V. Review of Fast Pyrolysis of Biomass and Product Upgrading. *Biomass Bioenergy* **2012**, *38*, 68–94. [\[CrossRef\]](#)
59. Hornung, A.; Jahangiri, H.; Ouadi, M.; Kick, C.; Deinert, L.; Meyer, B.; Grunwald, J.; Daschner, R.; Apfelbacher, A.; Meiller, M.; et al. Thermo-Catalytic Reforming (TCR)—An Important Link between Waste Management and Renewable Fuels as Part of the Energy Transition. *Appl. Energy Combust. Sci.* **2022**, *12*, 100088. [\[CrossRef\]](#)
60. Ungureanu, N.; Vlăduț, N.-V.; Biriș, S.-Ștefan; Gheorghită, N.-E.; Ionescu, M. Biomass Pyrolysis Pathways for Renewable Energy and Sustainable Resource Recovery: A Critical Review of Processes, Parameters, and Product Valorization. *Sustainability* **2025**, *17*, 7806. [\[CrossRef\]](#)
61. Ouadi, M.; Bashir, M.A.; Speranza, L.G.; Jahangiri, H.; Hornung, A. Food and Market Waste—A Pathway to Sustainable Fuels and Waste Valorization. *Energy Fuels* **2019**, *33*, 9843–9850. [\[CrossRef\]](#) [\[PubMed\]](#)
62. Di Gruttola, F.; Jahangiri, H.; Sajdak, M.; Majewski, A.; Borello, D.; Hornung, A.; Ouadi, M. Thermo-Catalytic Reforming (TCR) of Waste Solid Grade Laminate. *J. Clean. Prod.* **2023**, *419*, 138276. [\[CrossRef\]](#)
63. Ouadi, M.; Brammer, J.G.; Yang, Y.; Hornung, A.; Kay, M. The Intermediate Pyrolysis of De-Inking Sludge to Produce a Sustainable Liquid Fuel. *J. Anal. Appl. Pyrolysis* **2013**, *102*, 24–32. [\[CrossRef\]](#)
64. Kayan, R.R.; Jauhar, S.K.; Kamble, S.S.; Belhadi, A. Optimizing Bio-Hydrogen Production from Agri-Waste: A Digital Twin Approach for Sustainable Supply Chain Management and Carbon Neutrality. *Comput. Ind. Eng.* **2025**, *204*, 111021. [\[CrossRef\]](#)
65. Mahata, C.; Ray, S.; Das, D. Optimization of Dark Fermentative Hydrogen Production from Organic Wastes Using Acidogenic Mixed Consortia. *Energy Convers. Manag.* **2020**, *219*, 113047. [\[CrossRef\]](#)
66. Zhang, X.; Wang, Y.; Hu, J.; Zhang, Q.; Xuan, X.; Shi, L.; Sun, Y. A Molecular Descriptor-Based Correlation with the Composition of Acid-Pretreated Cornstalk Cultivation Medium for Biohydrogen Production Using a Machine Learning Approach. *Int. J. Hydrogen Energy* **2025**, *123*, 307–320. [\[CrossRef\]](#)

67. Moreno Cárdenas, E.L.; Zapata-Zapata, A.D.; Kim, D. Modeling Dark Fermentation of Coffee Mucilage Wastes for Hydrogen Production: Artificial Neural Network Model vs. Fuzzy Logic Model. *Energies* **2020**, *13*, 1663. [\[CrossRef\]](#)
68. Dang, V.T.; Chen, W.-H.; Hoang, A.T.; Guerrero-Pérez, M.O.; Rodríguez-Castellón, E.; López-Escalante, M.C.; Nguyen, M.C.; Bui, T.A.E.; Truong, T.H.; Nguyen, X.P. An Integrated Framework for Bio-Hydrogen Production Optimization Using Novel Metaheuristic Algorithms and Explainable Machine Learning Tuned via Grid Search. *Int. J. Hydrogen Energy* **2025**, *177*, 151555. [\[CrossRef\]](#)
69. Shomope, I.; Tawalbeh, M.; Al-Othman, A.; Almomani, F. Predicting Biohydrogen Production from Dark Fermentation of Organic Waste Biomass Using Multilayer Perceptron Artificial Neural Network (MLP–ANN). *Comput. Chem. Eng.* **2025**, *192*, 108900. [\[CrossRef\]](#)
70. Liu, C.; Li, F.; Zhang, P.; Balasubramanian, P. Augmented Machine Learning with Limited Data for Hydrogen Yield Prediction in Wastewater Dark Fermentation. *npj Clean Water* **2025**, *8*, 101. [\[CrossRef\]](#)
71. Nasr, N.; Hafez, H.; El Naggat, M.H.; Nakhla, G. Application of Artificial Neural Networks for Modeling of Biohydrogen Production. *Int. J. Hydrogen Energy* **2013**, *38*, 3189–3195. [\[CrossRef\]](#)
72. Monroy, I.; Guevara-López, E.; Buitrón, G. A Mechanistic Model Supported by Data-Based Classification Models for Batch Hydrogen Production with an Immobilized Photo-Bacteria Consortium. *Int. J. Hydrogen Energy* **2016**, *41*, 22802–22811. [\[CrossRef\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.