Energy and AI 18 (2024) 100414

Contents lists available at ScienceDirect

Energy and AI

journal homepage: www.sciencedirect.com/journal/energy-and-ai

BCLH2Pro: A novel computational tools approach for hydrogen production prediction via machine learning in biomass chemical looping processes

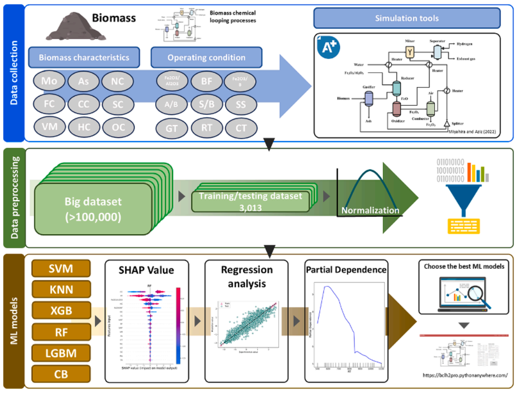
Thanadol Tuntiwongwat [a], Sippawit Thammawiset [b], Thongchai Rohitatisha Srinophakun [c], Chawalit Ngamcharussrivichai [d,e], Somboon Sukpancharoen [f,g,*]

[a] *Department of Mechanical Engineering, Faculty of Engineering, Khon Kaen University, Khon Kaen 40002, Thailand*
[b] *Department of Automation Robotics and Intelligent System, Faculty of Engineering, Khon Kaen University, Khon Kaen 40002, Thailand*
[c] *Department of Chemical Engineering, Faculty of Engineering, Kasetsart University, Bangkok 10900, Thailand*
[d] *Center of Excellence in Catalysis for Bioenergy and Renewable Chemicals (CBRC), Faculty of Science, Chulalongkorn University, Pathumwan, Bangkok 10330, Thailand*
[e] *Institute of Nanoscience and Nanotechnology (ION2), Universiti Putra Malaysia, Serdang 43400, Selangor, Malaysia*
[f] *Department of Agricultural Engineering, Faculty of Engineering, Khon Kaen University, Khon Kaen 40002, Thailand*
[g] *Center for Alternative Energy Research and Development, Khon Kaen University, Khon Kaen 40002, Thailand*

HIGHLIGHTS

- BCLH2Pro: Novel ML tool predicts H_2 yield in biomass chemical looping processes.
- CatBoost algorithm achieves over 98% accuracy in H_2 yield predictions.
- SHAP analysis reveals key factors: carbon content, reducer temp, Fe_2O_3/Al_2O_3 ratio.
- User-friendly web interface optimizes BCLpro operational parameters.

GRAPHICAL ABSTRACT



ARTICLE INFO

Keywords:
Artificial intelligence
Supervised learning
Process simulation
Chemical looping
Agricultural waste

ABSTRACT

This study optimizes biomass chemical looping processes (BCLpro), a technique for converting biomass to energy, through machine learning (ML) for sustainable energy production. The study proposes an integrated Fe_2O_3-based BCLpro combining steam gasification for H_2 production. Aspen Plus is used as the primary tool to generate extensive datasets covering 24 biomass types with 18 feature inputs in a supervised model. A methodology involving K-Nearest Neighbors (KNN), Extreme Gradient Boosting (XGB), Light Gradient Boosting Machine (LGBM), Support Vector Machine (SVM), Random Forest (RF), and CatBoost (CB) algorithms was employed to predict H_2 yields in the BCLpro, utilizing 10-fold cross-validation for robust model evaluation. Findings highlight the CB algorithm's superior performance, achieving up to 98% predictive accuracy, with carbon content, reducer temperature, and Fe_2O_3/Al_2O_3 mass ratio identified as crucial features. The algorithm has been developed into a

[*] Corresponding author. Department of Agricultural Engineering, Faculty of Engineering, Khon Kaen University, Khon Kaen 40002, Thailand.
E-mail address: sombsuk@kku.ac.th (S. Sukpancharoen).

https://doi.org/10.1016/j.egyai.2024.100414

Available online 13 August 2024

user-friendly tool, BCLH2Pro, accessible via a web server. This tool is designed to assist in reducing costs, optimizing biomass selection, and planning operational conditions to maximize H_2 yield in BCLpro systems. Access to the tool can be obtained through the following link: http://bclh2pro.pythonanywhere.com/.

1. Introduction

Over the last decade, escalating environmental concerns have stemmed from emissions such as carbon dioxide (CO_2), nitrogen dioxide (NO_2), and sulfur dioxide (SO_2) from fossil fuel combustion, contributing to global warming and climate change. Industrialization-driven CO_2 emissions necessitate prioritizing renewables for energy sector decarbonization and climate change mitigation. While solar, wind, hydro, tidal, and geothermal sources face weather-dependent challenges, biomass stands out for its abundance and environmental friendliness. Biomass, with ample chemical energy, emerges as an ideal resource for replacing fossil fuels, underscoring its pivotal role in a sustainable energy transition [1-5].

Biomass constitutes 37.2% of the Organization for Economic Cooperation and Development (OECD) total primary energy supply and isrecognized globally as a vital renewable resource. This category encompasses liquid biofuels, biogases, and renewable municipal waste as components of modern bioenergy. Derived from diverse sources like agricultural waste and industrial byproducts, biomass stands out for its adaptability, utilizing organic materials that would otherwise go to waste. Hydrogen (H_2) derived from biomass, is gaining attention as a promising clean energy source due to its high energy density and minimal emissions [6-9].

H_2 is gaining widespread attention as a promising energy source due to its high energy density and minimal greenhouse gas emissions, with its modern applications illustrated in Fig. 1. Its versatility arises from its ability to be generated from various low-carbon sources, making it an environmentally friendly option. With a higher heating value (HHV) of 141.8 MJ/kg, H_2 surpasses many other fuels in energy content, like gasoline with 44 MJ/kg. Steam gasification of biomass, which yielding approximately 53–55 %vol. of H_2, is a favored method for H_2 production. Chemical looping processes (CLP) provide an alternative route for H_2 production [5,9-13].

CLP revolutionize traditional reactions by utilizing a solid looping material's redox cycle, typically dividing the overall reaction into two sub-reactions in distinct reactors [14]. Iron oxide mediates redox reactions in chemical looping, allowing heat production from fuel combustion without releasing CO_2 emissions [15]. Chemical looping comprises two divisions: direct chemical looping (DCL) and syngas chemical looping (SCL) [16]. Biomass chemical looping processes (BCLpro) conversion efficiently harnesses biomass energy, converting it into H_2-rich syngas with solid metal oxides as oxygen carriers [17]. Thermochemical biomass conversion's key parameters [5], such as particle size, gasification temperature, equivalence ratio (ER), gasification agents, and steam/biomass (S/B) ratio, impact H_2 yield and production rate as shown in Fig. 2. Quantitatively assessing and optimizing these processes, either through experimental or simulation approaches, is challenging due to research limitations on specific biomass sources or model compounds. This complexity presents opportunity for the application of advanced data analysis techniques [18].

Recent advances in artificial intelligence (AI) have revolutionized data utilization in chemical engineering, particularly through machine learning (ML) [19]. ML, which is a statistical AI category, relies on existing datasets to make predictions and inferences, identifying patterns for complex tasks. This technology enhances insights into catalyst development, the reaction conditions refinement, and the reactor configuration optimization within the engineering domain [20,21].

ML operations, as outlined by [22], involve classification, regression, prediction, and clustering. Classification discerns input categories, while regression formulates models capturing input-output relationships, prediction forecasts future values, and clustering identifies and groups similar points in datasets. In academia, AI and ML algorithms are utilized to build predictive models for energy consumption, reducing uncertainties in renewable energy sources [23]. ML plays a crucial role in optimizing BCLpro in several ways:

1. Predictive Modeling: ML algorithms can predict H_2 yield based on various input parameters, allowing for optimization of process conditions without extensive experimental work.
2. Feature Importance: ML techniques like SHapley Additive exPlanations (SHAP) help identify the most influential factors in H_2 production, guiding focused research efforts.
3. Process Optimization: By analyzing complex relationships between multiple variables, ML can suggest optimal operating conditions for maximizing H_2 yield.
4. Rapid Iteration: ML models can quickly evaluate numerous scenarios, accelerating the process of finding optimal biomass compositions and operating conditions.
5. Cost Reduction: By reducing the need for extensive laboratory experiments, ML helps lower research and development costs in BCLpro optimization.
6. Scalability: ML models are easily scaled to incorporate new data, continuously improving predictions as more information becomes available.

These advantages make ML an invaluable tool in advancing BCLpro technology, enabling a more efficient and sustainable H_2 production from biomass resources.

Miyahira and Aziz introduced a tailored conversion system for rice husks, optimizing H_2 production efficiency. The system integrates superheated steam drying, steam gasification, chemical looping, and the Haber-Bosch process. Three systems were compared: dual fluidized bed chemical looping with H_2O and CO_2, and single fluidized bed chemical looping. Using Aspen Plus for modeling and optimization, their study revealed peak efficiencies of 51.80% for H_2, 38% for NH_3, and 0.651% for net power [16]. This comprehensive investigation advances the understanding of efficient biomass-derived H_2 production. However, the proposed integrated conversion system for rice husk biomass has limitations for practical applications: firstly, it considers only one type of biomass, and secondly, process modeling methods often rely on

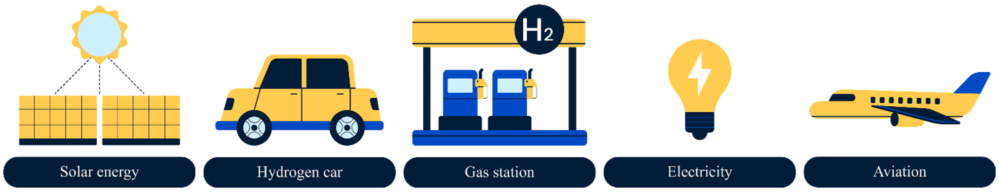


Fig. 1. The implementation of H_2 energy in various sectors.

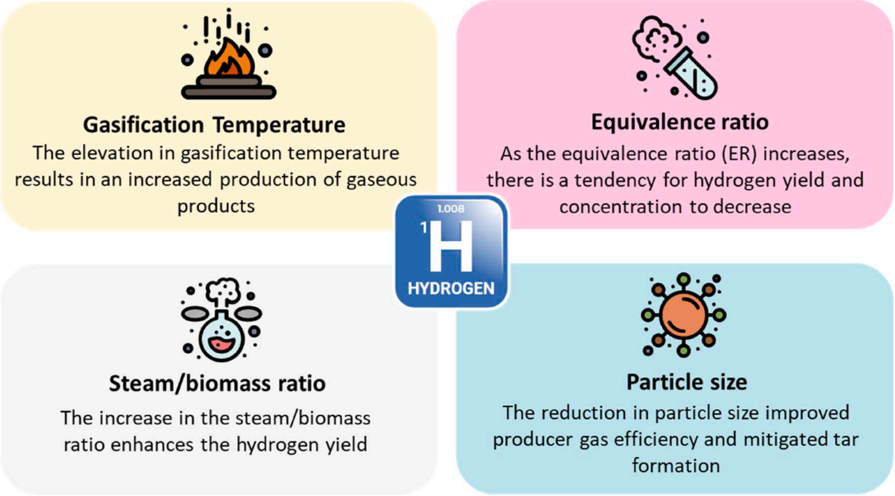

Gasification Temperature
The elevation in gasification temperature results in an increased production of gaseous products
Equivalence ratio
As the equivalence ratio (ER) increases, there is a tendency for hydrogen yield and concentration to decrease
1.008
1
H
HYDROGEN
Steam/biomass ratio
The increase in the steam/biomass ratio enhances the hydrogen yield
Particle size
The reduction in particle size improved producer gas efficiency and mitigated tar formation

Fig. 2. Effect of different factor to H_2 yield.

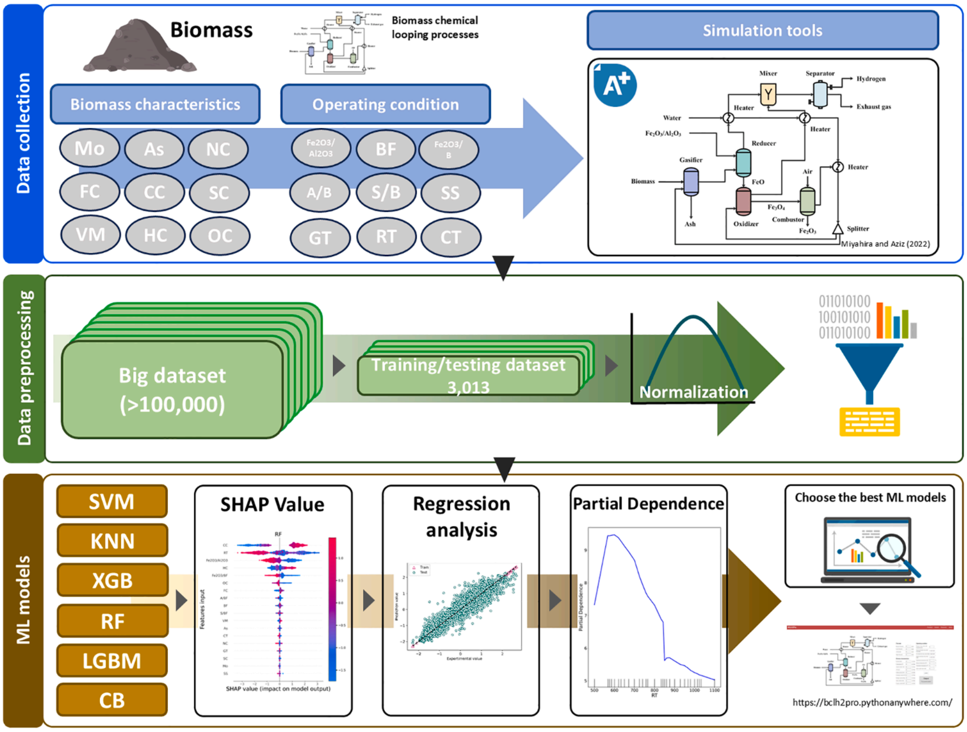


Fig. 3. A schematic flowchart of this study.

specialized software tools, such as Aspen Plus [24]. These factors may limit the broader applicability of the approach to diverse biomass sources and different modeling environments.

Aspen Plus, a widely recognized commercial software, exhibits high capabilities for modeling intricate chemical processes and optimizing complex processes like biomass gasification. Supported by validation databases and professional software assistance, its utilization, however, necessitates the acquisition and maintenance of a valid software license [25,26]. Therefore, it is imperative to develop reliable, accurate, robust, and efficient modeling approaches [24].

This study introduces BCLH2Pro, a novel and user-friendly tool specifically designed to assist researchers in the field of BCLpro. Developed without cost, BCLH2Pro leverages extensive datasets generated through simulations of BCLpro with Fe_2O_3/Al_2O_3 for H_2 production from biomass using Aspen Plus software. The tool incorporates a diverse range of 24 biomass feedstocks, including almond shell, coffee husk, and oat straw, and explores these under various operating conditions. A total of 18 input variables and one target output variable (H_2 yield) are employed in the analysis. Supervised ML techniques play a crucial role in BCLH2Pro's functionality. Six well-established algorithms, namely Support Vector Machine (SVM), K-Nearest Neighbors (KNN), eXtreme Gradient Boosting (XGB), Random Forest (RF), CatBoost (CB), and LightGBM (LGBM), and a 10-fold cross-validation approach. Model performance is meticulously evaluated using five key metrics: Coefficient of determination (R^2), Mean Absolute Error (MAE), Normalized Mean Absolute Error (NMAE), Root Mean Square Error (RMSE), and Normalized Root Mean Square Error (NRMSE). Ultimately, the model demonstrating the highest degree of accuracy and robustness is selected for developing the final BCLH2Pro tool.

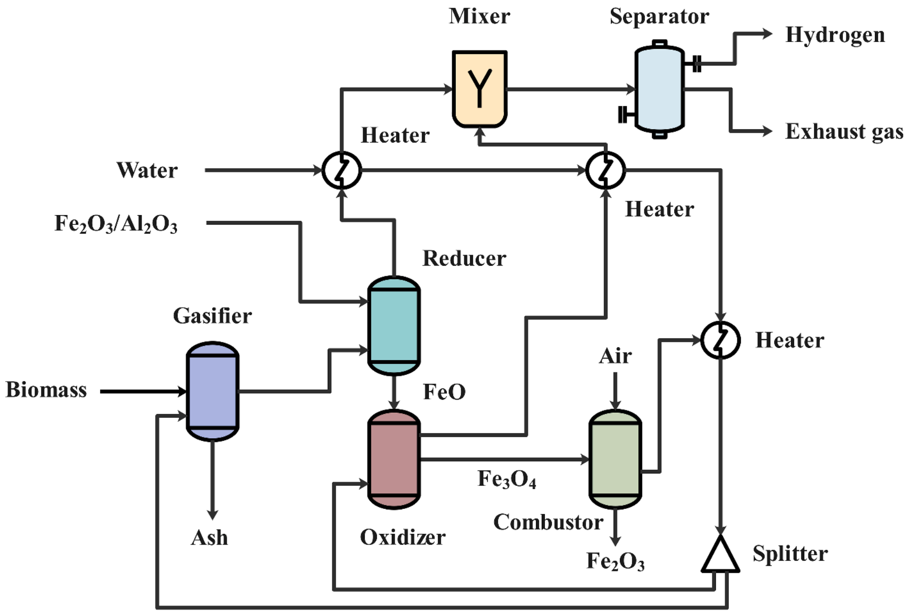


Fig. 4. Block diagram of BCLpro system.

2. Materials and methods

In this study, process and data simulations were generated using Aspen Plus V. 12.1. Data analysis and modeling employed ML techniques implemented through Python within Jupyter notebooks. The models were constructed utilizing the 'scikit-learn' library in Python [27]. All computations were executed on a system equipped with a 13th generation Intel(R) Core (TM) i9–13900KF processor operating at 3.00 GHz, complemented by 32 GB of installed RAM. The operating system was 64-bit, and the processor architecture is x64. A flowchart depicting the methodology is presented in Fig. 3.

2.1. *Biomass chemical looping processes (BCLpro) description*

This study modeled H_2 production efficiently using Miyahira and Aziz's system, primarily relying on biomass for its high energy potential [16]. The process incorporates steam gasification and chemical looping, detailed in Fig. 4., facilitating a systematic approach to optimize H_2 production with high energy efficiency in the BCLpro system, biomass undergoes gasification in the gasifier, transitioning from 30 °C and 1 bar to 500–1,100 °C. The resulting syngas moves to the reducer, reacting at 500–1,100 °C with Fe_2O_3/Al_2O_3 to produce CO_2 and H_2O. The reduced oxygen carrier proceeds to the oxidizer, generating H_2 and Fe_3O_4 by reacting with steam. The carrier then enters the combustor, forming Fe_2O_3 with air at 1,000–1,500 °C. Heat is transferred to the reducer. The exhaust gasses emanating from the reducer, oxidizer, and combustor serve the purpose of generating steam for the processes of gasification and oxidation. Additionally, these gases, along with CO_2 and H_2O from the reducer, enter a separator. H_2 is extracted, while other exhaust gases are discharged. This systematic process model optimizes H_2 production from biomass, with high energy efficiency.

The following reaction in the gasifier are considered to occur, the primary reactants are char and H_2, converting rapidly into CO_2 and H_2O, as depicted in Eqs. (1) - (6) [28]. The ultimate product stream emerges as a result of the reactions between gasification and oxidation products. H_2, carbon monoxide (CO), and methane (CH_4) production take place through processes such as the water gas shift reaction (Eq. (7)), methane reforming (Eq. (8)), char CO_2 gasification (Eq. (9)), and char steam gasification (Eq. (10)).

$$\text{Biomass} \rightarrow (H_2, CO, CO_2, CH_4)\ \text{gas} + \text{water} + \text{char} + \text{tar} \tag{1}$$

$$\text{Tar} \rightarrow CO_2+C+CH_4+CO \tag{2}$$

$$\text{Tar} + H_2O \rightarrow H_2+CO \tag{3}$$

$$C(\text{Solid}) + \frac{1}{2}O_2 \rightarrow CO \tag{4}$$

$$CO + \frac{1}{2}O_2 \rightarrow CO_2 \tag{5}$$

$$H_2 + \frac{1}{2}O_2 \rightarrow H_2O \tag{6}$$

$$CO + H_2O \leftrightarrow H_2 + CO_2 \tag{7}$$

$$CH_4 + H_2O \leftrightarrow CO + 3H_2 \tag{8}$$

$$C + CO_2 \rightarrow 2CO \tag{9}$$

$$C + H_2O \rightarrow H_2+CO \tag{10}$$

In the reduction stage, the syngas reacts with Fe_2O_3/Al_2O_3, producing CO_2 and H_2O, as depicted in Eqs. (11) – (15) [16].

$$Fe_2O_3 + CO \rightarrow 2FeO + CO_2 \tag{11}$$

$$FeO + CO \rightarrow Fe + CO_2 \tag{12}$$

$$Fe_2O_3 + H_2 \rightarrow 2FeO + H_2O \tag{13}$$

$$FeO + H_2 \rightarrow Fe + H_2O \tag{14}$$

$$4Fe_2O_3 + 3CH_4 \rightarrow 8Fe + 3CO_2 + 6H_2O \tag{15}$$

The oxidizer unit receives the reduced oxygen carrier, which then enters and reacts with steam, generating H_2 and Fe_3O_4. The reaction described in Eqs. (16) and (17) are considered within the oxidizer.

$$Fe + H_2O \rightarrow FeO + H_2 \tag{16}$$

$$3FeO + H_2O \rightarrow Fe_3O_4 + H_2 \tag{17}$$

In the combustor unit, the reduced oxygen carrier mixes with air and forms Fe_2O_3. The reaction described in Eq. (18) occurs within the combustor.

$$4Fe_3O_4 + O_2 \rightarrow 6Fe_2O_3 + N_2 \tag{18}$$

Table 1
Input and output variables with statistical properties for BCLpro modeling.

Variable	Name	Units	Boundary values	Average	S.D.
Input features					
Biomass characteristics					
Moisture	Mo	vol.%db.	2.27–15.00	6.95	3.62
Fixed carbon	FC	vol.%db.	5.57–75.90	22.51	12.81
Volatile matter	VM	vol.%db.	19.00–91.70	71.37	12.77
Ash	As	vol.%db.	0.88–19.91	6.12	5.37
Carbon content	CC	vol.%db.	28.85–54.02	43.00	7.16
Hydrogen content	HC	vol.%db.	3.26–8.84	5.82	1.27
Nitrogen content	NC	vol.%db.	0.15–2.86	0.77	0.5
Sulfur content	SC	vol.%db.	0.00–2.96	0.79	0.97
Oxygen content	OC	vol.%db.	34.99–50.62	43.49	4.02
Operating condition					
Fe_2O_3 to Al_2O_3 mass ratio	Fe_2O_3/Al_2O_3	kg/hr:kg/hr	0.30–0.80	0.52	0.15
Biomass flowrate	BF	kg/hr	2,496.11–11,339.81	7,160.78	2,496.11
Fe_2O_3 to biomass mass ratio	Fe_2O_3/B	kg/hr:kg/hr	6.67–39.68	12.51	5.09
Air to biomass mass ratio	A/B	kg/hr:kg/hr	0.69–52.91	1.96	4.09
Steam to biomass mass ratio	S/B	kg/hr:kg/hr	1.77–6.61	3.18	1.21
Steam split fraction	SS	–	0.30–0.70	0.5	0.04
Gasification temperature	GT	°C	500–1,100	814.55	214.12
Reducer temperature	RT	°C	500–1,100	745.1	169.27
Combustor temperature	CT	°C	1,000–1,500	1,250.33	168.86
Output target					
H_2 Yield	HY	%	0.495-14.928	7.40	3.07

The H_2 yield is determined using Eq. (19) [29,30].

$$H_2\text{yield}(\%) = \frac{H_2\ \text{mass flow output (kg/hr)}}{\text{Feedstock mass flow (dry basis)(kg/hr)}}\text{x100} \tag{19}$$

2.2. Feedstock data selection

This study extensively reviewed the literature to select a diverse range of biomass raw materials. A meticulous approach considered the distinctive properties of each biomass sample, intentionally selecting comparable characteristics to minimize potential bias in the training dataset. Proximate and ultimate analyses were conducted on 24 meticulously selected biomass samples from Kartal and Ozveren [31], quantifying parameters such as moisture (Mo), volatile matter (VM), ash (As), fixed carbon (FC), carbon content (CC), hydrogen content (HC), oxygen content (OC), nitrogen content (NC), and sulfur content (SC).

2.3. Data generation and preprocessing

The processes executed by BCLpro, which were simulated using Aspen Plus, underwent a meticulous validation process. Subsequently, the platform is prepared to serve as a tool for the generation of datasets. Information pertaining to the distinctive characteristics of each biomass is inputted into the feed line. Following this, adjustments are made to the operating conditions to diversify the data range specified for the variables, thus providing a comprehensive overview of the static features, as detailed in Table 1. A total of 106,795 datasets were generated, with 2,000-5,000 datasets for each of the 24 biomass types. To optimize processing time and resource allocation, 3,013 datasets were randomly selected, ensuring each biomass type is represented by 120–130 datasets. This balanced subset provides a representative sample for analysis.

Normalization is crucial for training the model as it ensures that data, especially with varying value ranges, are adjusted for balanced influence, preventing disproportionate impacts on prediction accuracy. In this study, normalization, described by Eq. (20) [32], transforms data into standardized ranges like −1 to 1 and 0 to 1 [33]. This process equalizes data attributes, facilitating analysis and ML by eliminating distortions arising from raw, unstandardized data collected from diverse sources [34,35].

$$X' = \frac{X - \mu}{\sigma} \tag{20}$$

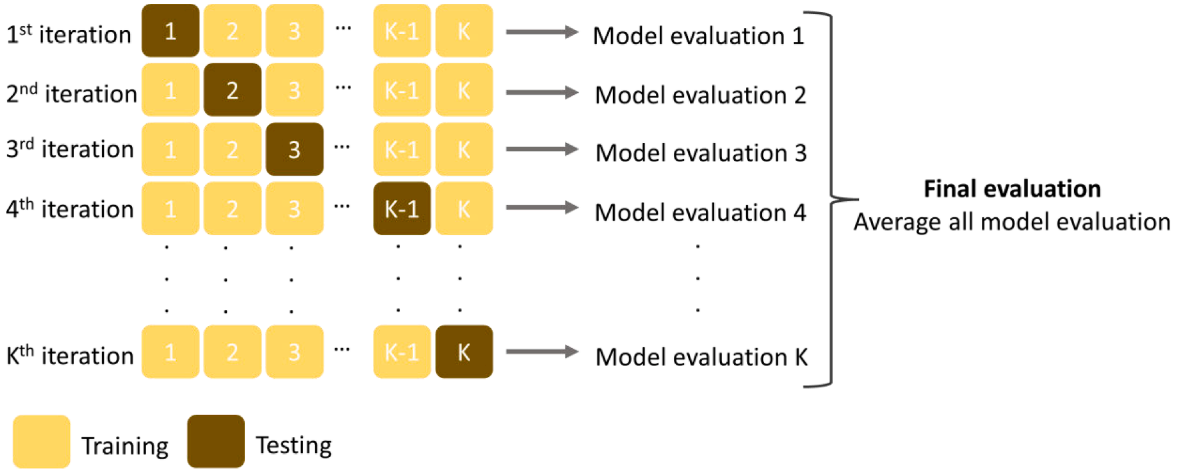


Fig. 5. K^{th}-fold cross validation.

In the provided formula, where X represents the overall data, $X^{'}$ denotes the transformed data, μ denotes the mean, and σ denote standard deviation (S.D.)

The Pearson Correlation Coefficient (PCC), expressed by Eq. (21), serves to assess the linear dependence between two variables, be it disparate inputs or their connection with outputs. A PCC value of 1 or −1 indicates a substantial correlation, while 0 signifies no correlation. The absolute PCC value gauges the relative significance of features influencing output variables, especially in H_2 production. Notably, in this context, x and y denote the variables of interest for PCC determination, and n signifies the number of datapoints [36].

$$PCC = \frac{\sum_{i=1}^{n}(x_i - \overline{x})\sum_{i=1}^{n}(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2\sum_{i=1}^{n}(y_i - \overline{y})^2}} \tag{21}$$

where $\overline{x}$ and $\overline{y}$ are the two variables of interest for which the PCC is to be determined, with x_i and y_i representing their respective values for the i^{th} sample, and n denotes the total number of data points in the dataset.

2.4. Model development

The input data is partitioned into ten subgroups using 10-fold cross-validation, as illustrated in Fig. 5. The remaining subgroup is used as test data, while nine subgroups serve as training data for each round of the models' training. Subsequently, these ten datasets are consecutively fed into the ML models (SVM, KNN, XGB, RF, LGBM, and CB). Conducting the majority of the training over 10 iterations helps eliminate inaccurate model assessments stemming from the unintentional partition of the sample datasets.

The effectiveness and precision of each model were evaluated using statistical metrics. These metrics include the R^2 defined in Eq. (22), MAE specified in Eq. (23), NMAE shown in Eq. (24), RMSE presented in Eq. (25), and NRMSE indicated in Eq. (26) [37,38]. The NRMSE assesses the RMSE across the entire range of observed variables and is computed as the ratio of RMSE to the average of the observed values.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(y_{a,i} - y_{p,i}\right)^2}{\sum_{i=1}^{n}\left(y_{a,i} - \overline{y}_a\right)^2} \tag{22}$$

$$MAE = \frac{1}{n}\left(\sum_{i=1}^{n}\left|y_{a,i} - y_{p,i}\right|\right) \tag{23}$$

$$NMAE = \frac{\frac{1}{n}\sum_{i=1}^{n}\left|y_{a,i} - y_{p,i}\right|}{\max\left(y_{a,i}\right) - \min\left(y_{a,i}\right)} \tag{24}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}\left(y_{a,i} - y_{p,i}\right)^2}{n}} \tag{25}$$

$$NRMSE = \frac{RMSE}{\overline{y}_a} \tag{26}$$

In the provided formula, where $y_{a,i}$ represents the actual value, $y_{p,i}$ is the predicted value, $\overline{y}_a$ denotes the average value of the actual value, and n is equal to the number of the studied data.

2.5. Shapley additive explanations (SHAP)

The SHAP method constitutes an interpretative approach applicable to ML models, facilitating the assessment of feature importance. This method allocates importance values to each feature for a specific prediction, enabling users to comprehend the individual contributions of features to the ultimate output. The computation of the SHAP value for a given feature is expressed by Eq. (27) [39]:

$$\varnothing_i = \sum_{S\subseteq F\setminus\{i\}}\frac{n_s!(n_F - n_s - 1)!}{n_F!}\left[f_{s\cup\{i\}}\left(x_{s\cup\{i\}}\right) - f_s(x_s)\right] \tag{27}$$

F represents the set of all features, with S being a subset of F. n_s and n_p denote the number of features in F and S, respectively. The sign of the SHAP value indicates whether a feature has a positive or negative effect on the model output. The absolute SHAP value quantifies the magnitude of a specific feature's impact. By calculating the average absolute SHAP value across all samples, a comprehensive assessment of input feature importance can be achieved.

2.6. Selected machine learning algorithm

2.6.1. K-nearest neighbor (KNN)

The KNN algorithm, a method used for both regression and classification, categorizes sample data points by utilizing votes from their nearest neighbors, based on a pre-trained database [40]. It assigns the most frequent class for categorical output. The variable 'k', a small positive integer, denotes the number of considered neighbors, and for regression, averaging is applied. Increasing 'k' minimizes variance but introduces bias. Mathematically, classification KNN is expressed in Eq. (28). For KNN regression, the output is defined as the average of the k nearest values [41].

$$D_{(X,\ Y)} = \sqrt{\sum_{i=1}^{k}(x_i - y_i)^2} \tag{28}$$

2.6.2. Support vector machine (SVM)

The SVM is a supervised learning algorithm employed for both regression and classification problems [42]. SVM has found extensive application across various domains, owing to its numerous reliable learning characteristics and its ability to predict positive trial outcomes. The theoretical foundation supporting the SVM algorithm is elucidated in Eq. (29) [43].

$$T = \{(x_i y_i)|\ i = 1, 2, \ldots n\} \tag{29}$$

where n-dimensional feature vectors x_i and y_i are defined in the real number domain, with $x_i \in X$, and $y_i \in \{-1, +1\}$. When the dataset under analysis exhibits a linear relationship, Eqs. (30),(31) can be applied.

$$W^T X + b = 0 \tag{30}$$

$$W = (W_1; W_2; \ldots; W_d) \tag{31}$$

In the given equation, W represents the hyperplane, and b signifies the distance between the hyperplane and the origin. The formal representation for the distance from any point X to the hyperplane can be expressed as Eq. (32).

$$\gamma = \frac{|\omega^T X + b|}{||\omega||} \tag{32}$$

2.6.3. Random forest (RF)

The RF algorithm utilizes ensemble learning to generate multiple Decision Trees (DTs). When applied to a dataset, it organizes the data into a hierarchical tree structure. Each node in this structure undergoes further classification based on specific criteria, ultimately leading to the final output [44,45].The calculation is detailed in Eq. (33).

$$f(x) = \sum_{m=1}^{M}\frac{1}{m}f_m(x) \tag{33}$$

where, f_m represents the m^{th} tree in the set of all decision trees M.

2.6.4. eXtreme gradient boosting (XGB)

XGB is an efficient ML model that iteratively trains decision trees,

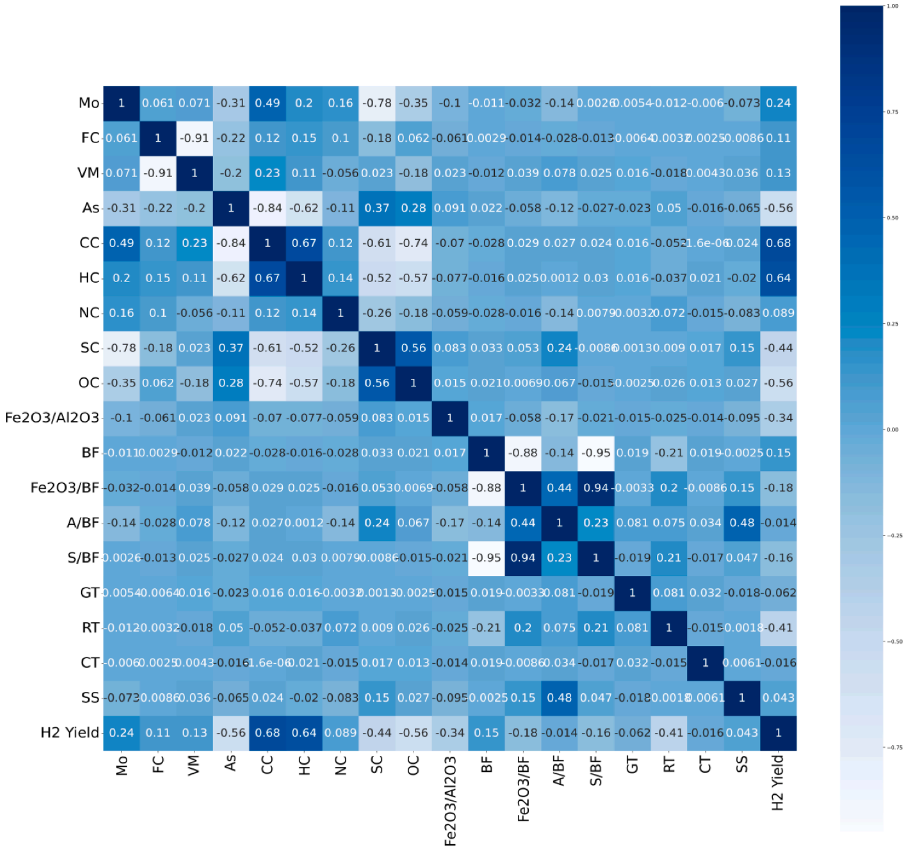


Fig. 6. Pearson correlation analysis matrix.

focusing on high-error samples. It improves accuracy by progressively adjusting the weight of each tree's contribution, as shown in Eq. (34) [46].

$$\widehat{y_i} = \sum_{k=1}^{K} f_k(x_i),\ f_k(x_i) \in F \tag{34}$$

where x_i represents the input value of the i^{th} sample, $\widehat{y_i}$ is the predicted value for the i^{th} sample, K denotes the total number of trees, f_k signifies a function within the set F of all functions, F represents the set space of all trees, and k indicates the k^{th} tree.

Eq. (35) formally expresses the objective function of XGBoost:

$$X = \sum_{i=1}^{n} l(y, \widehat{y}) + \sum_{k=1}^{K} \Omega(f_k) \tag{35}$$

Here, $l(y, \widehat{y})$ represents the error between the model's prediction and the actual value, while $\Omega(f_k)$ denotes the regularization term that governs the model's complexity

To reduce overfitting, a penalty term is added to the regularization function, as shown in in Eq. (36).

$$\Omega(f_k) = \gamma T + \lambda \frac{1}{2} \sum_{j=1}^{T} \omega_j^2 \tag{36}$$

λ is the control leaf node fraction, T the total leaf nodes, γ the control leaf node count, and ω_j is the j^{th} leaf node fraction.

2.6.5. Light gradient boosting machine (LGBM)

LGBM exhibits outstanding efficiency, accuracy, and speed, attributed to the integration of two unique data sampling and categorization approaches. This combination enhances the efficiency and accuracy of data scanning, sampling, grouping, and categorization when compared to similar methods [47]. In the LGBM algorithm, y_i is considered the target value, $\widehat{y_i}^{(t)}$ is the predicted value for the i^{th} sample at the current iteration t, and $\widehat{y_i}^{(t-1)}$ is the predicted value for the i^{th} sample from the previous iteration (t-1). The objective function of the LGBM model is articulated in Eq. (37) [48].

$$Obj^{(t)} = \sum_{i=1}^{n} \left(y_i, \widehat{y_i}^{(t)}\right) + \sum_{i=1}^{t} \Omega(f_i) \tag{37}$$

This objective function is further expanded using the Taylor formula, resulting in Eq. (38):

$$Obj^{(t)} = \sum_{i=1}^{n} \left[l\left(y_i, \widehat{y_i}^{(t-1)}\right) + g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i) \right] + \Omega(f_t) \tag{38}$$

where $g_i = \partial_{\widehat{y_i}^{\text{(t-1)}}} l(y_i, \widehat{y_i}^{\text{(t-1)}})$, $h_i = \partial^2_{\widehat{y_i}^{\text{(t-1)}}} l(y_i, \widehat{y_i}^{\text{(t-1)}})$. The accumulation of n samples is utilized to traverse all leaf nodes, yielding the final objective function of the LGBM model, as expressed in Eq. (39).

$$Obj^{(t)} = \sum_{j=1}^{s} \left[G_j w_j + \frac{1}{2}(H_j + \lambda) w_j^2 \right] \tag{39}$$

where S indicates the number of leaf nodes, and w signifies the leaf weight.

In this equation: $G_j = \sum_{i \in I_j} g_i$, and $H_j = \sum_{i \in I_j} h_i$ and I represents the sample set in leaf node j.

2.6.6. CatBoost (CB)

The CB algorithm is efficient, precise, and adept at handling categorical characteristics. It builds an ensemble of weak decision trees, integrating them through gradient boosting. This process involves adding extra trees to rectify errors made by earlier trees. For predicting continuous values, CB applies the formula specified in Eq. (40) [49].

Table 2
Optimized hyperparameters of six machine learning algorithms.

Models	Hyperparameters estimator	Boundary value	Optimization value
KNN	‘leaf_size’	[1,2,…,15]	3
	‘n_neighbors’	[1,2,…,10]	5
	‘p’	[1,1.4,…,2.6]	1.8
	‘weights’	[’uniform’, ’distance’]	distance
	‘algorithm’	[’auto’,’ball_tree’,’kd_tree’,’brute’]	kd_tree
SVM	‘C’	[1,1.2,…,2.8]	2.8
	‘kernal’	[’linear’]	linear
	‘gamma’	[’scale’,’auto’]	scale
RF	‘max_depth’	[1,2,…,15]	14
	‘n_estimators’	[1,2,…,30]	29
XGB	‘min_child_weight’	[1,2,…,5]	1
	‘gamma’	[1,2,3]	1
	‘subsample’	[0.2,0.4,…,1.0]	0.8
	‘colsample_bytree’	[0.6,0.8,1.0]	1
	‘n_estimators’	[500,1000]	1000
	‘max_depth’	[1,2,…,5]	4
LGBM	‘colsample_bytree’	[0.6,0.8,1.0]	0.8
	‘n_estimator’	[500,1000]	1000
	‘max_depth’	[1,2,3,4]	3
	‘num_leaves’	[2,3,4,5]	4
CB	‘depth’	[1,2,…,5]	4
	‘l2_leaf_reg’	[0.5,1.0,5.0]	0.5
	‘learning rate’	[0.001,0.01,0.004]	0.01
	‘min_child_samples’	[1,4,8,16,32]	8

$$y = f(x) = \sum_{i=1}^{n} \alpha i \; hi(x) \tag{40}$$

The variable y represents the predicted value, while x denotes the input features. The output function $f(x)$ is expressed as a linear combination of the basis functions $hi(x)$, where the coefficients αi determine the weight of each basis function in the linear combination.

3. Results and discussion

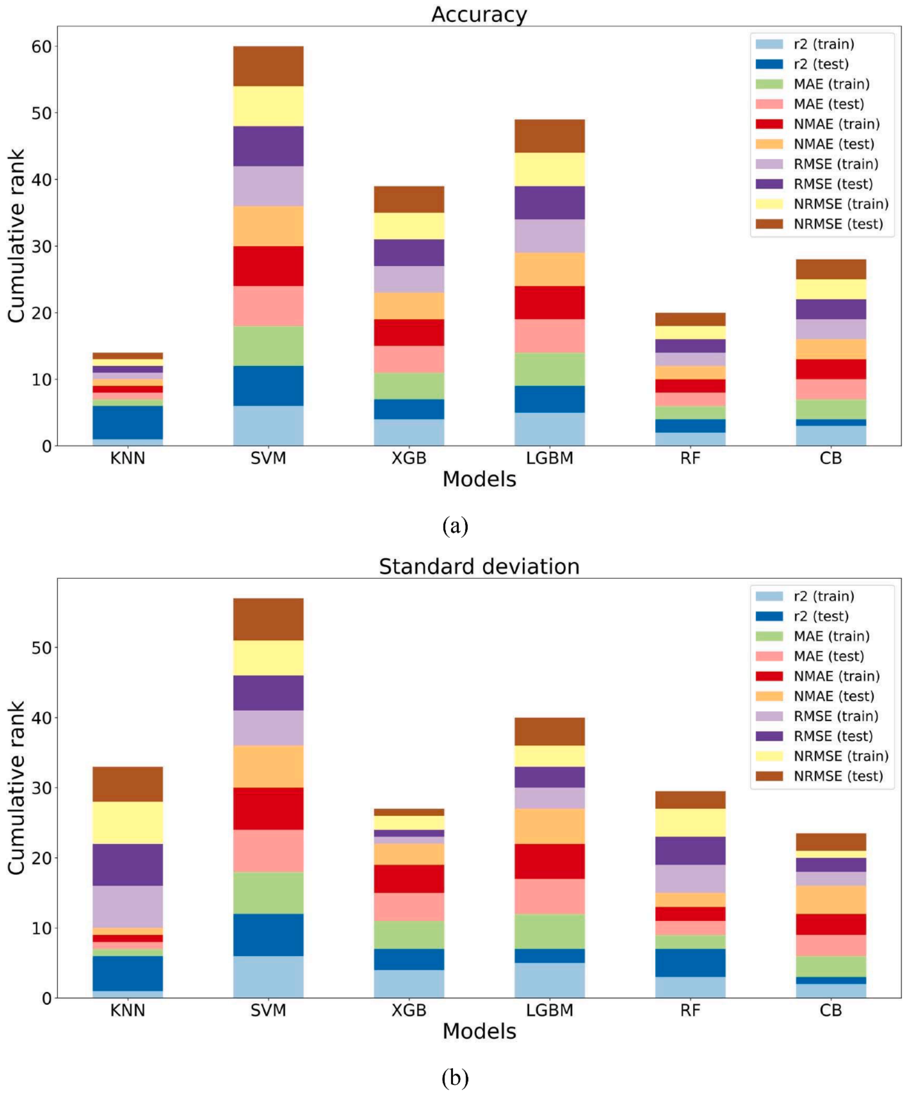
3.1. Pearson correlation coefficient (PCC) analysis

In order to investigate potential feature overlap and construct predictive models for estimating H_2 yield, the PCC was employed to assess the extent of collinearity among independent variables [50]. The presence of positive and negative signs serves to elucidate direct or inverse correlations between pairs of independent variables. Fig. 6 visually represents the degree of linear dependency among the 18 input variables.

Ideally, PCC values of 1 indicate a high degree of correlation in the data, while a PCC of 0 suggests no correlation between the variables. Generally, variables with a |PCC| greater than 0.6 are regarded as correlated entities [51]. The positive and negative signs delineate the nature of correlations between two independent variables. For instance, the positive correlation between S/BF feature and Fe_2O_3/BF feature is characterized by a |PCC| value of 0.94, whereas the negative correlation between the isosteric heat of FC feature and VM feature is indicated by a |PCC| value of 0.91.

Table 3
Performance metrics for train and test sets of ML algorithms.

Models	Data	Performance metrics for regression				
		R^2 (S.D.)	MAE (S.D.)	NMAE (S.D.)	RMSE (S.D.)	NRMSE (S.D.)
KNN	Train	0.99987 (0.00007)	0.0008 (0.00042)	0.00011 (0.00006)	0.02954 (0.01557)	0.00412 (0.00217)
	Test	0.86547 (0.01218)	0.0008 (0.00042)	0.00011 (0.00006)	0.02954 (0.01557)	0.00410 (0.00216)
SVM	Train	0.77800 (0.00234)	1.01704 (0.00445)	0.14162 (0.00056)	1.36986 (0.00847)	0.19076 (0.00103)
	Test	0.77358 (0.01931)	1.01704 (0.00445)	0.14168 (0.00316)	1.36986 (0.00847)	0.19083 (0.00454)
XGB	Train	0.98688 (0.00021)	0.25106 (0.00240)	0.03496 (0.00038)	0.33301 (0.00270)	0.04637 (0.00046)
	Test	0.97963 (0.00377)	0.25106 (0.00240)	0.03497 (0.00059)	0.33301 (0.00270)	0.04638 (0.00072)
LGBM	Train	0.95759 (0.00060)	0.46208 (0.00306)	0.06435 (0.00047)	0.59875 (0.00387)	0.08338 (0.00057)
	Test	0.95089 (0.00363)	0.46208 (0.00306)	0.06437 (0.00132)	0.59875 (0.00387)	0.08341 (0.00177)
RF	Train	0.99686 (0.00018)	0.10229 (0.00190)	0.01424 (0.00025)	0.16277 (0.00478)	0.02267 (0.00065)
	Test	0.98150 (0.00455)	0.10229 (0.00190)	0.01425 (0.00047)	0.16277 (0.00478)	0.02268 (0.00092)
CB	Train	0.99437 (0.00016)	0.15446 (0.00209)	0.02151 (0.00027)	0.21808 (0.00321)	0.03037 (0.00042)
	Test	0.98466 (0.00360)	0.15446 (0.00209)	0.02152 (0.00062)	0.21808 (0.00321)	0.03038 (0.00092)



(a)

(b)

Fig. 7. Model performance comparison: (a) Cumulative rank based on accuracy metrics; (b) Cumulative rank based on S.D. of metrics.

However, all other features have PCC < 0.6 and are not highly correlated. Elimination of any one variable may result in the sum of composition biomass percentages not being equal to 100%. The PCC analysis serves as a preliminary screening method for features within the raw dataset, aiding in the comprehension of overlapping input variables. Moreover, it helps mitigate resource requirements and prediction timing when executing the predictive models.

3.2. Accuracy evaluation of hydrogen yield estimation

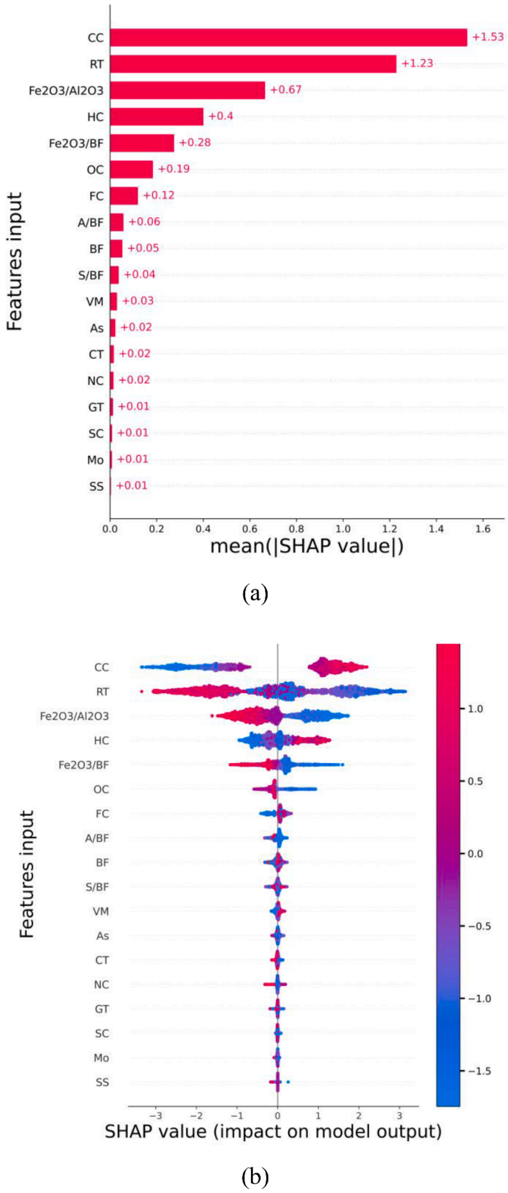
Six ML algorithms (SVM, KNN, XGB, RF, LGBM, CB) were customized and optimized with hyperparameters for H_2 yield prediction analysis in this study. The distinctive performances of each algorithm were evaluated to identify the model with optimal predictive accuracy [50]. Table 2 provides the optimized hyperparameters for each algorithm.

Table 3 presents the results pertaining to the predictive accuracy of diverse ML models for forecasting H_2 yield in BCLPro, encompassing SVM, KNN, XGB, RF, LGBM, and CB. Based on the experimental findings, it is discerned that within the training dataset, the KNN method exhibited the highest accuracy at 99.987% with a S.D. of 0.00007, implying superior robustness compared to alternative algorithms. Nevertheless, RF, CB, XGB, and LGBM methods also proved to be effective, boasting accuracies exceeding 95%. In the test dataset, the CB method demonstrated a predictive accuracy of 98.466%, while RF, XGB, and LGBM similarly exhibited commendable accuracies. Conversely, Support SVM methods demonstrated lower accuracy on such datasets, evident from the R^2 value in both training and testing phases, which only reached 77%. Consequently, in the process of selecting a model for further development as a predictive tool, BCLH2pro must undergo ranking, amalgamating the cumulative ranking for each model, with 1 representing the highest rank and 6 the lowest. This determination is based on the R^2, MAE, NMAE, RMSE, and NRMSE values illustrated in Fig. 7(a), as well as on the S.D. depicted in Fig. 7(b).

From Fig. 7(a) and (b), it is evident that the KNN, RF, and CB methods possess the lowest cumulative rankings, designating them as the most favorable models. Similarly, in terms of robustness, the CB method accumulates the least rank, succeeded by XGB and RF. However, in the decision-making process concerning model selection, consideration must be given to the model that aligns well with the dataset to avoid overfitting. Significant discrepancies in R^2 values between the training and test datasets, as observed in Table 3 for the KNN model, indicate potential overfitting. Consequently, it is deduced that the CB method exhibits a more favorable fit with the dataset than KNN. As a result, CB is selected for further development as a tool named BCLH2Pro.

3.3. Shapley additive explanations (SHAP)

The best-performing RF model was used in conjunction with the



(a)

(b)

Fig. 8. SHAP analysis of feature importance: (a) Mean SHAP values for each feature, (b) Distribution of SHAP values across features.

SHAP algorithm to evaluate the importance of input variables. As shown in Fig. 8(a), the factors that most significantly influence prediction accuracy are CC, RT, and the mass ratio of Fe_2O_3 to Al_2O_3. These results help prioritize the importance of biomass components in H_2 production. CC, HC, OC, and FC are the most critical variables, while SC and Mo in biomass have the least influence. Additionally, as illustrated in Fig. 8(b), a pronounced monotonic correlation exists between biomass characteristics and operational conditions, as indicated by SHAP values. In contrast, the relationships between SHAP values and other original experimental features exhibit greater complexity. This clear and concise progression of feature relationships constitutes a crucial element contributing to the improved training performance of ML models. The transparency in feature evolution streamlines the identification of shallow nonlinear relationships during the model training process, enabling a more comprehensive exploration of significant feature relationships [39].

3.4. Learning curve of the machine learning algorithms

Fig. 9 illustrates the learning curve of the six algorithms during the development of H_2 yield predictions for BCLpro. Notably, the accuracy of H_2 yield predictions was higher when applied to training datasets compared to test datasets. The increase in the number of training data points resulted in improved prediction accuracy. A smaller training dataset led to higher training error and lower R^2 values, indicating limited generalization capability and higher test error. As the volume of training data increased, H_2 yield prediction also improved, reducing bias errors. Remarkably, beyond 2,500 testing data points, the accuracy of H_2 yield prediction did not significantly increase from the 0.95 R^2 value for most algorithms, with the exception of SVM, which continued to show improvement.. Therefore, this study utilizes approximately 3,013 data sets from a total dataset exceeding 106,795 entries, ensuring accurate model construction and minimizing delays in computer processing [52].

3.5. Regression analysis plots for ML models

Fig. 10 illustrates a comparative analysis between predicted and experimental values, utilizing normalized data for H_2 yield through the application of six ML algorithms: KNN, SVM, RF, XGB, LGBM, and CB shown Fig. 10(a)–(f), respectively. The proximity of data points to regression lines serves as an indicator of enhanced predictability within the developed models. In the context of experimental values, the CB emerges as a superior model for predicting data across all ML algorithms due to its elevated accuracy and low error rates. It exhibits the highest R^2 values in both the training and testing datasets, amounting to 0.99437 and 0.98466, respectively. The black dotted line in the results symbolizes the point where predicted values from the ML model precisely align with the test values. Upon scrutinizing Fig. 10(c), (e), and (f), it becomes apparent that the scatter plots XGB, RF, and CB manifest a distribution pattern for the train and test datasets closely aligned with the black dotted line. These findings underscore the distribution characteristics of predictive data and the precision of the CB model, positioning it as a preferable choice among the ML algorithms due to its superior R^2 values and minimal error rates.

3.6. One-dimensional partial dependence plots (1D-PDPs)

PDPs in Fig. 11 illustrate the influence of the six most important input features (CC, RT, Fe_2O_3/Al_2O_3, HC, Fe_2O_3/BF, and OC) on the predicted H_2 yield obtained using the RF model. PDPs help explain the relationship between each input feature and the predicted H_2 yield, making it easier to understand how changes in each variable affect the H_2 yield.

When considering the relationship between biomass CC and H_2 yield, as shown in Fig. 11(a), a non-linear pattern is observed. The relationship can be divided into three stages: a gradual increase in H_2 yield within the CC range of 30-40%, a rapid increase from 40-45% due to the evaporation of volatiles and cellulose decomposition [53], and a stabilization when CC exceeds 45% [54]. This analysis indicates that the optimal CC range for H_2 production is 40-45%, providing valuable insights for biomass selection and process condition optimization.

RT is another crucial factor affecting H_2 yield, as depicted in Fig. 11 (b). The optimal temperature range for H_2 production is found to be 550-650°C, where the H_2 yield reaches its peak. At temperatures below 550°C, the H_2 yield is relatively low, possibly due to insufficient energy for efficient gasification reactions. Conversely, when temperatures exceed 650°C, the H_2 yield decreases significantly, suggesting that excessively high temperatures may adversely affect the production process.

The ratio between Fe_2O_3 and Al_2O_3 is another interesting variable, as shown in Fig. 11(c). The plot reveals that the H_2 yield tends to decrease continuously as the Fe_2O_3/Al_2O_3 ratio increases from 0.3 to 0.8, with the most pronounced decline occurring between 0.3 and 0.5. This

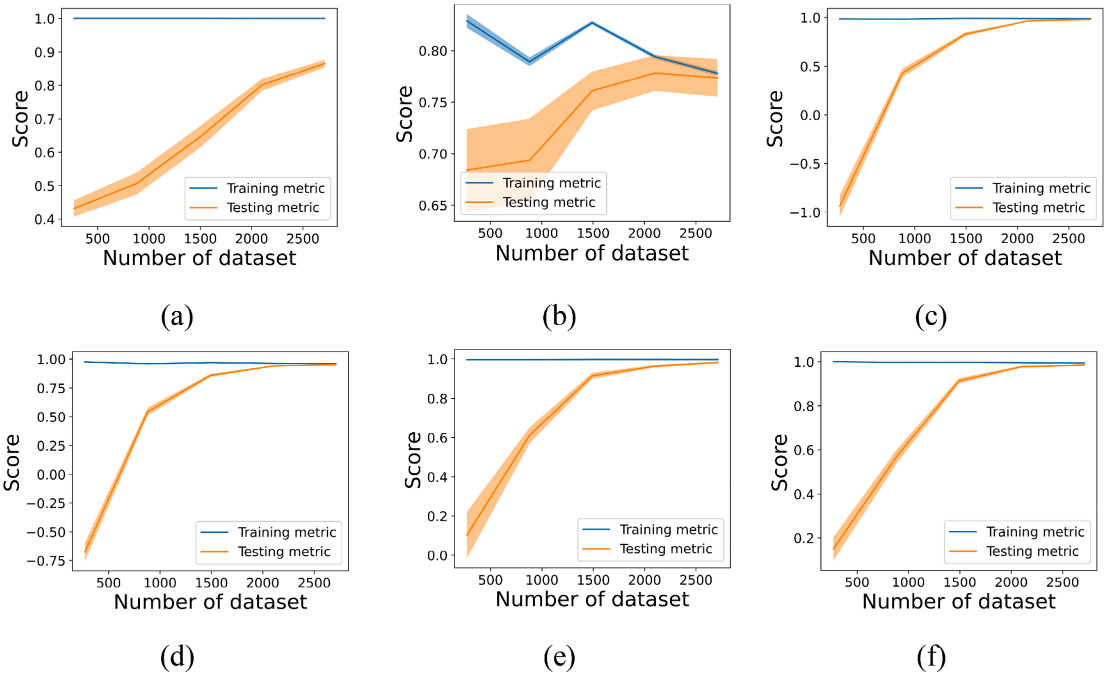


Fig. 9. Learning curves for six machine learning models: (a) KNN, (b) SVM, (c) XGB, (d) LGBM, (e) RF, and (f) CB.

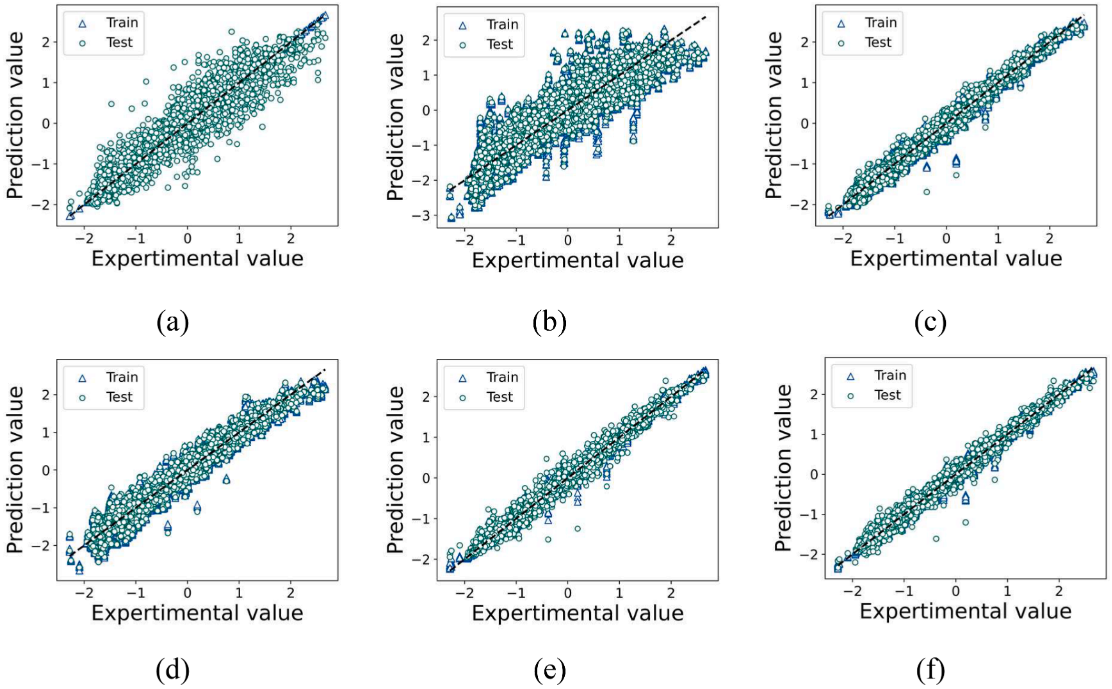
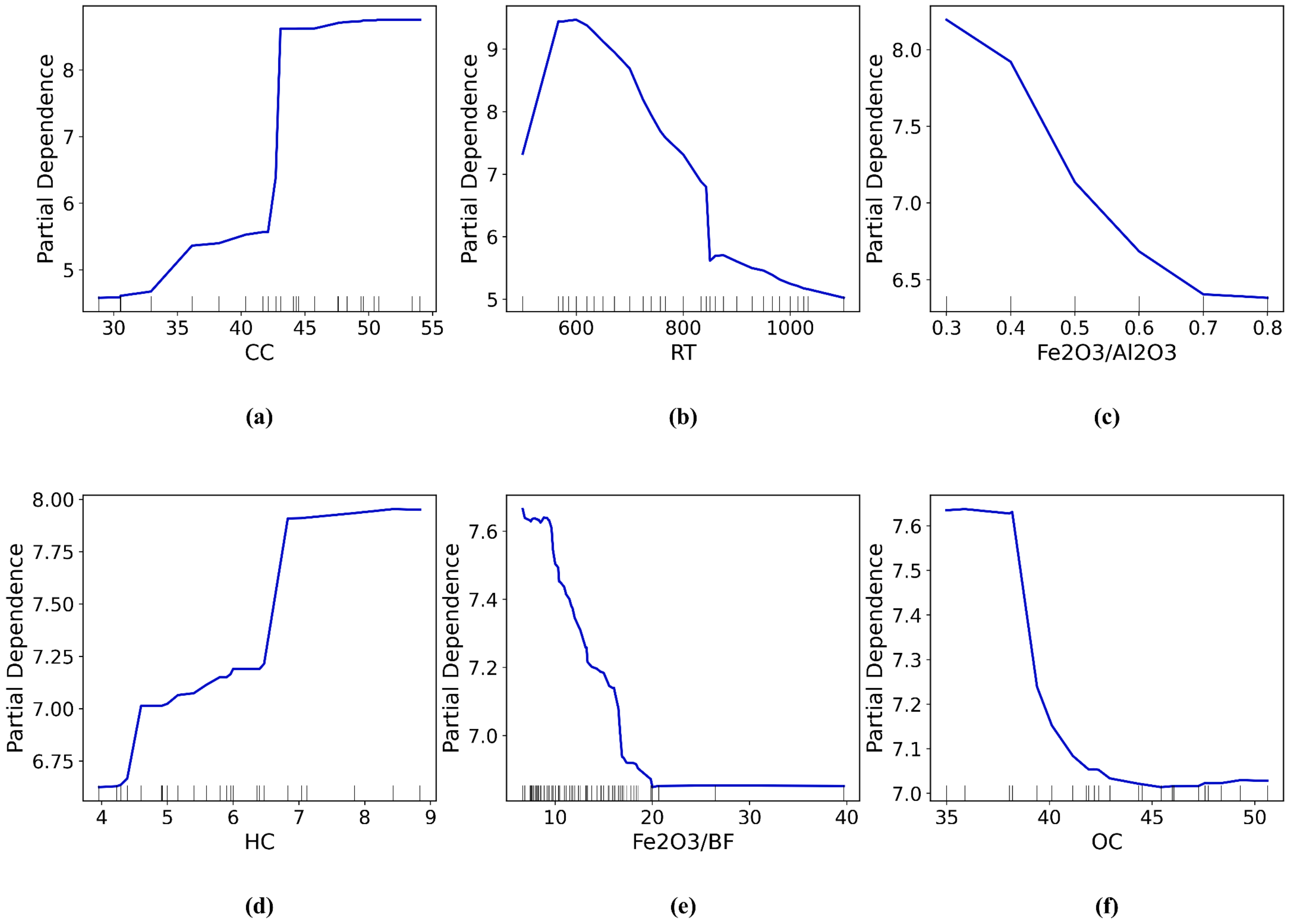


Fig. 10. Comparison of predicted vs. actual H_2 yield values for ML models: (a) KNN, (b) SVM, (c) XGB, (d) LGBM, (e) RF, and (f) CB.

phenomenon may be attributed to changes in oxygen transfer capacity and oxygen carrier stability as the proportion of Fe_2O_3 increases. It is observed that lower ratios (approximately 0.3-0.4) yield the highest H_2 production, possibly due to an optimal balance between oxygen transfer capability and material stability.

When analyzing the effect of initial HC in biomass on H_2 yield, as shown in Fig. 11(d), a complex three-stage relationship is observed. In the first stage (HC 4-5%), a gradual increase in H_2 yield is noted, attributed to moisture expulsion and hemicellulose decomposition, resulting in syngas formation and elevated H_2 levels. The second stage (HC 5-7%) exhibits a steeper slope with slight fluctuations, indicating intensified devolatilization and cellulose breakdown, leading to increased syngas production and higher H_2 yield. In the final stage (HC $> 7\%$), a rapid increase in H_2 yield is observed, followed by stabilization, likely due to lignin decomposition and fixed carbon formation [55].

The ratio between Fe_2O_3 and the BF also has a significant impact on H_2 yield, as illustrated in Fig. 11(e). As the ratio increases from 8 to approximately 20, a continuous and substantial decrease in H_2 yield is observed. At lower ratios ($Fe_2O_3/BF < 10$), insufficient oxygen leads to incomplete oxidation, promoting H_2 production. As the ratio increases (Fe_2O_3/BF 10-15), excess oxygen accelerates the oxidation of hydrogen to water and shifts the equilibrium towards CO_2 production (Eqs. (11)-



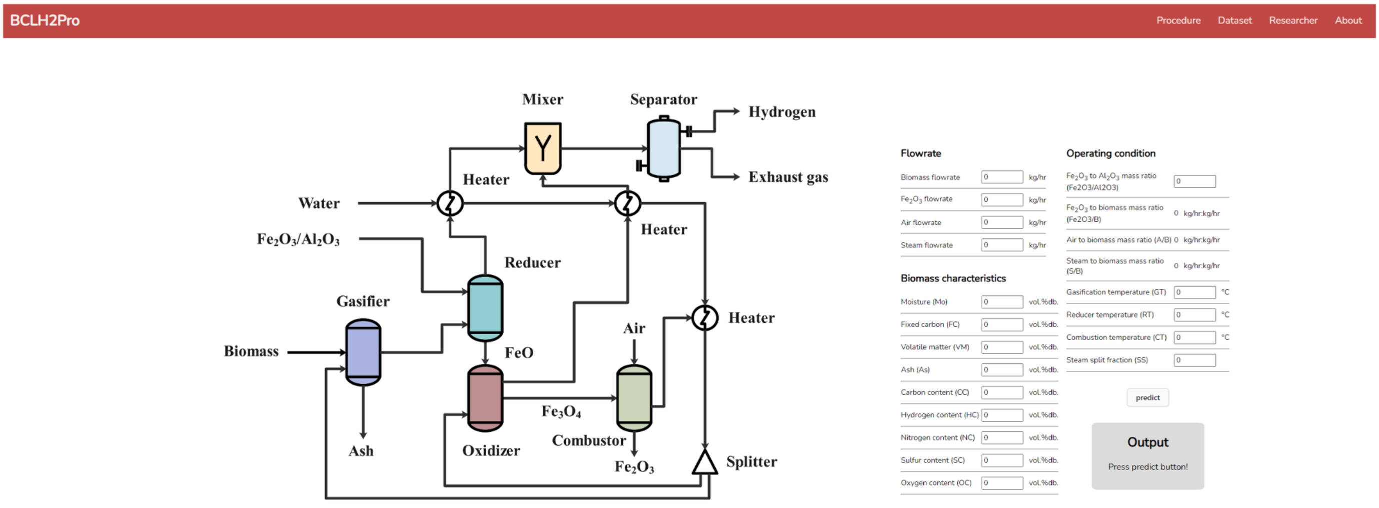
(a) (b) (c)

(d) (e) (f)

Fig. 11. 1D-PDPs of key features influencing H_2 yield predictions: (a) CC, (b) RT, (c) Fe_2O_3/Al_2O_3, (d) HC, (e) Fe_2O_3/BF, (f) OC.

(18)). H_2 yield stabilizes when the Fe_2O_3/BF ratio exceeds 20, potentially indicating the cessation of biomass decomposition or the presence of excess oxygen in the system.

Fig. 11(f) demonstrates the complex relationship between biomass OC and H_2 yield, which can be divided into two main stages. In the first stage (OC < 38%), the H_2 yield remains high and relatively constant, indicating optimal conditions for syngas synthesis and hydrogen release from the biomass structure. However, in the second stage (OC > 38%), a significant decrease in H_2 yield is observed as OC increases further. This may be due to enhanced complete combustion, increased water vapor formation, altered thermodynamic equilibrium favoring oxidation reactions, and over-oxidation of the oxygen carrier. Therefore, the data



Flowrate

Biomass flowrate 0 kg/hr

Fe_2O_3 flowrate 0 kg/hr

Air flowrate 0 kg/hr

Steam flowrate 0 kg/hr

Biomass characteristics

Moisture (Mo) 0 vol.%db.

Fixed carbon (FC) 0 vol.%db.

Volatile matter (VM) 0 vol.%db.

Ash (As) 0 vol.%db.

Carbon content (CC) 0 vol.%db.

Hydrogen content (HC) 0 vol.%db.

Nitrogen content (NC) 0 vol.%db.

Sulfur content (SC) 0 vol.%db.

Oxygen content (OC) 0 vol.%db.

Operating condition

Fe_2O_3 to Al_2O_3 mass ratio (Fe2O3/Al2O3) 0

Fe_2O_3 to biomass mass ratio (Fe2O3/B) 0 kg/hr:kg/hr

Air to biomass mass ratio (A/B) 0 kg/hr:kg/hr

Steam to biomass mass ratio (S/B) 0 kg/hr:kg/hr

Gasification temperature (GT) 0 °C

Reducer temperature (RT) 0 °C

Combustion temperature (CT) 0 °C

Steam split fraction (SS) 0

Press predict button!

Fig. 12. Screenshots captured from the BCLH2pro web server interface.

suggest that the optimal OC range for maximizing H_2 yield is approximately 36-38%, and beyond this point, production efficiency decreases markedly.

3.7. BCLH2Pro web server

To enhance accessibility and promote collaborative engagement within the scientific community, the BCLH2Pro web server has been designed and implemented. This user-friendly platform is openly accessible online at http://bclh2pro.pythonanywhere.com/. Detailed, step-by-step guidelines have been meticulously provided to facilitate users in effectively utilizing the BCLH2Pro web server for obtaining specific and desired research outcomes. The unrestricted availability of this resource contributes to the seamless integration of advanced computational tools, fostering a conducive environment for scientific exploration and discovery. The user interface of BCLH2Pro presents the primary interface and showcases the dataset employed for training, as illustrated in Fig. 12. Users can follow these steps to utilize the tool:

1. Access the web server and input the biomass data and operational parameters in the designated fields.
2. Select the "Output" button to initiate the process and obtain the H_2 yield value.

4. Conclusions

The integration of advanced ML algorithms and the user-friendly BCLH2Pro web server marks a significant improvement in BCLpro. Aspen Plus generated a comprehensive dataset, which was strategically reduced for efficient analysis. Learning curve analysis confirmed high prediction accuracy despite the data reduction. The CB algorithm, identified through meticulous analysis, demonstrates superior predictive power, achieving up to 98% accuracy in H_2 yield predictions. SHAP analysis identified CC, RT, and Fe_2O_3/Al_2O_3 ratio as the most influential factors for H_2 production. Optimal conditions for high H_2 yield were determined to be 40-45% vol.%db., 550-600°C, and 0.3-0.4, respectively. BCLH2Pro has emerged as a valuable resource for the research community, aiding in biomass selection and operational setup in BCLPro. This application is expected to minimize trial-and-error experimentation, thereby reducing laboratory costs across various industries. However, the study's reliance on simulated data presents limitations. Future research should focus on validating these findings with experimental data and expanding the scope to include more diverse biomass types and operational conditions. Despite these constraints, this study provides a robust foundation for optimizing BCLpro systems and for advancing efficient, sustainable H_2 production from biomass resources.

CRediT authorship contribution statement

Thanadol Tuntiwongwat: Writing – original draft, Resources, Methodology, Formal analysis, Data curation, Investigation, Validation. **Sippawit Thammawiset:** Validation, Software. **Thongchai Rohitatisha Srinophakun:** Writing – review & editing, Project administration, Software, Supervision. **Chawalit Ngamcharussrivichai:** Writing – review & editing, Project administration. **Somboon Sukpancharoen:** Writing – review & editing, Writing – original draft, Software, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization, Project administration, Validation, Visualization, Supervision.

Declaration of competing interest

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

Data availability

Data will be made available on request.

Acknowledgement

This research project is supported by Research and Graduate Studies, Khon Kaen University.

References

[1] Khan KA, Quamar MM, Al-Qahtani FH, Asif M, Alqahtani M, Khalid M. Smart grid infrastructure and renewable energy deployment: a conceptual review of Saudi Arabia. Energy Strategy Rev 2023;50:101247.
[2] Tun ZM, Christwardana M, Adiguna R, Hadiyanto H, Mini Windarta JA. Review on the biomass energy implementation from economic perspective in Indonesia. J Bioresourc Environ Sci 2023;2(1):1–8.
[3] Osman AI, Chen L, Yang M, Msigwa G, Farghali M, Fawzy S, Rooney DW, Yap PS. Cost, environmental impact, and resilience of renewable energy under a changing climate: a review. Environ Chem Lett 2023;21(2):741–64.
[4] Kumar S., Rathore K. Renewable energy for sustainable development goal of clean and affordable energy. Int J Mater Manuf Sustain Technol 2(1):115.
[5] Lanjekar PR, Panwar NL. Agrawal C. A comprehensive review on hydrogen production through thermochemical conversion of biomass for energy security. Bioresour Technol Rep 2023;21:101293.
[6] Jaroenkhasemmeesuk C, Tippayawong N, Shimpalee S, Ingham DB, Pourkashanian M. Improved simulation of lignocellulosic biomass pyrolysis plant using chemical kinetics in Aspen Plus® and comparison with experiments. Alexandria Eng J 2023;63:199–209.
[7] Sezer S, Kartal F, Özveren U. Prediction of chemical exergy of syngas from downdraft gasifier by means of machine learning. Thermal Sci Eng Progress 2021; 26:101031.
[8] Junsittiwate R, Srinophakun TR, Sukpancharoen S. Techno-economic, environmental, and heat integration of palm empty fruit bunch upgrading for power generation. Energy Sustain Dev 2022;66:140–50.
[9] Zhou X, Jin H, Li N, Ma X, Ma Z, Lu P, Yao X, Chen S. New process combining Fe-based chemical looping and biomass pyrolysis for cogeneration of hydrogen, biochar, bio-oil and electricity with in-suit CO2 separation. Molecules 2023;28(6): 2793.
[10] Zare AA, Yari M, Nami H, Mohammadkhani F. Low-carbon hydrogen, power and heat production based on steam methane reforming and chemical looping combustion. Energy Convers Manag 2023;279:116752.
[11] Campari A, Ustolin F, Alvaro A, Paltrinieri N. A review on hydrogen embrittlement and risk-based inspection of hydrogen technologies. Int J Hydrogen Energy 2023.
[12] Dawood F, Anda M, Shafiullah GM. Hydrogen production for energy: an overview. Int J Hydrogen Energy 2020 Feb;45(7):3847–69.
[13] Argyris PA, De Leeuwe C, Abbas SZ, Amieiro A, Poultson S, Wails D, Spallina V. Chemical looping reforming for syngas generation at real process conditions in packed bed reactors: an experimental demonstration. Chem Eng J 2022;435: 134883.
[14] Goel A, Moghaddam EM, Liu W, He C, Konttinen J. Biomass chemical looping gasification for high-quality syngas: a critical review and technological outlooks. Energy Convers Manag 2022;268:116020.
[15] Donat F, Kierzkowska A, Muller CR. Chemical looping partial oxidation of methane: reducing carbon deposition through alloying. Energy Fuels 2022;36(17): 9780–4.
[16] Miyahira K, Aziz M. Hydrogen and ammonia production from low-grade agricultural waste adopting chemical looping process. J Clean Prod 2022;372: 133827.
[17] Liu G, Sun Z, Zhao H, Mao X, Yang B, Shang J, Wu Z. Chemical looping reforming of toluene as bio-oil model compound via NiFe2O4@ SBA-15 for hydrogen-rich syngas production. Biomass Bioenergy 2023;174:106851.
[18] Zhang W, Chen Q, Chen J, Xu D, Zhan H, Peng H, Pan J, Vlaskin M, Leng L, Li H. Machine learning for hydrothermal treatment of biomass: a review. Bioresour Technol 2023;370:128547.
[19] Thebelt A, Wiebe J, Kronqvist J, Tsay C, Misener R. Maximizing information from chemical engineering data sets: applications to machine learning. Chem Eng Sci 2022;252:117469.
[20] Yüksel N, Börklü HR, Sezer HK, Canyurt OE. Review of artificial intelligence applications in engineering design perspective. Eng Appl Artif Intell 2023;118: 105697.
[21] Khaleel M, Ahmed AA, Alsharif A. Artificial Intelligence in Engineering. Brilliance: Res Artif Intell 2023;3(1):32–42.
[22] Moein MM, Saradar A, Rahmati K, Mousavinejad SH, Bristow J, Aramali V, Karakouzian M. Predictive models for concrete properties using machine learning and deep learning approaches: a review. J Build Eng 2023;63:105444.
[23] Asghar Z, Hafeez K, Sabir D, Ijaz B, Bukhari SS, Ro JS. RECLAIM: renewable energy based demand-side management using machine learning models. IEEE Access 2023;11:3846–57.

[24] Umenweke GC, Afolabi IC, Epelle EI, Okolie JA. Machine learning methods for modeling conventional and hydrothermal gasification of waste biomass: a review. Bioresour Technol Rep 2022;17:100976.

[25] Tangsriwong K, Lapchit P, Kittijungjit T, Klamrassamee T, Sukjai Y, Laoonual Y. Modeling of chemical processes using commercial and open-source software: a comparison between Aspen Plus and DWSIM. In: InIOP Conference Series: Earth and Environmental Science. 463. IOP Publishing; 2020, 012057.

[26] Catalanotti E, Porter RT, Mahgerefteh H. An aspen plus kinetic model for the gasification of biomass in a Downdraft Gasifier. Chem Eng Trans 2022;92:679–84.

[27] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Scikit-learn Vanderplas J. Machine learning in Python. J Mach Learn Res 2011;12:2825–30.

[28] Cortazar M, Santamaria L, Lopez G, Alvarez J, Zhang L, Wang R, Bi X, Olazar M. A comprehensive review of primary strategies for tar removal in biomass gasification. Energy Convers Manag 2023;276:116496.

[29] Marcantonio V, De Falco M, Capocelli M, Bocci E, Colantoni A, Villarini M. Process analysis of hydrogen production from biomass gasification in fluidized bed reactor with different separation systems. Int J Hydrogen Energy 2019;44(21):10350–60.

[30] Lahafdoozian M, Khoshkroudmansouri H, Zein SH, Jalil AA. Hydrogen production from plastic waste: a comprehensive simulation and machine learning study. Int J Hydrogen Energy 2024;59:465–79.

[31] Kartal F, Özveren U. A deep learning approach for prediction of syngas lower heating value from CFB gasifier in Aspen plus®. Energy 2020;209:118457.

[32] Singh D, Singh B. Investigating the impact of data normalization on classification performance. Appl Soft Comput 2020;97:105524.

[33] Singh P, Bardhan A, Han F, Samui P, Zhang W. A critical review of conventional and soft computing methods for slope stability analysis. Model Earth Syst Environ 2023;9(1):1–7.

[34] Jo JM. Effectiveness of normalization pre-processing of big data to the machine learning performance. J Korea Inst Electronic Commun Sci 2019;14(3):547–52.

[35] Umar MA, Zhanfang C. Effects of feature selection and normalization on network intrusion detection. Authorea Preprints; 2023.

[36] Li Y, Gupta R, You S. Machine learning assisted prediction of biochar yield and composition via pyrolysis of biomass. Bioresour Technol 2022;359:127511.

[37] Sumayli A. Development of advanced machine learning models for optimization of methyl ester biofuel production from papaya oil: gaussian process regression (GPR), multilayer perceptron (MLP), and K-nearest neighbor (KNN) regression models. Arab J Chem 2023;16(7):104833.

[38] Sudharshan K, Naveen C, Vishnuram P, Krishna Rao Kasagani DV, Nastasi B. Systematic review on impact of different irradiance forecasting techniques for solar energy prediction. Energies (Basel) 2022;15(17):6267.

[39] Feng S, Sun X, Chen G, Wu H, Chen X. LBE corrosion fatigue life prediction of T91 steel and 316 SS using machine learning method assisted by symbol regression. Int J Fatigue 2023;177:107962.

[40] Ye NY, Saparbaev RK, Omonov II. A brief review of machine learning algorithms. O`zbekistonda Fanlararo Innovatsiyalar Va Ilmiy Tadqiqotlar Jurnali 2023;2(15): 411–7.

[41] Ghunimat D, Alzoubi AE, Alzboon A, Hanandeh S. Prediction of concrete compressive strength with GGBFS and fly ash using multilayer perceptron algorithm, random forest regression and k-nearest neighbor regression. Asian J Civil Eng 2023 Jan;24(1):169–77.

[42] Manatura K, Chalermsinsuwan B, Kaewtrakulchai N, Kwon EE, Chen WH. Machine learning and statistical analysis for biomass torrefaction: a review. Bioresour Technol 2023;369:128504.

[43] Çakir M, Yilmaz M, Oral MA, Kazanci HÖ, Oral O. Accuracy assessment of RFerns, NB, SVM, and kNN machine learning classifiers in aquaculture. J King Saud Univ-Sci 2023;35(6):102754.

[44] Ali SW, Asif M, Zia MY, Rashid M, Syed SA, Nava E. CDSS for early recognition of respiratory diseases based on ai techniques: a systematic review. Wirel Pers Commun 2023:1–23.

[45] Tapeh AT, Naser MZ. Artificial intelligence, machine learning, and deep learning in structural engineering: a scientometrics review of trends and best practices. Arch Comput Methods Eng 2023;30(1):115–59.

[46] Zhou B, Chen X, Li G, Gu P, Huang J, Yang B. Xgboost–sfs and double nested stacking ensemble model for photovoltaic power forecasting under variable weather conditions. Sustainability 2023;15(17):13146.

[47] Ahmad A, Yadav AK, Singh A. Application of machine learning and genetic algorithms to the prediction and optimization of biodiesel yield from waste cooking oil. Korean J Chem Eng 2023;40(12):2941–56.

[48] Chen H, Li X, Feng Z, Wang L, Qin Y, Skibniewski MJ, Chen ZS, Liu Y. Shield attitude prediction based on Bayesian-LGBM machine learning. Inf Sci (Ny) 2023; 632:105–29.

[49] Kumar V, Kedam N, Sharma KV, Mehta DJ, Caloiero T. Advanced machine learning techniques to improve hydrological prediction: a comparative analysis of streamflow prediction models. Water (Basel) 2023;15(14):2572.

[50] Sukpancharoen S, Katongtung T, Rattanachoung N, Tippayawong N. Unlocking the potential of transesterification catalysts for biodiesel production through machine learning approach. Bioresour Technol 2023;378:128961.

[51] Hai A, Bharath G, Patah MF, Daud WM, Rambabu K, Show P, Banat F. Machine learning models for the prediction of total yield and specific surface area of biochar derived from agricultural biomass by pyrolysis. Environ Technol Innov 2023;30: 103071.

[52] Wongchai W, Onsree T, Sukkam N, Promwungkwa A, Tippayawong N. Machine learning models for estimating above ground biomass of fast growing trees. Expert Syst Appl 2022;199:117186.

[53] Yan X, Hu J, Zhang Q, Zhao S, Dang J, Wang W. Chemical-looping gasification of corn straw with Fe-based oxygen carrier: thermogravimetric analysis. Bioresour Technol 2020;303:122904.

[54] Onsree T, Tippayawong N, Zheng A, Li H. Pyrolysis behavior and kinetics of corn residue pellets and eucalyptus wood chips in a macro thermogravimetric analyzer. Case Stud Thermal Eng 2018;12:546–56.

[55] Rasaq WA, Golonka M, Scholz M, Białowiec A. Opportunities and challenges of high-pressure fast pyrolysis of biomass: a review. Energies (Basel) 2021;14(17): 5426.