**INTERNATIONAL JOURNAL OF ADVANCED RESEARCH**
*International Peer-Reviewed Research Journal*
**ISSN (Online): 3106-7980 | ISSN (Print): 3106-7972**
**Volume 3 • Issue 5 • September 2026**

# AI-DRIVEN OPTIMIZATION OF AGRICULTURAL RESIDUE BIOREFINERIES FOR SIMULTANEOUS PRODUCTION OF BIOFUELS, PLATFORM CHEMICALS, AND VALUE-ADDED BIOMATERIALS

**Zakir Ullah[*1], Afsheen Zakir[2]**

[*1]*Department of Agriculture, University of Agriculture Peshawar*
[2]*Department of Agriculture Extension, Women University Mardan*

[*1]zakirullah499@gmail.com, [2]afsheenzakir972@gmail.com

**DOI:** https://doi.org/10.5281/zenodo.23183257

**Keywords**
*Artificial intelligence; Machine learning; Lignocellulosic biorefinery; Agricultural residues; Biofuels; Platform chemicals; Biomaterials; Digital twins; Multi-objective optimization; Circular bioeconomy..*

**Article History**
Received: 01 June 2026
Accepted: 23 August 2026
Published: 30 September 2026

**Copyright** @Author
**Corresponding Author: * Zakirullah**

**Abstract**

*A shift to circular bioeconomy requires the sustainable utilization of agricultural residues, which are available in the annual global scale of about 998 MT. Lignocellulosic biorefineries continue to struggle with the issues of feedstock heterogeneity and recalcitrance, energy intensive pretreatments and the economic viability of single-product systems. This review focuses on the disruptive impact of artificial intelligence (AI) and machine learning (ML) in enhancing the performance of integrated biorefineries for the co-production of biofuels, platform chemicals, and value-added biomaterials. Complex models, such as artificial neural networks (ANNs), eXtreme Gradient Boosting (XGB), Random Forest (RF), Support Vector Regression (SVR), and bio-inspired metaheuristics, were developed for predictive modeling, real-time process control, and multi-objective optimization of the biorefinery process from feedstock pretreatment to biopolymer product recovery. AI-guided pretreatment processes suppress feedstock variability as well as inhibitor production, while ML algorithms maximize enzymatic saccharification, fermentation kinetics, and thermochemical conversion route. In platform chemicals synthesis, hybrid machine learning (ML) schemes integrated with thermodynamic screening enables high-throughput catalyst screening and biphasic system parameter optimization for 5-HMF, furfural and levulinic acid synthesis. For biomaterials, transfer learning and neural network models are used to improve processes such as extraction of nanocellulose, macromolecular tailoring of lignin, and fermentation of polyhydroxyalkanoates. Digital twins, techno-economic analysis, and life cycle assessment, at the systems level, provide the mechanisms for dynamic flow routing based on market values and carbon efficiency objectives. Data heterogeneity, small dataset issues, and model interpretability are also challenged by a number of methods such as transfer learning, physics-informed neural networks (PINNs), and explainable AI (XAI) frameworks. The review ends with stating that AI-based optimization is a game changing*

*technology that empowers autonomous, economically feasible, and environmentally sustainable biorefineries.*

## 1. INTRODUCTION

### 1.1 Context and Global Imperative for Lignocellulosic Biorefineries

The move away from an economy that depends on fuels to a sustainable circular bioeconomy will be one of the most important technological changes, in the twenty-first century. Agricultural waste is one of the growing types of biomasses. It is created by farming over the world at a rate of 998 million tonnes each year (Jayabal & Prabhakar 2026). Wheat straw, which is 400 to 529 million tonnes, bagasse, which's 279 to 300 million tonnes, corncob waste, which is 200 to 230 million tonnes and rice husks which is 100 to 120 million tonnes are some of the biggest parts of this waste. These food-waste lignocellulosic residues are a carbon-plentiful resource. They can replace petroleum-based fuels, industrial solvents and synthetic materials (Coronado-Contreras et al. 2025).

Even though there is much potential large-scale processing of lignocellulosic materials has always had problems. The tough structure of plant cell walls makes it difficult to use these materials. Lignocellulose is a complex, heterogeneous structure comprising cellulose microfibrils embedded in a cross linked network of hemicellulose and lignins. Energetic, thermal or chemical pretreatments are needed to break down this structural barrier (John et al., 2025). Yet all these stages often lead to the generation of microbial and enzymatic inhibitors, are heavily energy-intensive, and introduce process bottlenecks which enfeeble the economic feasibility. Recently, more advanced machine learning models that dynamically relate feedstock variations to ideal processing severities are emerging with the goal to reduce energy usage and the production of inhibitors (Pusuluru et al., 2026).

### 1.2 The Tripartite Co-Production Paradigm: Biofuels, Chemicals, and Biomaterials

To achieve commercial solvency, modern biorefineries are pivoting away from single-output models such as standalone bioethanol production toward integrated, multi-product cascade systems (Parthipan et al., 2026). The tripartite co-production strategy systematically fractionates agricultural residues into three distinct value streams:

- **Biofuels**: High-volume, moderate-margin liquid and gaseous energy carriers, including fuel-grade bioethanol, biomethane via anaerobic digestion, and syngas via gasification.
- **Platform Chemicals**: Medium-volume, high-value C5 and C6 platform molecules such as 5-hydroxymethylfurfural (5-HMF), furfural, levulinic acid, and succinic acid which serve as precursors for green solvents, biplastics, and pharmaceuticals (Manhongo, 2023).
- **Value-Added Biomaterials**: Precision-engineered solid materials, including cellulose nanocrystals (CNCs), functionalized structural lignins, and polyhydroxyalkanoates (PHAs) accumulated via microbial fermentation.

Maximizing carbon efficiency in these three streams means keeping a close watch on reactor conditions, chemical balances and the flow of each stream. In this model that works together as a whole machine learning algorithms adjust how the streams are directed. They do this by using real-time predictions about how fast reactions happen and, by looking at what the market values (Liu et al. 2025).

**INTERNATIONAL JOURNAL OF ADVANCED RESEARCH**
*International Peer-Reviewed Research Journal*
**ISSN (Online): 3106-7980 | ISSN (Print): 3106-7972**
**Volume 3 • Issue 5 • September 2026**

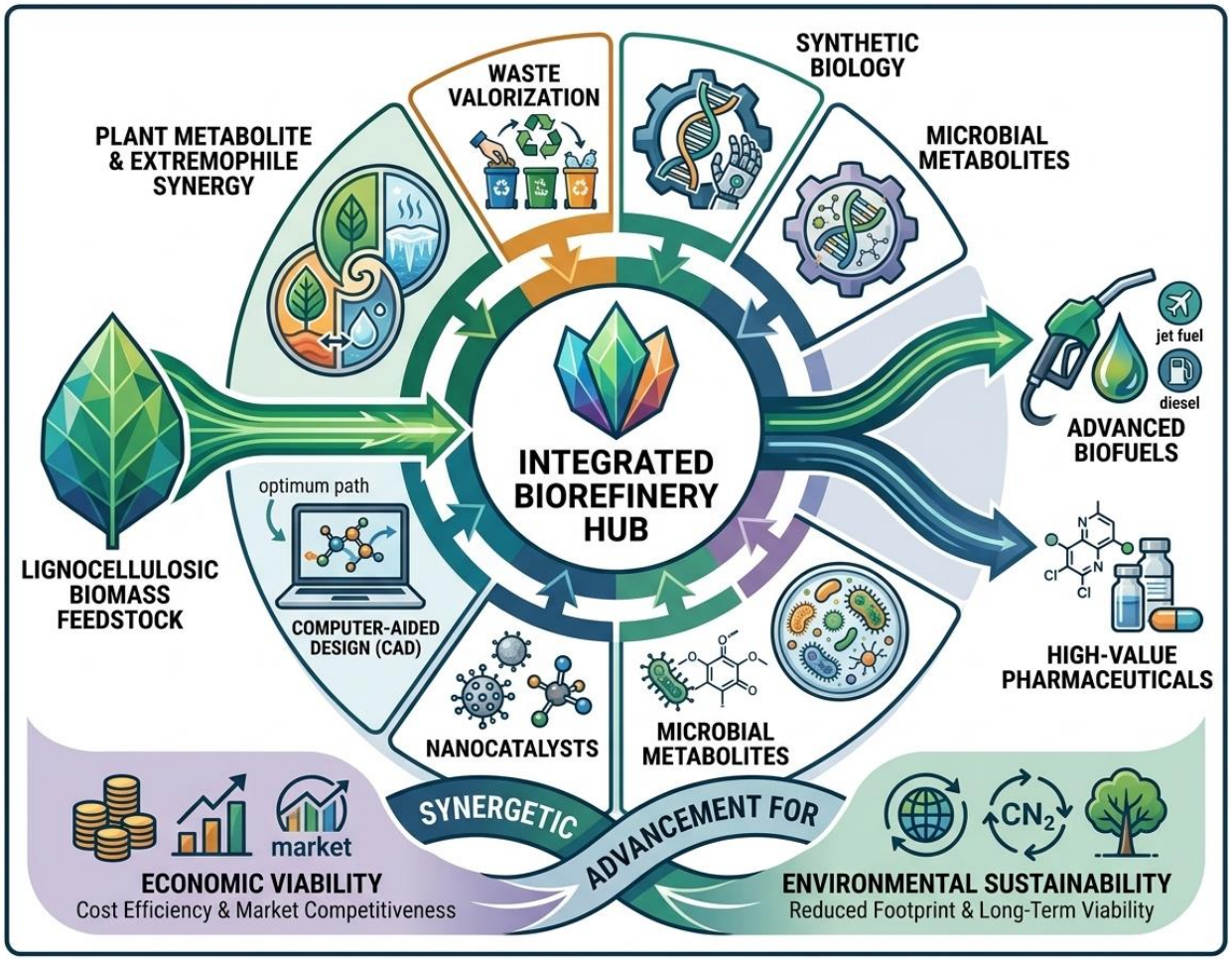



**Figure 1: A diagram of an integrated biorefinery hub. It shows where input feedstock streams enter how AI helps design the process and how the hub makes biofuels, platform chemicals and value-added biomaterials.**

### 1.3 Role of Artificial Intelligence and Machine Learning in Addressing Process Complexity

The multivariate nature of integrated biorefineries is problematic for conventional empirical or univariate process optimization. Feedstock compositions continuously vary according to type of crop, soil conditions, harvesting method, seasonal storage. In addition, reaction rates (reaction kinetics) during pretreatment, enzymatic saccharification and catalytic conversion process are also highly non-linear (Xie et al., 2026).

Artificial intelligence (AI) and machine learning (ML) methodologies offer a structure to tackle

such issues. Using -dimensional data machine learning algorithms such as artificial neural networks, XGBoost support vector regression and bio-inspired metaheuristics can uncover hidden complex relationships between inputs and outputs. These AI-based tools enable predictive modeling, real-time process control and multi-objective optimization across all stages of the biorefinery process chain. This digital

transformation bridges the gap, from laboratory bench scale to scale (Lilaria et al., 2026).

## 2. Agricultural Residue Feedstock Dynamics and Pretreatment Optimization

### 2.1 Compositional Heterogeneity and Substrate Recalcitrance

There are differences in the main biopolymer parts of agricultural waste materials. For example sugarcane bagasse has 45% to 50% cellulose, 25% to 30% hemicellulose, 17% to 25% lignin 2.4% to 9% ash. Corn stover and cereal straws have amounts of cellulose compared to lignin and different types of inorganic ash (Pilarski & Pilarska 2025). The way these materials resist breakdown is because of how cellulose is organized how hemicellulose protects the structure how lignin connects to carbohydrates and the parts that make up the lignin (like guaiacyl, syringyl and p-hydroxyphenyl units) (Chen et al., 2026).

Changes in the makeup of these materials affect how strong the reactions are, during processing because they change how strong the reactions need to be. When the same processing settings are used on types of starting materials more energy is used less sugar is produced and more harmful substances are created (Ayele et al. 2026).

### 2.2 Machine Learning for Predictive Pretreatment Severity and Fractionation

Machine learning architectures are utilized to compensate for feedstock heterogeneity by modeling pretreatment dynamics. Rather than using fixed empirical severity parameters, ML models treat as inputs vectors of raw feedstock chemistry (cellulose, hemicellulose, lignin, ash content), process conditions (temperature, residence time, liquid-to-solid ratio, chemical catalyst concentration), and heating modes (Xu et al., 2026).

Recent results suggest that composition-based modeling approaches, which represent the post pretreatment biomass as a “state” in the form of an intermediate feature vector, are capable of providing high prediction accuracy for multiple conversion routes (Zhong et al.,2025). Metaheuristic optimization algorithms-based ensemble methods adjusted with bio-inspired optimization algorithms such as the Grey Wolf Optimizer (GWO), Artificial Bee Colony (ABC), and Differential Evolution (DE) are superior than traditional linear regression. A GWO-optimized GB model with a RMSE of 4.38, mean absolute error (MAE) of ≈2.19 and coefficient of determination (R 2) of ≈0.99 for predicting biomass conversion product under varying pretreatment severities Xie et al., (2026).

### 2.3 Mechanistic and Energy Trade-Offs Across Pretreatment Modalities

The choice of suitable pretreatment technology is of the trade-off between sugar release and energy input, and generation of inhibitors. Hydrothermal pretreatment (HTP) – liquid hot water, steam explosion, etc. – is known to dissolve hemicellulose, cleave LCC linkages (Podgorbunskikh et al., 2026). However, high-severity HTP leads to substantial amounts of inhibitors, such as furfural from pentose degradation, 5-HMF from hexose degradation, aliphatic organic acids, and insoluble pseudo-lignin polymerization materials. These extras are complex and expensive to purify (Mehta, 2026).

Microwave-assisted pretreatment (MWP) is based on fast dielectric bulk heating that shorten reaction residence times from hours to minutes. The MWP process requires less total thermal energy than for conductive HTP e.g. 40.1" kJ/g” for MWP vs 70.85" kJ/g” for conventional HTP. However, MWP requires quality electricity, which has different economic trade-offs than use of low-quality waste steam in a biorefinery’s combined heat and power (CHP) system (Di Domenico et al., 2025).

Mechanochemical ball milling (BM) breaks down biomass. Causes the cellulose to lose its crystalline structure without creating chemical degradation products. However, the high amount of power it needs (up to 2.8 kWh per kg) limits its use as a standalone process and encourages combining it

with chemical processing methods (Wu et al. 2026). An artificial neural network (ANN) model created for steam exploded alkaline pretreated softwood had an R squared value of 0.9805. This model was able to predict the results of enzymatic hydrolysis under different chemical concentrations (0.5%, to 2.0% NaOH) and various residence times (Gao et al. 2026).

**Table 1. Comparative Analysis of Pretreatment Modalities and Machine Learning Predictive Performance.**

| Pretreatment Modality | Primary Mechanistic Action | Typical Energy Demand | Inhibitor Generation Risk | Representative ML Architecture | Model Performance Metrics | Primary Input Features |
|---|---|---|---|---|---|---|
| **Hydrothermal (HTP / Steam Explosion)** | Hemicellulose solubilization, LCC disruption | 70.85 kJ/g | High (furfural, 5-HMF, pseudo-lignin) | Artificial Neural Network (ANN) | $R^2 = 0.9805$ | Steam time, NaOH %, reaction temperature |
| **Microwave-Assisted (MWP)** | Rapid dielectric volumetric heating | 40.10 kJ/g | Moderate (mitigated by short residence) | Random Forest (RF) | $R^2 \approx 0.911$ | Microwave power, exposure time, solvent ratio |
| **Ball Milling (BM)** | Mechanical shear, crystallinity destruction | Up to 2.80 kWh/kg | Negligible (no chemical degradation) | Ensemble Decision Trees | $R^2 > 0.90$ | Milling duration, media size, rotation speed |
| **Hybrid Chemical / Solvolysis (DES / IL)** | Selective delignification, ether bond cleavage | Variable (solvent recovery dependent) | Low to Moderate | GWO-Tuned Gradient Boosting | $R^2 \approx 0.99$, RMSE = 4.38 | Liquid-to-solid ratio, temperature, composition |

## 3. AI-Guided Enzymatic Saccharification and Biochemical Conversion

### 3.1 Enzyme-Substrate Mechanics and Non-Productive Lignin Binding

After pretreatment the structural polysaccharides cellulose and hemicellulose must be changed into sugars such as glucose, xylose and arabinose by breaking them down with cellulase and hemicellulase enzyme mixtures. A major problem is that cellulolytic enzymes attach uselessly to leftover lignin, which blocks saccharification. This issue remains a constraint, to enzymatic saccharification (Wyman et al., 2005). It is induced by hydrophobic forces, π"~" π stacking of tyrosine residues with lignin phenolic moieties in the enzymes catalytic or carbohydrate-binding domains (CBDs) and in the phenolic moieties of lignin, and can be weakened by hydrogen bonding and electrostatic interaction. Guaiacyl-rich (G-type) lignins have an excellent binding capacity for β-glucosidase (for example, from T. reesei) which inhibits the release of cellobiose from cellulose and also causes product inhibition (Saha, 2003). Non-productive binding can also be inhibited by addition of inexpensive sacrificial additives such as soy protein conjugates from defatted soy flour. These additives selectively associate with active lignin sites through hydrophobic and electrostatic

interactions, resulting in protective shielding of cellulose microfibrils and preservation of the catalytic activities of cellulase to degrade structural

**3.2 Machine Learning Models for Hydrolysis Yield Prediction**

Enzymatic hydrolysis is a multi-stage process governed by multi-factor interactions, including solid loading, enzyme loading (filter paper units per gram of glucan), substrate accessibility, temperature, pH, and incubation time, in which these factors influence each other in a non-linear way. The AI models inherently learn these non-linearities without needing to have the complicated numerical solutions to these complex differential rate equations (Cheng & Rosentrater, 2025).

Such algorithms as ANNs, RF regression, and SVR are good candidates for predicting fermentable sugar titers. The current bibliometric analyses indicate that ANNs are applied to about 90% of the documented studies on bioethanol pathway modeling whereas RSM (≈25%) and RF (≈10%) are some of the other models encountered (Yu et al., 2013). The RF models developed to predict glucose release from ionic-liquid-pretreated sugarcane bagasse exhibited R^2 values greater than 0.91. This type of model enables a scientist to find promising enzyme ratios and processing times before running lab experiments (Jayabal & Prabhakar, 2026).

**3.3 Optimization of Fermentation and Microbial Consortia Dynamics**

Once fermentable sugars are liberated, micro-organisms convert the hydrolysate into biofuels or biochemical precursors. Wild-type *Saccharomyces cerevisiae* ferments hexoses (glucose, fructose, mannose) via the Embden-Meyerhof-Parnas glycolytic pathway according to the stoichiometry:

$$C_6H_{12}O_6 \longrightarrow 2C_2H_5OH + 2CO_2$$

Wild-type yeast strains do not use pentose sugars, like xylose. Arabinose very well even though these sugars are a big part of hemicellulose hydrolysates (Posilipo, 2024). To make use of these sugars researchers have created industrial strains that have special pathways. These pathways use xylose reductase and xylitol dehydrogenase. They use polysaccharides to monomeric glucose (Kucharska et al., 2018).

xylose isomerase. These changes help yeast to break down both five-carbon sugars and six-carbon sugars (Yu et al., 2013).

Machine-learning tools can help manage sugar use in engineered yeast cultures and mixed microbial communities. These tools use measurements such as redox potential, dissolved oxygen, pH, temperature organic loading and nitrogen levels to adjust feeding schedules in fed-batch and continuous reactors. By doing this they help keep the system running. Predictive control can also limit the buildup of byproducts, like glycerol, acetic acid and lactic acid. This leads to ethanol yields and faster production rates (Falowo et al. 2026).

**4. Thermochemical and Catalytic Conversion to Biofuels and Platform Chemicals**

**4.1 Pyrolysis, Gasification, and Anaerobic Digestion Optimization**

In addition to biochemical processes, thermochemical methods allow whole, non-grounded agricultural residues to be quickly converted. Anaerobic digestion (AD) is a multi-stage biological process (hydrolysis, acidogenesis, acetogenesis, and methanogenesis) that converts heterogeneous organic matter to biogas ( 〖"CH" 〗_4 and 〖"CO" 〗_2) (Hidalgo et al., 2025). Since AD depends on a variety of bacterial and archaeal species acting together, the stability of the process is vulnerable to the build-up of VFAs and sudden changes in OLR (based on Acosta et al., 2016). ML models the operating parameters to estimate the amount of biomethane produced and to avoid the souring of digester (Hu & Chen, 2025).

Gasification is the thermal conversion of biomass in an environment with limited supply of oxygen at high temperature (650^∘ "C~" 900^∘ "C" ) to produce a fuel gas (syngas) mainly consisting of ("CO" , "H" _2, 〖"CO" 〗_2, 〖"CH" 〗_4). Co-gasification with alternate waste such as agricultural residues (e.g., sugarcane bagasse, corncob, wood chips) and secondary waste will

enhance gas quality by compensating the C:H ratios and mitigating bed reactivity fluctuations (Nagappan et al., 2025). Controlling reaction processes, e.g., the endothermic Boudouard process ("C"+〖"CO" 〗_2⇌2"CO" ) and the steam gasification process ("C"+"H" _2 "O"⇌"CO"+"H" _2) can be closely affected by the design of the reactor, the gasifying medium, temperature, and length of staying. ML prediction models analyze input perturbations for predicting the quality of syngas without performing costly trial runs (Danso-Boateng & Achaw, 2022).

**Table 2. Syngas and Biogas Composition Characteristics Across Selected Feedstocks and Process Temperatures.**

| Feedstock Type | Processing Temperature (∘C) | Feed C (wt%, dry) | Feed H (wt%, dry) | Feed O (wt%, dry) | Output CO2 (vol %) | Output H2 (vol %) | Output CH4 (vol %) | Output CO (vol %) |
|---|---|---|---|---|---|---|---|---|
| **Pure Cellulose** | 900 | 40.7 | 5.6 | 53.7 | 40.0 | 70.5 | 6.0 | 23.5 |
| **Corncob** | 750 | 46.9 | 5.9 | 43.6 | 14.2 | 38.9 | 7.1 | 38.2 |
| **Hazelnut Shells** | 848 | 50.5 | 5.9 | 42.1 | 12.3 | 15.5 | 2.5 | 20.0 |
| **Olive Pomace** | 800 | 48.1 | 5.2 | 44.6 | 26.0 | 16.2 | 2.0 | 5.5 |
| **Pecan Waste** | 650 | 17.2 | 2.1 | 14.2 | 65.8 | 81.6 | 1.6 | 0.1 |
| **Pine Wood** | 400 | 50.5 | 7.1 | 41.1 | | 20.0 | | 10.0 |
| **Sawdust** | 750 | 48.0 | 6.0 | 45.4 | 35.3 | 31.4 | 8.4 | 20.6 |
| **Sawdust** | 800 | 48.0 | 6.0 | 45.4 | 35.5 | 37.2 | 7.6 | 17.4 |
| **Wheat Straw** | 900 | 47.5 | 5.7 | 40.1 | 23.3 | 34.8 | 3.6 | 37.9 |
| **Sugarcane Bagasse** | 750 | 46.1 | 5.3 | 44.3 | 24.3 | 59.1 | 0.8 | 15.7 |
| **Wood Chips** | 800 | 43.6 | 5.2 | 5.1 | 9.2 | 12.2 | 1.2 | 15.9 |

**4.2 Catalytic Dehydration Pathways to Platform Chemicals (5-HMF, Furfural, Levulinic Acid)**

Acid-catalyzed dehydration converts pentose and hexose sugars into key C5 and C6 platform chemicals. C5 pentoses (xylose and arabinose) undergo acidic dehydration via a 1,2-enediol intermediate or a 2,5-anhydroxylose furanose intermediate to yield furfural. C6 hexoses (glucose and fructose) dehydrate via the loss of three water molecules to form 5-HMF (Árvai et al., 2024). Subsequent rehydration of 5-HMF in an aqueous acidic environment breaks open the furan ring, yielding levulinic acid (LA) and formic acid in equimolar amounts:

$$C_6H_{12}O_6 \xrightarrow[-3H_2O]{H^+} C_6H_6O_3 \text{ (5-HMF)} \xrightarrow[+2H_2O]{H^+} C_5H_8O_3 \text{ (Levulinic A}$$

$CH_2O_2$ (Formic Acid)

Heteropolyacid catalysts specifically Keggin-type structures such as phosphotungstic acid ($H_3PW_{12}O_{40}$) and silicotungstic acid ($H_4SiW_{12}O_{40}$) exhibit high Brønsted acidity and structural stability for these reactions. Biphasic solvent configurations combining water with organic co-solvents (such as methyl isobutyl ketone [MIBK] or $\gamma$-valerolactone [GVL]) continuously

extract 5-HMF or furfural into the organic phase, protecting them from secondary degradation reactions (ivec et l., 2019).

Machine learning models, coupled with thermodynamic screening frameworks like COSMO-RS, accelerate catalyst selection and biphasic system optimization. Hybrid architectures such as Gaussian Wavelet Support Vector Regression combined with Grey Wolf or Whale Optimization Algorithms optimize reaction parameters (temperature, solvent-to-water phase ratios, acid catalyst loading, and residence time). These systems achieve high target chemical yields (e.g., 40.9% mol 5-HMF yield with 59.8% mol selectivity) while minimizing side reactions (Chen et al., 2024).

**4.3 Reductive Catalytic Fractionation (RCF) and Lignin Monomer Yield Modeling**

Reductive catalytic fractionation, also known as "lignin-biorefining uses solid catalysts like Ru/C, Pd/C or Ni/C with proton-donating solvents and heat. The process breaks ether bonds in natural lignin, such as β-O-4 and α-O-4 linkages. It creates single-ring materials like propylguaiacol and propylsyringol while keeping cellulose mostly whole for more processing (Madadi et al. 2025).

Getting the results from RCF means finding the right mix of conditions. A machine-learning model that used extreme gradient boosting was trained on 3,451 results from 54 studies. It predicted lignin monomer yields with $R^2$ values between 0.80 and 0.86; RMSE values were between 3.99 and 8.31 and predictions were off by 2.0% to 2.6% compared to experiments (Webber et al. 2025). The study found that temperature, time and hydrogen pressure were responsible for 40%, to 57% of the models predictions. The type of feedstock contributed 25% to 43% and the catalyst and solvent properties made up 14% to 21% (Kenny et al., 2022).

Active-learning approaches, including Bayesian optimization, can also guide process adjustments in systems such as the AquaSolv Omni biorefinery. By varying hydrothermal temperature and reaction severity (P-factor), these methods help identify conditions that balance competing outcomes. The goal is to produce lignin with desired molecular weight and hydroxyl content while retaining a predicted proportion of β-O-4 linkages (Anderson et al., 2018).

**5. Machine Learning Applications in Value-Added Biomaterials Synthesis**

**5.1 Nanocellulose Extraction and Crystallinity Predictive Modeling**

Cellulose microfibrils obtained from preprocessed agricultural residues may be transformed into high-value CNC or CNF using controlled acid hydrolysis, enzymatic treatment, or mechanical blending. CNCs have high mechanical strength, specification surface, and adaptable surface chemistry for use as reinforcing materials for composites, barrier coatings and hydrogels (Wang et l., 2025).

Nanocellulose Crystallinity Index ( “CrI” ) is a vital quality indicator that controls the tensile strength, biological/thermal stability, and degradability of the nanocellulose. Concentration of the crystalline part resulting from cellulosic strong acid hydrolysis in fact, selectively hydrolyzes the amorphous regions in cellulose. Machine learning methods are used to relate the reaction conditions (acid strength, time, temperature, and liquid-to-solid ratio) to both the mass yield and the ultimate "CrI" (Abu-Ellail et al., 2026).

Transfer learning techniques have demonstrated promising performance on biomaterials where the datasets are limited. Models pretrained on well-known feedstock profiles (e.g., hardwood pulp) can be further fine-tuned on small datasets based on ag residue feedstocks such as wheat straw or sugarcane bagasse. Transfer learning-based yield prediction models achieved an RMSE of 2.84 and a R^2 of 0.84 whereas corresponding crystallinity prediction models achieved RMSE of 2.94 and R^2 of 0.83 (Nair & Verma, 2025).

**5.2 Lignin Macromolecular Tailoring and Polymer Blend Optimization**

Advances in technology have changed lignin from a waste material that was burned into a useful

polymer for bioplastics, resins, carbon fibers and polyurethanes. The value of lignin depends on its structure, which includes its molecular weight distribution, dispersity, aliphatic and phenolic hydroxyl content and thermally induced glass transition temperature (Tg) (Kapuge Dona & Smith 2025).

Sequential ML models are used to predict the properties of multi-component polymer blends that have functionalized lignins. As an example ML regressors predict the -linear mechanical properties of composites made from technical lignin and composites of polylactic acid (PLA) polyhydroxybutyrate (PHB) or other organic polymers. With the inclusion of adhesion, domain dispersion and thermal history ML models suggest the best blend ratios that improve tensile strength, elongation, at break and UV-blocking capability (Isnard et al., 2022).

**5.3 Polyhydroxyalkanoates (PHA) and Biopolymer Fermentation Tuning**

Because of advances, in technology industrial lignin has changed from a low-value fuel used in boilers to a molecule that can be used in bioplastics, resins, carbon fibers and polyurethanes. The uses of lignin depend on its chemical properties. These properties include the distribution of weight the spread of molecular sizes the amount of aliphatic and phenolic hydroxyl groups and the thermal glass transition temperature (T_g) (Kapuge Dona & Smith 2025). Machine learning models are used to estimate the physical properties of multi-component polymer blends that include chemically modified lignins. For example, ML regressors encode the composite non-linear mechanical response of composite mixtures of technical lignin with polylactic acid (PLA), polhydroxybutyrate (PHB) or synthetic polymers. Accounting for interfacial adhesion, domain dispersion, and thermal history, ML models also predict the optimal blend ratios that maximize tensile strength,Polyhydroxyalkanoates (PHAs) are the bio-based and biodegradable polyesters which are produced by microbial species (e.g. Cupriavidus necator) as the intracellular carbon and energy storage material (Koller, 2018).

Under nitrogen- or phosphorus-limiting conditions, microorganisms produce PHAs from volatile fatty acids, sugar hydrolysates, or organic side streams Machine learning (ML) models are developed to optimize PHA production by predicting microbial growth kinetics, substrate utilization rates, polymer accumulation rates, and monomer composition (e.g., fraction of 3-hydroxybutyrate to 3-hydroxyvalerate) (Hernández-Herreros et al., 2025).

Neural networks with bioprocess sensors in feedback loops have been used for dynamic adjustment of nutrient flows during fermentation that increase cellular polymer accumulation and minimize substrate toxicity elongation at break, and UV-blocking properties (Isnard et al., 2022).

**6. Systems-Level AI Integration: Digital Twins, Techno-Economic Analysis (TEA), and Life Cycle Assessment (LCA)**

**6.1 Multi-Objective Optimization for Co-Product Valorization Pathways**

Systems-level coordination is needed to achieve a balanced trade-off across simultaneous unit operations for a mature digital biorefinery. Hemi-cellulosic sugars are catabolized to furfural rather than fermented to bioethanol, decreasing the amount of sugar available for bioethanol fermentation; and harsh delignification leads to higher sugar yields but may result in degradation of the structural integrity of recovered lignin (Devarajan et al., 2026). Real-time application of process analytical technology (PAT) streams data continuously into digital twin frameworks, which drives multi-objective evolutionary algorithms to dynamically re-assign stream routing based on market valuations and carbon efficiency goals (Ejiyi et al., 2026).

Pareto-frontier based multi-objective optimization algorithms such as -dominated Sorting Genetic Algorithm II (NSGA-II) along with ANN surrogates are used to explore product slate configurations. These models consider real-time market prices, feedstock availability and utility costs to constantly change process streams, for

plant profitability and to meet sustainability goals (Nabi et al. 2026).

### 6.2 Techno-Economic Analysis (TEA) and Minimum Selling Price (MSP) Sensitivity

Techno-Economic Analysis (TEA) is used to evaluate the commercial viability of biorefinery configurations through the calculation of capital expenditures (CAPEX), operating expenditures (OPEX), Net Present Value (NPV), and Minimum Selling Prices (MSP). These models have traditionally been based on static spreadsheet formulations that do not adequately capture operational dynamic uncertainties (Ranganathan, P., & Savithri, 2019).

ML regressors are now integrated with TEA platforms to perform dynamic sensitivity analyses. By directly linking process prediction models to financial calculations, researchers can now observe the impact of variations in feedstock cost, enzyme pricing, catalyst lifespan, and sugar conversion efficiencies on product MSPs. The industrial application of ML-guided RCF technology processing 140 million tons of biomass per year is expected to create $4,729 million in socio-economic savings, which shows the efficiency gains that can be achieved through algorithmic optimization (Martinez-Hernandez et al., 2019).

### 6.3 Life Cycle Assessment (LCA) and Decarbonization Metrics

Life Cycle Assessment (LCA) quantifies the environmental performance (e.g., global warming potential, fossil energy demand, water consumption) of the entire value chain from farm gate to end-of-life of a product. AI-enhanced LCA frameworks combine operational plant data with regional supply chain data (Fernández-González et al., 2023). These models monitor energy flows, chemical consumption, and emission characteristics under varying processing conditions. On a system level, the implementation of machine-learning-optimized biorefineries processing 140 million tons of biomass per year has the potential to eliminate 20.6 million tons of 〖"CO" 〗_2 equivalent emissions per year, resulting in measurable industrial decarbonization (Bach et a., 2023).

**INTERNATIONAL JOURNAL OF ADVANCED RESEARCH**
*International Peer-Reviewed Research Journal*
**ISSN (Online): 3106-7980 | ISSN (Print): 3106-7972**
**Volume 3 • Issue 5 • September 2026**

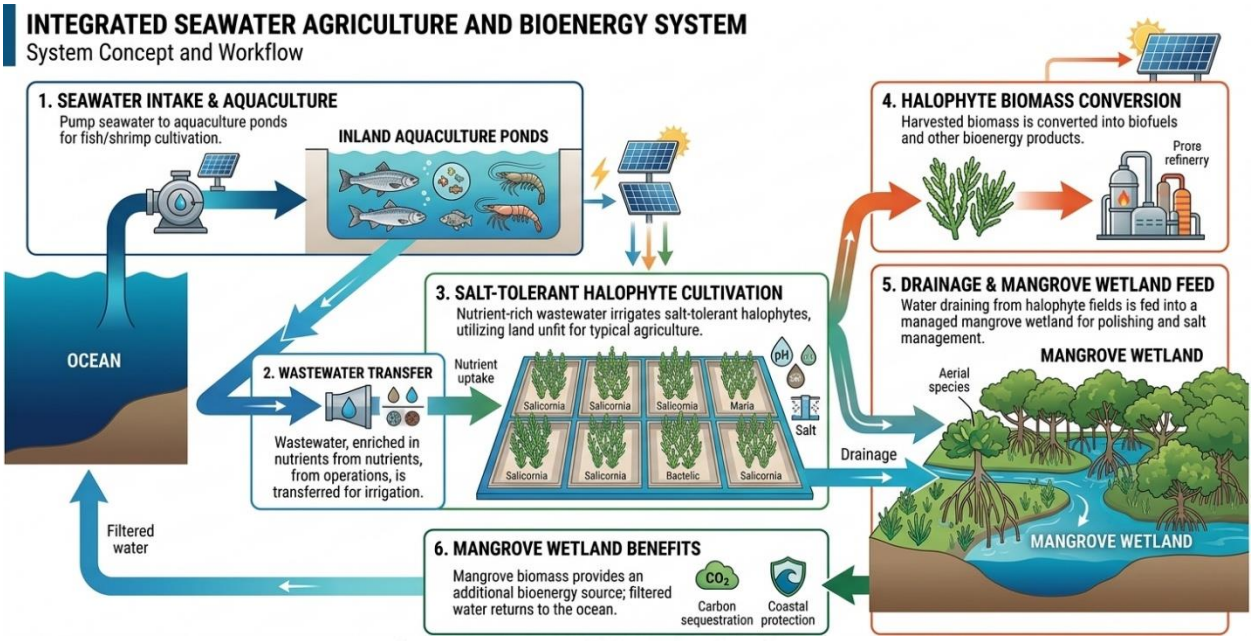



**Figure 2: Workflow and process concept for an integrated system that uses seawater for agriculture and bioenergy. This system connects aquaculture, growing plants that, like saltwater and storing carbon in wetlands.**

**Table 3. Performance Metrics of Prominent Machine Learning Algorithms Across Biorefinery Processing Stages**

| Processing Stage | Target Output Variable | Optimal Algorithm Architecture | R2 Score | RMSE Metric | Key Input Vector Parameters |
|---|---|---|---|---|---|
| **Pretreatment (Mild Alkaline / Steam Explosion)** | Hydrolysis Rate / Sugar Release | Artificial Neural Network (ANN) | 0.9805 | Not Reported | Steam duration, NaOH concentration, reaction time |
| **Pretreatment (Biphasic Fractionation)** | Delignification & Sugar Yield | GWO-Tuned Gradient Boosting | 0.9900 | 4.3800 | Temperature, liquid-to-solid ratio, residence time |
| **Enzymatic Saccharification** | Bioethanol Titer / Sugar Yield | Random Forest (RF) | 0.9110 | Not Reported | Enzyme loading, substrate loading, ionic liquid pretreatment severity |
| **Reductive Catalytic Fractionation (RCF)** | Lignin Monomer Yield | eXtreme Gradient Boosting (XGBoost) | 0.8000–0.8600 | 3.9900–8.3100 | Temperature, reaction time, hydrogen pressure, catalyst type, substrate ratios |
| **Biomaterial Extraction (Nanocellulose)** | Cellulose Nanocrystal Yield | Transfer-Learned Regressor | 0.8400 | 2.8400 | Acid concentration, reaction duration, hydrolysis temperature |

| | | | | | |
|---|---|---|---|---|---|
| **Biomaterial Extraction (Nanocellulose)** | Crystallinity Index (CrI) | Transfer-Learned Regressor | 0.8300 | 2.9400 | Biomass source features, acid-to-solid ratio, mechanical shear energy |

## 7. Challenges, Opportunities, and Future Perspectives

### 7.1 Data Heterogeneity, Small Dataset Limitations, and Transfer Learning

A big challenge when using AI in biorefining is how experimental data is created. The research in biorefining is spread out. Inconsistent. Different studies use feedstocks, pretreatment methods, reactor sizes and analysis tools. Because of this most studies only have datasets—fewer, than 100 data points. With little data complex AI models tend to overfit meaning they learn the noise instead of the real patterns (Niu et al. 2021).

To address the lack of data, researchers are using transfer learning, multi-task learning, and synthetic data generation through Generative Adversarial Networks (GANs) or Physics-Informed Neural Networks (PINNs). Transfer learning takes advantage of the knowledge gained from training on large benchmark datasets (e.g., pure cellulose conversion models) and applies it to smaller target domains, which are specific to certain feedstocks, requiring less training (Plested et al., 2026).

### 7.2 Explainable AI (XAI) and Hybrid First-Principles Modeling

Deep ANNs and sophisticated ensemble models are among the black-box ML models that may not be interpretable, hindering their deployment among industrial plant operators. The integration of Explainable AI (XAI) frameworks like SHapley Additive exPlanations (SHAP) and Local Interpretable Model-agnostic Explanation (LIME) affords researchers the ability to provide quantification of non-linear feature interactions and feature importance ranking graphs (Rajulapati et al., 2022).

Moreover, integrating mechanistic first-principles models with data-driven ML models results in resilient hybrids. In such hybrid frameworks, basic conservation laws (mass, energy and momentum balances) constrain the baseline physical solution space, while data-driven ML layers concurrently learn unobservable biochemical reaction kinetics, unproductive enzyme inhibition, and multiphase flow through porous media models (Sivaram & Venkatasubramanian, 2022).

### 7.3 Industrial Scale-Up and Autonomous Biorefinery Control

The end goal of AI application is a self-sufficient biorefinery. Achieving this goal requires the use of process analytical technology (PAT) (for example near-infrared (NIR) spectroscopy, Raman spectroscopy, in-line acoustic sensors) to continuously check feedstock composition, hydrolysate sugar profiles and microbial health parameters (Butean et al., 2025).

Sending live PAT data into edge-computer Digital Twin models allows time predictive control. These systems can also react to changes, in agricultural residues by changing reactor temperature, chemical application rate and residence time as needed. This constant feedback makes sure that target conversion efficiency is kept reactor instabilities are prevented and long-term cost competitiveness is secured (Fröner-Lacerda et al., 2024).

## Conclusion

This review is to provide a general insight into AI and ML application in agricultural residue biorefinery, which may ultimately impact the biorefinery design, its operation, and process optimization. The application of sophisticated predictive models in each stage pretreatment, saccharification, fermentation, thermochemical conversion, and biomaterial synthesis facilitates unparalleled control of process complexity and product quality. AI methodologies improve carbon utilization through tripartite co-production streams by modeling feedstock heterogeneity via composition-specific models and

reducing non-productive enzyme binding through tailored additive strategies. The use of digital twins in tandem with multi-objective evolutionary algorithms enables the dynamic rerouting according to market prices, feedstock delivery, and sustainability indicators, contributing to profit and decarbonization target maximization. Techno-economic analysis coupled with ML regressors enables dynamic sensitivity analyses, disclosing the trade-offs of $4,729 million/yr in global-scale socioeconomic savings. In addition, AI-enabled LCA analyses predict up to 20.6 million t of CO2—equivalent emissions could be reduced annually. Although challenges such as limited data, model interpretability, and validation in industry remain, a new generation of approaches including transfer learning, physics-informed neural networks, explainable AI, and online process analytical technology are closing the gap between laboratory research and commercial application. This is the long-term goal of this digital transformation: fully autonomous biorefineries that can optimize themselves on the fly as feedstock properties and market conditions change. This goal will be met through interdisciplinary efforts of bioprocess engineers, data scientists, and industrial stakeholders to create benchmark datasets, validate hybrid models at pilot scale, and design robust control architectures. In conclusion using intelligence to improve processes offers a practical way to build lignocellulosic biorefineries that are economically sound, good for the environment and helpful to society. These biorefineries can make a difference, in global efforts to cut carbon emissions and support a circular bioeconomy.

## REFERENCES

Abu-Ellail, F. F., Sakr, E. A., Kumar, T., Nour, S. A., Salim, R. G., El-Sayed, G. M., ... & Ying, X. (2026). Integrating artificial intelligence and conventional approaches in sugarcane bagasse biorefineries: a review towards a circular bioeconomy. *Frontiers in Bioengineering and Biotechnology*, *14*, 1879932.

Anderson, E. M., Stone, M. L., Hülsey, M. J., Beckham, G. T., & Román-Leshkov, Y. (2018). Kinetic studies of lignin solvolysis and reduction by reductive catalytic fractionation decoupled in flow-through reactors. *ACS Sustainable Chemistry & Engineering*, *6*(6), 7951-7959.

Árvai, C., Medgyesi, Z., Lui, M. Y., & Mika, L. T. (2024). The chemistry of levulinic acid: its potential in the production of biomass-based chemicals. *Advanced Synthesis & Catalysis*, *366*(23), 4846-4888.

Ayele, G., Admassu, H., Mosisa, G., Desalegn, A., & Abeje, M. (2026). Nutritional, Functional, and Safety Aspects of Single-Cell Protein Produced via Microbial Fermentation From Fruit-Processing Waste: A Comprehensive Review. *Comprehensive Reviews in Food Science and Food Safety*, *25*(3), e70464.

Bach, V. (2023). Life cycle assessment in the context of decarbonization and carbon neutrality. *The International Journal of Life Cycle Assessment*, *28*(7), 741-745.

Butean, A., Cutean, I., Barbero, R., Enriquez, J., & Matei, A. (2025). A review of artificial intelligence applications for biorefineries and bioprocessing: from data-driven processes to optimization strategies and real-time control. *Processes*, *13*(8), 2544.

Chen, C., Lv, M., Hu, H., Huai, L., Zhu, B., Fan, S., ... & Zhang, J. (2024). 5-Hydroxymethylfurfural and its downstream chemicals: a review of catalytic routes. *Advanced Materials*, *36*(37), 2311464.

Chen, Y., Li, B., Guo, X., Wang, S., Liu, Y., & Tang, Z. (2026). From Agricultural Waste to Industrial Feedstock: A Review on Multiphase Conversion Mechanisms and Material Reconstruction of Tomato Residues. *Agronomy*, *16*(12), 1177.

Cheng, M. H., & Rosentrater, K. A. (2025). Furthering the Application of a Low-Moisture Anhydrous Ammonia Pretreatment of Corn Stover. *Processes*, *13*(8), 2643.

Coronado-Contreras, S. A., Ibarra-Manzanares, Z. G., Casas-Rodríguez, A. D., Pastrana-Pastrana, Á. J., Sepúlveda, L., & Rodríguez-Herrera, R. (2025). Bio-circular economy and digitalization: Pathways for biomass valorization and sustainable biorefineries. *Biomass*, *6*(1), 1.

Danso-Boateng, E., & Achaw, O. W. (2022). Bioenergy and biofuel production from biomass using thermochemical conversions technologies—a review. *Aims Energy*, *10*(4), 585-647.

Devarajan, Y., T, R., Ks, S. R., Ps, R. R., Tripathy, S., Kohli, D., ... & Mehar, K. (2026). Waste-to-hydrogen ($WtH_2$) hubs: integrated conversion pathways, digital architectures, and deployment frameworks for circular industrial decarbonization. *International Journal of Sustainable Energy*, *45*(1), 2659444.

Di Domenico, G., Cioccolo, E., Bianchini, L., Venanzi, R., Colantoni, A., Picchio, R., ... & Di Stefano, V. (2025). A systematic review of mechanical pretreatment techniques of wood biomass for bioenergy. *Energies*, *18*(13), 3294.

Ejiyi, C. J., Djukem, D. L. W., Diokpo, C. N., Eze, F. O., Muoka, G. W., Sam, F., ... & Bamisile, O. O. (2026). Artificial Intelligence Methods for Clean Energy and Environmental Engineering Systems: A Cross-Sectoral Computational Review. *Archives of Computational Methods in Engineering*, 1-34.

Falowo, O. A., Oladipo, B., Adeyemo, P. O., Oyekola, O. O., & Ojumu, T. V. (2026). Covalent immobilization of amyloglucosidase on magnetic-silver nanoparticles for rapid and reusable saccharification of cassava peel. *Carbohydrate Research*, 110073.

Fernández-González, J., Rumayor, M., Domínguez-Ramos, A., Irabien, A., & Ortiz, I. (2023). The relevance of life cycle assessment tools in the development of emerging decarbonization technologies. *JACS Au*, *3*(10), 2631.

Fröner-Lacerda, L. R., Sganzerla, W. G., Lacerda, V. F., Sillero, L., Solera, R., Pérez, M., & Forster-Carneiro, T. (2024). Design, fabrication, automation, and scaleup of anaerobic reactors for waste management and bioenergy recovery. *Biofuels, Bioproducts and Biorefining*, *18*(5), 1093-1106.

Gao, J., Zhou, F., Li, X., Deng, Y., Wu, Z., Feng, S., ... & Xiao, Z. (2026). A Review of Grain Boundaries: Formation Mechanism, Synthesis Strategy, and Application in Electrocatalysis. *EcoEnergy*, e70074.

Hernández-Herreros, N., Rodriguez, A., Rivero-Buceta, V., Rojas, A., & Prieto, M. A. (2025). Flexible feeding strategy for high-yield PHA bioprocessing in Cupriavidus necator H16 from anaerobically fermented industrial wastewater. *Bioresource Technology*, *434*, 132774.

Hidalgo, D., Urueña, A., Martín-Marroquín, J. M., & Díez, D. (2025). Integrated approach for biomass conversion using thermochemical routes with anaerobic digestion and syngas fermentation. *Sustainability*, *17*(8), 3615.

Hu, Y., & Chen, Z. (2025). Thermochemical conversion of sewage sludge: Progress in pyrolysis and gasification. *Water*, *17*(12), 1833.

Isnard, F., Monga Mulunda, M., Rubens, M., Marien, H., Harumashi, T., Wakabayashi, K., ... & Vendamme, R. (2022). Stabilization of hybrid adhesives and sealants by thermodynamic tuning of molecularly optimized lignin bio-additives: small changes, big effects. *Biomacromolecules*, *23*(8), 3174-3185.

Jayabal, R., & Prabhakar, P. (2026). Advanced strategies in food and agricultural waste valorisation: integrating nanoengineering, artificial intelligence and smart bioconversion for sustainable resource recovery. *Journal of the Science of Food and Agriculture*, *106*(8), 4507-4520.

John, S. U., Onu, C. E., Ezechukwu, C. M. J., Nwokedi, I. C., & Onyenanu, C. N. (2025). Multi-product biorefineries for biofuels and value-added products: advances and future perspectives. *Academia Green Energy*, *2*(1).

Kapuge Dona, N. L., & Smith, R. C. (2025). Tailoring polymer properties through lignin addition: A Recent perspective on lignin-derived polymer modifications. *Molecules*, *30*(11), 2455.

Kenny, J. K., Brandner, D. G., Neefe, S. R., Michener, W. E., Román-Leshkov, Y., Beckham, G. T., & Medlin, J. W. (2022). Catalyst choice impacts aromatic monomer yields and selectivity in hydrogen-free reductive catalytic fractionation. *Reaction Chemistry & Engineering*, *7*(12), 2527-2533.

Koller, M. (2018). A review on established and emerging fermentation schemes for microbial production of polyhydroxyalkanoate (PHA) biopolyesters. *Fermentation*, *4*(2), 30.

Kucharska, K., Rybarczyk, P., Hołowacz, I., Łukajtis, R., Glinka, M., & Kamiński, M. (2018). Pretreatment of lignocellulosic materials as substrates for fermentation processes. *Molecules*, *23*(11), 2937.

Lilaria, K., Tiwari, S., & Jadhav, S. K. (2026). From Statistical Design to Machine Learning and Omics-Driven Strategies: A Multi-Dimensional Perspective on Cellulase Optimization. *Archives of Computational Methods in Engineering*, *33*(5), 7225-7252.

Liu, T., He, M., Shi, R., Yin, H., & Luo, W. (2025). Biofuel–pharmaceutical Co-Production in integrated biorefineries: strategies, challenges, and sustainability. *Fermentation*, *11*(6), 312.

Madadi, M., Kargaran, E., Hashemi, S. S., Sun, C., Denayer, J. F., Karimi, K., ... & Gupta, V. K. (2025). Scalable lignin monomer production via machine learning-guided reductive catalytic fractionation of lignocellulose. *Advanced Science*, *12*(42), e10496.

Manhongo, T. T. (2023). *A comparative investigation of the technoeconomic feasibility and sustainability of mango waste biorefineries in South Africa: a process modeling approach* (Doctoral dissertation, Stellenbosch: Stellenbosch University,).

Martinez-Hernandez, E., Ramírez-Verduzco, L. F., Amezcua-Allieri, M. A., & Aburto, J. (2019). Process simulation and techno-economic analysis of bio-jet fuel and green diesel production—Minimum selling prices. *Chemical Engineering Research and Design*, *146*, 60-70.

Mehta, N. (2026). Synergistic pretreatment: hybrid strategies for maximum lignocellulose valorization. *Biodegradation*, *37*(3), 74.

Nabi, M. H. B., Akter, L., Chowdhury, Y., Any, M. A., Islam, N., Murtini, E. S., ... & Zzaman, W. (2026). Citrus Essential Oils in Food Safety and Health: Advances in AI-Assisted Authentication, Antimicrobial Applications, Sustainable Extraction, and Regulatory Translation. *Food Safety and Health*.

Nagappan, B., Subbiah, G., Paliwal, R. K., Choudhury, S., Rai, K., Mehar, K., ... & Kamakshi Priya, K. (2025). Data-driven optimization of biomass conversion pathways: integrating thermochemical processes. *International Journal of Chemical Reactor Engineering*, *23*(11), 1309-1326.

Nair, L. G., & Verma, P. (2025). Harnessing carbon potential of lignocellulosic biomass: advances in pretreatments, applications, and the transformative role of machine learning in biorefineries. *Bioresources and Bioprocessing*, *12*(1), 97.

Niu, S., Liu, Y., Wang, J., & Song, H. (2021). A decade survey of transfer learning (2010–2020). *IEEE Transactions on Artificial Intelligence*, *1*(2), 151-166.

Parthipan, P., Santhoshkumar, M., Shanmugapriya, D., Duraisamy, N., Alam, M. M., Al-Sehemi, A. G., & Sathishkumar, K. (2026). Algal Biorefineries for a Circular Bioeconomy: Integrating Nanobiomaterial Synthesis, Multi-Product Valorization and Sustainable Manufacturing. *Waste and Biomass Valorization*, 1-18.

Pilarski, K., & Pilarska, A. A. (2025). Kinetics and energy yield in anaerobic digestion: Effects of substrate composition and fundamental operating conditions. *Energies*, *18*(23), 6262.

Plested, J., Phiri, M., & Gedeon, T. (2026). Deep transfer learning for image classification: a survey. *Artificial Intelligence Review*, *59*(3), 100.

Podgorbunskikh, E., Skripkina, T., & Bychkov, A. (2026). Mechanical Pretreatment of Plant Biomass: Mechanisms, Energy Efficiency, Technologies, and Life Cycle Assessment. *Polysaccharides*, *7*(2), 38.

Posilipo, A. (2024). *Optimization of Biochemical Saccharification of Lignocellulosic Biomass* (Doctoral dissertation, PhD Thesis Università degli Studi di Napoli Federico II).

Pusuluru, S., Velmurugan, K., Punna, S. K., Ravikumar, M., Kumar, M., Almukhlifi, H. A., ... & Menaa, F. (2026). Advancing biofuel production through algal biorefineries: Optimization of polysaccharide utilization, AI-driven innovations, and sustainability assessments. *Biomass and Bioenergy*, *204*, 108366.

Rajulapati, L., Chinta, S., Shyamala, B., & Rengaswamy, R. (2022). Integration of machine learning and first principles models. *AIChE Journal*, *68*(6), e17715.

Ranganathan, P., & Savithri, S. (2019). Techno-economic analysis of microalgae-based liquid fuels production from wastewater via hydrothermal liquefaction and hydroprocessing. *Bioresource technology*, *284*, 256-265.

Saha, B. C. (2003). Hemicellulose bioconversion. *Journal of industrial microbiology and biotechnology*, *30*(5), 279-291.

Sivaram, A., & Venkatasubramanian, V. (2022). XAI-MEG: Combining symbolic AI and machine learning to generate first-principles models and causal explanations. *AIChE Journal*, *68*(6), e17687.

Šivec, R., Grilc, M., Huš, M., & Likozar, B. (2019). Multiscale modeling of (hemi) cellulose hydrolysis and cascade hydrotreatment of 5-hydroxymethylfurfural, furfural, and levulinic acid. *Industrial & Engineering Chemistry Research*, *58*(35), 16018-16032.

Wang, S., Du, W., Liu, W., Li, J., Yi, Z., Chen, Z., ... & Li, M. (2025). Machine learning-driven prediction and optimization of microcrystalline cellulose and cellulose nanocrystal production from Miscanthus biomass. *Industrial Crops and Products*, *237*, 122180.

Webber, M. S., Puente-Urbina, A., Myers, L. C., Woodworth, S. P., Pellerin, E. J., Jang, J. H., ... & Román-Leshkov, Y. (2025). Lignin-derived phenolic compounds and water are effective cosolvents for reductive catalytic fractionation. *ACS Sustainable Chemistry & Engineering*, *13*(44), 19199-19209.

Wu, Z., Sahaluddin, M., Bani Mustafa, D., Ghosh, S., & Bedewy, M. (2026). Laser-induced graphene: from precursor chemistry to process control and throughput–resolution–performance trade-offs. *Nanoscale*, *18*(13), 6647-6686.

Wyman, C. E., Decker, S. R., Himmel, M. E., Brady, J. W., Skopec, C. E., & Viikari, L. (2005). Hydrolysis of cellulose and hemicellulose. *Polysaccharides: Structural diversity and functional versatility*, *1*, 1023-1062.

Xie, D., Ma, H., Wang, J., Cárdenas-Oscanoa, A. J., Wang, S., & Huang, C. (2026). Multi-Scale Modeling of Enzymatic Hydrolysis Efficiency Based on Machine Learning: From Hydrothermal Pretreatment Conditions to Lignin Structural Properties. *ACS Sustainable Chemistry & Engineering*, *14*(12), 6204-6222.

Xie, D., Ma, H., Wang, J., Cárdenas-Oscanoa, A. J., Wang, S., & Huang, C. (2026). Multi-Scale Modeling of Enzymatic Hydrolysis Efficiency Based on Machine Learning: From Hydrothermal Pretreatment Conditions to Lignin Structural Properties. *ACS Sustainable Chemistry & Engineering*, *14*(12), 6204-6222.

Xu, H., Zeng, X., Bin, Y., Wu, R., Qin, C., Liang, C., ... & Liu, B. (2026). Interpretable modeling of biomass fractionation under acidic pretreatment via multi-step data augmentation and an entropy-weighted TOPSIS ensemble. *Bioresource Technology*, 135324.

Yu, Q., Zhuang, X., Lv, S., He, M., Zhang, Y., Yuan, Z., ... & Tan, X. (2013). Liquid hot water pretreatment of sugarcane bagasse and its comparison with chemical pretreatment methods for the sugar recovery and structural changes. *Bioresource technology*, *129*, 592-598.

Yu, Q., Zhuang, X., Lv, S., He, M., Zhang, Y., Yuan, Z., ... & Tan, X. (2013). Liquid hot water pretreatment of sugarcane bagasse and its comparison with chemical pretreatment methods for the sugar recovery and structural changes. *Bioresource technology*, *129*, 592-598.

Zhong, Z., Gorish, B. M. T., Bai, Y., Abdelmula, W. I. Y., Dang, W., & Zhu, D. (2025). Predicting lignin removal efficiency in deep eutectic solvent-based biomass fractionation: an explainable machine learning approach. *Green Chemistry*, *27*(36), 11036-11054.