

Machine learning optimization for enhanced biomass-coal co-gasification

Junting Pan^a, Hossein Shahbeik^b, Alireza Shafizadeh^c, Shahin Rafiee^c, Milad Golvirdizadeh^c, Seyyed Alireza Ghafarian Nia^c, Hossein Mobli^c, Yadong Yang^a, Guilong Zhang^d^{**}, Meisam Tabatabaei^b^{***}, Mortaza Aghbashlo^c^{*}

^a State Key Laboratory of Efficient Utilization of Arid and Semi-arid Arable Land in Northern China, Institute of Agricultural Resources and Regional Planning, Chinese Academy of Agricultural Sciences, Beijing, 100081, China

^b Higher Institution Centre of Excellence (HICoE), Institute of Tropical Aquaculture and Fisheries (AKUATROP), Universiti Malaysia Terengganu, 21030 Kuala Nerus, Terengganu, Malaysia

^c Department of Mechanical Engineering of Agricultural Machinery, Faculty of Agricultural Engineering and Technology, College of Agriculture and Natural Resources, University of Tehran, Karaj, Iran

^d Agro-Environmental Protection Institute, Ministry of Agriculture and Rural Affairs, Tianjin, 300191, China

ARTICLE INFO

Keywords:

Co-gasification
Biomass
Coal
Machine learning
Syngas
Hydrogen production

ABSTRACT

The co-gasification of biomass feedstocks with coal offers a promising approach to enhancing syngas quality while mitigating the environmental impacts of traditional coal gasification. However, experimental determination of the optimal biomass/coal blending ratio and operational parameters is often resource-intensive. To address this challenge, modeling techniques are invaluable for optimizing biomass-coal co-gasification. This study aims to develop a machine learning (ML) model to optimize biomass-coal co-gasification. Additionally, an evolutionary algorithm is employed for multi-objective optimization, targeting maximum H₂ production and optimal performance for the Fischer-Tropsch process. A comprehensive dataset from reputable literature sources, covering a wide range of biomass/coal blending ratios under various process conditions, was compiled. The dataset underwent statistical analysis, and mechanistic discussions were included to elucidate the effects of each parameter on the process. Among the four ML models applied, gradient boosting regression demonstrated the best performance during the testing phase, achieving an R² exceeding 0.92 and MAE and RMSE values lower than 2.92 and 3.39, respectively. For H₂ production, optimal results were observed with steam yields and temperatures near 1480 °C, while air and temperatures around 1570 °C yielded the best outcomes for the Fischer-Tropsch process. A biomass/coal blending ratio between 50 % and 70 % was found to be suitable for almost all gasifying agents under both criteria. The process was also analyzed techno-economically based on optimal conditions, revealing that steam exhibits superior techno-economic performance compared to other gasifying agents.

1. Introduction

A key component of sustainable development is the availability of electricity that provides basic services and improves the quality of life for poor urban and rural communities [1]. Approximately 63 % of the world's electricity comes from fossil fuels, while approximately 30 % comes from low-carbon energy sources [2]. Coal is the preferred fuel for electricity generation worldwide due to its widespread availability and cost-effectiveness [3]. More specifically, coal is ranked as the

second-largest energy source and the largest source of electricity [4]. Nevertheless, most power plants burn coal in pulverized form, which limits their overall efficiency [5]. In addition, coal combustion produces a lot of pollutants, such as particulate matter (PM), nitrogen oxides (NOx), sulfur oxides (SOx), and carbon oxides (COx) [6]. Therefore, the need for coal power generation that is more efficient and cleaner is growing.

Thermochemical processes, such as gasification, are favored over combustion in coal-to-energy conversion due to their numerous

* Corresponding author.

** Corresponding author.

*** Corresponding author.

E-mail addresses: zhangguilong@caas.cn (G. Zhang), meisam.tabatabaei@umt.edu.my (M. Tabatabaei), maghbashlo@ut.ac.ir, aghbashlomortaza@gmail.com (M. Aghbashlo).

Table 1

Summary of the features of ML models used to model biomass-coal co-gasification.

Feedstock(s)	Inputs	Outputs	ML model(s)	Ref.
Coal and biomass	Carbon, oxygen, hydrogen, fixed carbon, moisture content, volatile matter (for both biomass and coal), temperature, ash	Syngas yield, syngas lower heating value (LHV)	Artificial neural network, multiple linear regression, support vector machine, decision tree	[16]
High-ash coal and biomass	Gasifier temperature, coal and biomass feed rates, fixed carbon, volatile matter, ash, moisture, air/fuel ratio, steam/fuel ratio, carbon, hydrogen, oxygen (for both biomass and coal), reaction rate constant	Syngas yield, carbon conversion efficiency, syngas heating value, cold gas efficiency	Genetic programming, multilayer perceptron, support vector regression	[17]
Coal and sawdust	Temperature, coal blending ratio, equivalence ratio	Hydrogen production	Support vector machine regression, Gaussian process regression, and artificial neural network	[18]
Coal and municipal solid wastes	Carbon, oxygen, hydrogen, nitrogen, fixed carbon, ash content, volatile matter, temperature, blending ratio, steam/fuel ratio, equivalent ratio	Syngas composition, H ₂ /CO ratio, syngas LHV, tar yield, carbon conversion efficiency	Extreme gradient boosting regression, gradient boosting regression, random forest regression, histogram-based gradient boosting regression	[19]

advantages [4]. Unlike combustion, gasification involves partial oxidation, resulting in the generation of hydrogen and carbon monoxide, along with byproducts like carbon dioxide, water vapor, and methane [6]. Gasification technologies typically employ dry reactors with conventional or microwave heat sources, while in certain cases, wet reactors (hydrothermal gasification) are utilized for gasifying the feedstocks [7, 8]. In both techniques, the produced gas (syngas) can be easily used in gas turbines and combined-cycle power plants for electricity generation and in chemical processing systems for liquid fuel and chemical synthesis [5]. Gasification outperforms combustion in coal-to-energy conversion in terms of thermal efficiency and environmental impact [6]. The promising features of coal gasification are expected to drive its thermal power output to reach 350 MWth [9]. Nevertheless, gasification faces some challenges despite being mature [3,6]. These challenges include low carbon conversion, low feedstock reactivity, low-quality syngas, high greenhouse gas emissions, high NO_x/SO_x emissions, and tar formation [6,9].

In gasification, biomass can replace coal as an alternative renewable energy source while reducing reliance on fossil fuels [3]. Biomass fuel is made from biological sources such as forest residues, agricultural residues, animal wastes, and municipal organic wastes. Compared to fossil fuels, biomass feedstocks have several environmental benefits because of their lower carbon footprint and lower pollutant gas emissions (i.e., SO_x and NO_x) [10]. Although biomass is a promising gasification feedstock, it poses some challenges in gasification systems due to its low bulk density, low energy density, high tar formation potential, and high transport, handling, and storage costs [4,11]. Furthermore, biomass feedstocks are heterogeneous, causing greater fluctuations in their composition and quality and more complexity in their conversion processes [3]. However, coal and biomass co-gasification presents a promising solution to the challenges associated with their separate gasification [12]. Through this process, coal can be partially substituted with renewable biomass, thereby reducing dependency on fossil fuels [6].

The co-gasification of biomass and coal offers several environmental benefits, including a reduced carbon footprint, suppressed SO_x and NO_x emissions, and improved syngas quality [5,13]. In this process, the interaction between biomass and coal leads to synergistic effects of varying forms and intensities [9]. Alkaline and alkaline earth metals, present in higher concentrations in biomass feedstocks, can act as cost-free catalysts for coal gasification, thereby reducing industrial catalyst costs [5,9]. Additionally, the higher oxygen content in biomass compared to coal enhances its reactivity during gasification [6]. The high temperatures during co-gasification are advantageous for decomposing biomass-derived tar into useful gases [12,14]. While biomass gasification is typically feasible only on a small scale, co-gasification can address this limitation [12]. The issue of seasonality can also be mitigated, as plants can operate on coal alone when biomass feedstocks are unavailable [12]. Consequently, the biomass and coal co-gasification process has great potential for achieving technically feasible, economically viable, and environmentally clean co-conversion of these

feedstocks [9].

Modeling and optimizing biomass and coal co-gasification pose significant challenges due to the multitude of variables influencing process performance. These variables include temperature, type of gasifying agents and their concentration, biomass/coal blending ratio, and the chemical properties of both biomass and coal [6]. The optimum operational parameters and biomass/coal blending ratio for a given coal and biomass are typically determined by various experiments. The procedure should be repeated for other biomass feedstocks and coal types with different chemical properties. Much time and resources are required to conduct biomass and coal co-gasification experiments. It is not always possible to model and optimize the co-gasification process based on experimental measurements. The abovementioned issues can be addressed using machine learning (ML) technology based on abundant experimental data. This data-driven tool enables insight discovery and prediction about target properties quickly [15]. Many advancing ML capabilities have been demonstrated, such as low experimental costs, low time constraints, and high generalization abilities.

Some studies have explored the use of ML for modeling the biomass-coal co-gasification process, owing to its capability to handle the complexities inherent in such systems. Table 1 provides a summary of the features of ML approaches employed in modeling the biomass-coal co-gasification process. ML techniques have consistently demonstrated effectiveness in analyzing biomass-coal co-gasification, revealing hidden trends and patterns. Using these models could significantly reduce the time and effort required to identify optimal gasification conditions for biomass-coal co-gasification, potentially minimizing the need for chemically intensive and costly experiments. However, existing models often overlook critical parameters, such as biomass nitrogen and sulfur content, biomass/coal blending ratio, equivalent ratio, and gasifying agent type. Furthermore, existing models may predict limited outputs and lack comprehensive feature importance analysis and multi-objective optimization. Thus, this research presents an ML-driven method to model and optimize biomass-coal co-gasification, with the goal of overcoming the recognized shortcomings in published literature.

The study begins with a comprehensive review and screening of published literature to gather the necessary data for developing ML models. Various statistical techniques are employed to analyze the collected data, accompanied by mechanistic discussions to elucidate their relationships. Subsequently, four ML models—Kernel Ridge Regression (KRR), Decision Tree Regression (DTR), Random Forest Regression (RFR), and Gradient Boosting Regression (GBR)—are trained to model the composition (H₂, CO, CO₂, and CH₄) and heating value of syngas based on coal and biomass composition and gasification conditions. Feature importance analysis is conducted to ascertain the influence of input variables on output responses. After selecting the best-performing model, a heuristic algorithm is employed to identify the optimal coal and biomass composition and gasification conditions. These optimal conditions are then subjected to techno-economic analysis to assess their economic viability and potential for commercialization.

Table 2

Optimization objectives targeting the adaptation of the biomass-coal co-gasification process according to two distinct criteria.

Item	H ₂ production	Degree of importance	Fischer-Tropsch process	Degree of importance
Syngas yield	Maximize	4	Maximize	1
H ₂	Maximize	2	Maximize	4
CO ₂	Maximize	1	Minimize	1
CO	Minimize	1	In range (50 % of H ₂)	1
CH ₄	Minimize	1	Minimize	1
Syngas LHV	Maximize	1	Not important	1

This study presents a novel approach that leverages ML for optimizing syngas production in biomass-coal co-gasification. By integrating statistical analyses, ML models, and heuristic algorithms, the research goes beyond traditional methods to identify the most economically viable and environmentally friendly configurations for this process. This innovation paves the way for the development of commercially attractive biomass-coal co-gasification technologies. The novelty of this study lies in its comprehensive application of ML and heuristic optimization to the biomass-coal co-gasification process, which is not commonly seen in existing literature. Recent advancements in the field have emphasized the potential of ML to predict gasification outcomes and optimize operating conditions, but they often rely on simpler models or smaller datasets. Furthermore, the integration of a heuristic algorithm for multi-objective optimization and subsequent techno-economic analysis enhances the practical applicability of the findings, addressing the critical challenges of cost and scalability in biomass-coal co-gasification technology. Overall, this study contributes significantly to the advancement of technically feasible, economically viable, and environmentally friendly co-gasification of biomass and coal.

2. Research methodology

The primary objective of this study is to enhance and refine ML models tailored for the complex process of co-gasification involving biomass and coal. To accomplish this aim, a meticulously planned approach was implemented. The study commenced with an extensive review and analysis of existing literature on biomass-coal co-gasification, aimed at achieving a comprehensive understanding of the subject and using ML techniques for process modeling. The process began with thorough library research and systematic examination of numerous published papers to identify the most relevant and well-supported sources for further investigation and evaluation. Subsequently, the study focused on determining the critical operational parameters and feedstock properties influencing biomass-coal co-gasification. A dataset was then compiled by extracting pertinent data from eligible papers. Moving forward, the dataset underwent refinement, including data imputation, outlier removal, and normalization, before being analyzed and integrated into ML techniques for modeling. Four ML models, KRR, DTR, RFR, and GBR, were used to construct predictive models for biomass and coal co-gasification processes. Subsequently, the most effective ML model was identified based on its superior statistical performance. Feature importance analysis was conducted on the selected ML model to identify the most influential input parameters. Subsequently, a multi-objective optimization process was conducted using an evolutionary algorithm (i.e., genetic algorithm) to determine the optimum conditions for maximizing H₂ production and facilitating suitable conditions for the Fischer-Tropsch process. Finally, a techno-economic analysis was performed to evaluate the potential for process commercialization based on the multi-objective optimization results proposed by the genetic algorithm.

2.1. Data compilation

To ensure the validity and robustness of developing an ML model, obtaining a comprehensive and representative dataset for modeling and analysis is paramount. The process of acquiring suitable data for studying biomass-coal co-gasification involved a meticulous and systematic approach, primarily using Scopus as the main database and Elsevier or Google Scholar as secondary search databases. For Scopus, a systematic search strategy was implemented, employing relevant keywords such as “co-gasification”, “biomass”, and “coal” across article titles, abstracts, and keywords spanning from 2005 to 2023. Subsequently, eligible papers were selected based on their relevance to the subject matter and the availability of required data. To ensure coherence and a targeted approach in model development, studies focusing on specific gasification techniques, such as catalytic gasification, were excluded. The next step entailed a comprehensive evaluation of the chosen papers, giving precedence to those that focused specifically on experimental criteria related to biomass-coal co-gasification. Emphasis was placed on papers encompassing all essential input parameters crucial to the requirements of this study. Selecting the appropriate input factors from experimental research is critical to accurately represent the complex physical and chemical aspects of biomass-coal co-gasification, which depend on biomass-coal interactions and operational parameters.

The input parameters introduced to the ML modeling included the elemental composition of biomass and coal (carbon, hydrogen, nitrogen, oxygen, and sulfur), all measured in wt%. These inputs provided a foundational quantitative understanding of the different types of materials used as feedstocks in the process. Additionally, ash content, moisture content, volatile matter, and fixed carbon content of biomass and coal, all measured in wt% through proximate analysis, could help to better understand the unique properties of these materials. Operational parameters of the co-gasification process, including temperature (°C), biomass/coal blending ratio (%), equivalent ratio (–), and type of gasifying agent, served as other independent input parameters. Five types of gasifying agents, namely carbon dioxide, steam, air, and nitrogen, were reported in the published literature for biomass-coal co-gasification. In the dataset, each gasifying agent was assigned a numerical value based on its presence in the experiment. For instance, if only one type of gasifying agent, such as air, was used, it was represented as 1 in the Excel dataset, while the other types were assigned a value of 0 in the corresponding row. In cases where two gasifying agents were employed, they were assigned equal ratios based on the values reported in the paper, such as 0.5 each. The flow rate of the gasifying agent was omitted from the modeling due to the limited reported data. The corresponding selected output parameters for modeling included syngas yield (%), H₂ content (mol%), CO₂ content (mol%), CH₄ content (mol%), CO content (mol%), and syngas LHV (MJ/Nm³).

An Excel file was prepared to serve as the dataset, encompassing all input and output values required for the study (Supplementary Excel File). The data used in this study were collected from relevant papers, extracting details from text, tables, and graphs. Information found in text or tables was directly recorded in an Excel file, whereas data from graphs were digitized using a tool called “Plot Digitizer”. This software was calibrated to translate graphical pixels into corresponding x and y coordinates. To ensure uniformity, input values belonging to the same parameter but expressed in different units were standardized and transformed to a common unit. Subsequently, the dataset underwent rigorous filtering to identify and remove outliers, which is crucial for preventing overfitting in ML model development. The Grubbs technique, implemented in Origin software, was utilized to identify outliers in the dataset for each specific input at a significance level of 0.01. After outlier removal, the imputation of missing data became essential in the data compilation process. Missing data related to biomass or coal elemental composition and proximate analysis parameters were imputed using the MissForest methodology. The imputed data can be identified in the

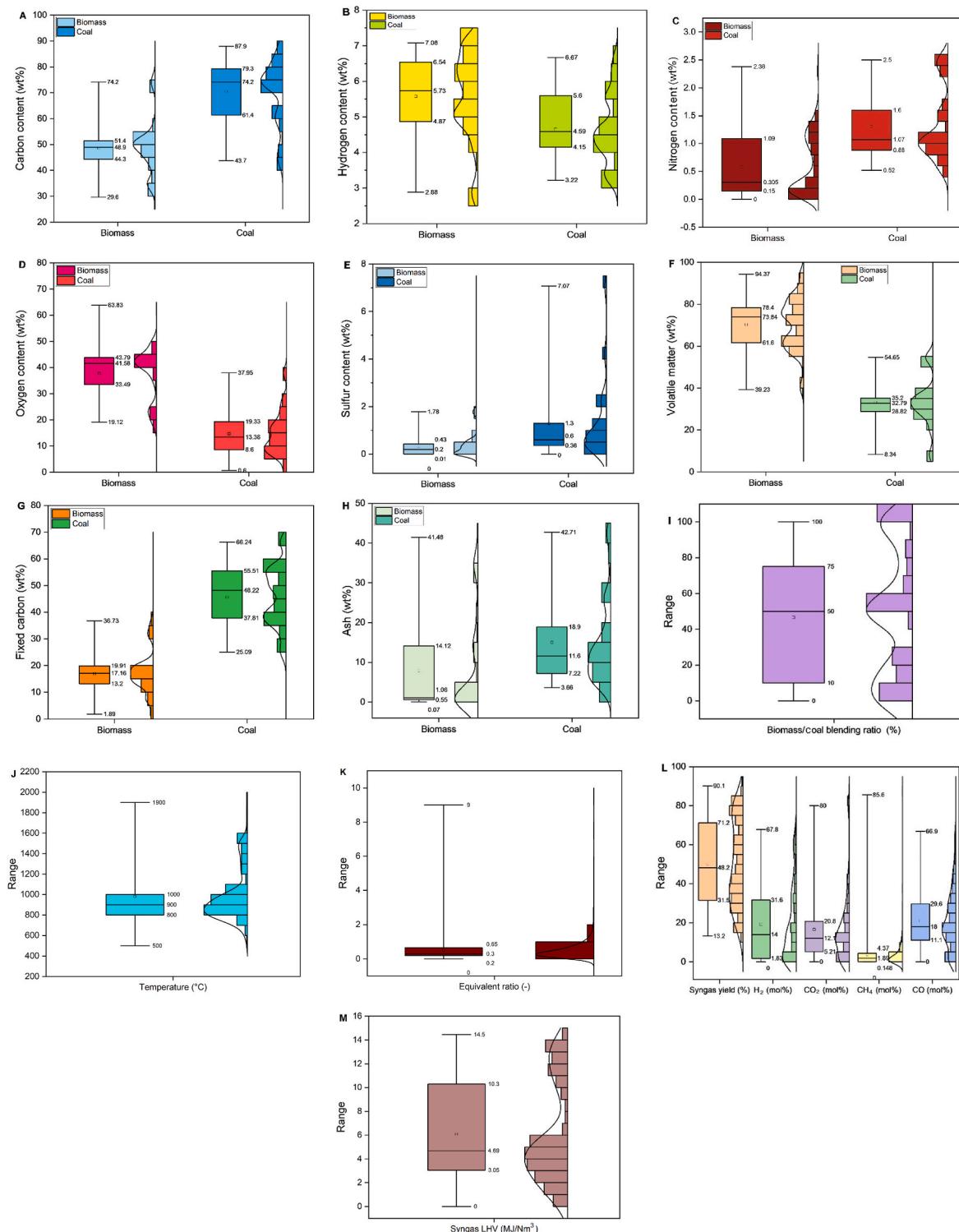


Fig. 1. Statistical evaluation of modeling database (A) carbon, (B) hydrogen, (C) nitrogen, (D) oxygen, (E) sulfur, (F) volatile matter, (G) fixed carbon, (H) ash, (I) biomass/coal blending ratio, (J) temperature, (K) equivalent ratio, (L) syngas yield and compositions, and (M) syngas LHV.

dataset by green-inked cells. A detailed overview of the dataset, comprising 458 data rows, is provided in the “Supplementary Excel File” accompanying this study.

2.2. Machine learning modeling

To enhance training stability, reduce biases, improve interpretability, and enhance model performance, it is essential to normalize the

training dataset before model training. In this study, all input and output parameters were normalized to a range of 0–1. Four types of ML models were considered for developing the predictive model of biomass-coal co-gasification: KRR, DTR, RFR, and GBR. For each model, the dataset was divided into training and testing phases. Specifically, 80 % of the total data was designated for the training phase, with the remaining 20 % set aside for testing purposes.

KRR belongs to the family of kernel methods, a group of pattern

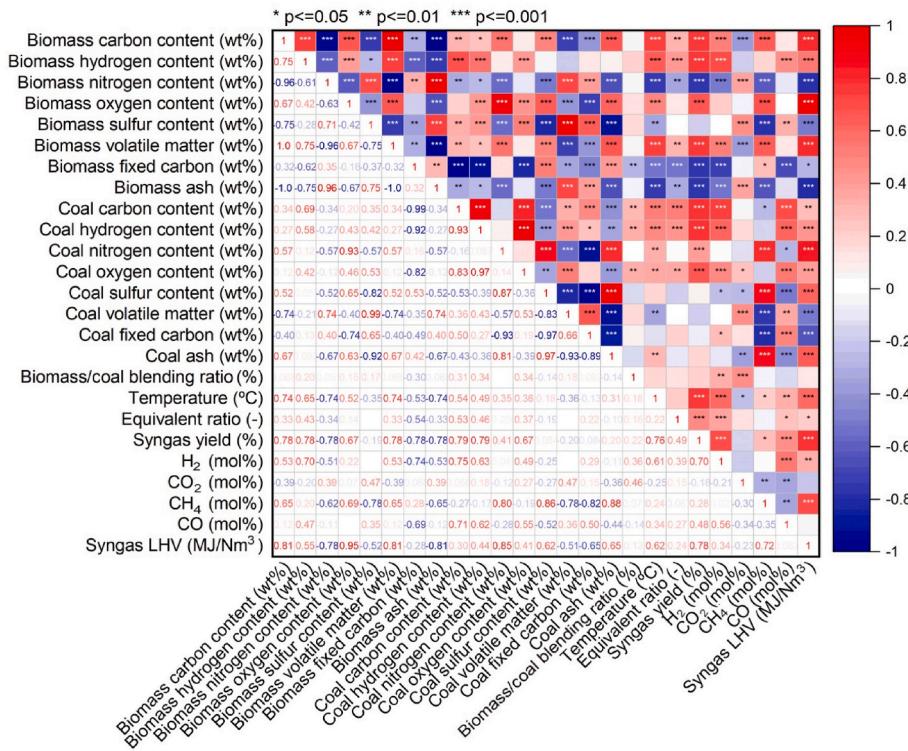


Fig. 2. The heatmap displays Spearman correlation coefficients and their associated p-values for all inputs and outputs. Red indicates positive correlations, blue indicates negative correlations, and the color intensity reflects the strength of the associations.

analysis techniques, with the support vector machines being its most well-known approach. KRR demonstrates good performance, especially when there is a nonlinear structure in the data, as it is developed based on kernel-based methods [20]. Kernel methods derive their name from the utilization of a kernel function, enabling them to operate in a high-dimensional space known as the reproducing kernel Hilbert space without the need for explicit computations in that space. This capability is enabled by the kernel trick, which allows the computation of inner products of mapped data points using a kernel function, often nonlinear in nature [21]. KRR is particularly appealing due to its simple implementation, fast processing speed, and accuracy. It employs squared error loss as the loss function combined with ℓ^2 regularization [22].

DTR is a non-parametric supervised ML algorithm widely used for analyzing large datasets in both classification and regression tasks. This algorithm operates by utilizing independent variables to predict the dependent variable (often referred to as the target variable) through the creation of straightforward decision rules. DTR constructs a decision tree through a recursive process of segmenting the training dataset into subsets based on the similarity of variables. The algorithm comprises branches, nodes, and leaves, where each node represents a decision point, and each leaf represents an outcome. The algorithm assesses the impact of each variable on the target variable by splitting the tree into new branches at each node [23]. The model structure, including nodes, branches, and leaves, symbolizes the entire sample, dataset characteristics, and outcomes, respectively. At each node, purity is measured to ensure optimal splits, thus enhancing accuracy. Common techniques to measure purity include Gain-ratio, Gini index, and Chi-square [23]. Pruning is employed to refine the tree by utilizing average target values and eliminating redundant or non-critical branches, thereby optimizing the final results [24].

RFR operates as an ensemble technique that integrates decision trees with an element of randomness. This methodology entails generating multiple decision trees using random subsets of data and variable features. The predictions from these individual trees are then independently generated and averaged to produce the final output [25]. RFR is

adept at handling both linear and nonlinear relationships while effectively managing high-dimensional datasets comprising categorical and numerical attributes. It proficiently captures intricate interactions among input variables, mitigates overfitting, and demonstrates resilience against outliers and missing values. Within the algorithm, each tree construction involves selecting the best feature or split point from a randomly chosen subset of input features, thus reducing individual tree strength to mitigate correlations. This strategy diminishes generalization error by diminishing correlations, allowing trees to grow without pruning, and contributing to computational efficiency. RFR also employs out-of-bag elements for unbiased estimation of generalization error, preventing overfitting as the number of trees increases in the RFR model [26]. Some of the most critical hyperparameters within the RFR algorithm include the number of trees, randomly selected predictors, and the minimum observations at terminal nodes (node size). It is noteworthy that higher values for the number of trees can enhance stability in estimating variable importance [27].

GBR is an ensemble model consisting of a series of iteratively arranged tree models, where each subsequent model learns from the errors of its predecessor. In this method, an ensemble of decision trees is employed, with each tree acting as a weak learner. However, when integrated into an ensemble, these trees collectively form a potent predictive tool. The concept of “gradient boosting” specifically refers to the application of the gradient descent technique to optimize the fitting process [28]. In GBR, the gradient descent loss function is employed to minimize errors by iteratively updating the initial estimations with new estimations. As a result, a final model emerges that consolidates all preliminary estimations while assigning appropriate weights to yield an optimized predictive model.

The k-fold cross-validation method was employed in this study in order to assess the performance of ML models and select the best model configuration. The dataset was divided into k equal folds, with the model trained on k-1 folds and validated against the remaining fold iteratively. This rotation ensures that every data point serves as validation once, reducing bias and variance in evaluations and avoiding

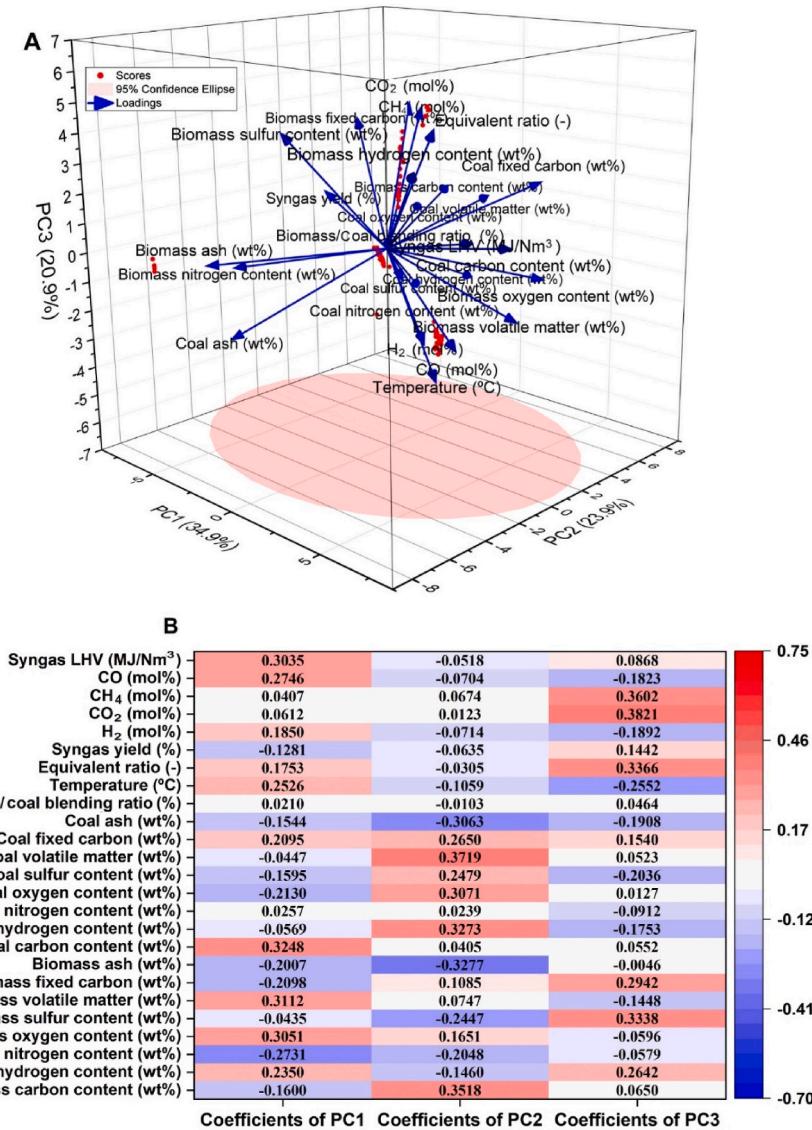


Fig. 3. Principal component analysis of the database (A) 3D plot illustrating the influence of inputs and outputs across the entire dataset (B) heat map showing the coefficient of each descriptor with the top three principal components.

overfitting during model development. Averaging the performance across these iterations provides a robust estimation of the generalizability and performance of the model, offering a reliable assessment of its predictive capabilities while effectively managing the available data for both training and validation purposes [29]. It is important to note that small k-fold values may impact results and affect accuracy, while excessively large K values increase computational costs. Therefore, in this study, 5-fold cross-validation was considered to strike a balance between model accuracy and computational cost.

Identifying the best-performing model among the developed models for further improvement and analysis is crucial, particularly when evaluating multi-output models, such as biomass and coal co-gasification processes. Recognizing the diversity in output magnitudes holds significant importance. While the coefficient of determination (R^2) remains valuable for assessing individual output correlations, it is essential to complement its assessment with other statistical criteria. Hence, root-mean-square-error (RMSE) and mean absolute error (MAE) are also considered for evaluating the performance of the developed models [30].

2.3. Model interpretation and optimization

Indeed, while ML models are known for their accuracy in predictions, understanding the rationale behind these predictions is equally crucial, especially for domain experts. Model interpretability helps to shed light on the relationship between independent inputs and output predictions, making decisions more transparent and trustworthy. One popular technique for model interpretation is SHapley Additive exPlanations (SHAP) analysis, proposed by Lundberg and Lee and developed based on game theory principles [31]. SHAP methodology provides meaningful explanations by attributing a score to each feature, indicating its contribution to the prediction. In this analogy, the inputs are considered as players, and the prediction is the payout. SHAP analysis determines the contribution of each “player” to the overall prediction, thereby elucidating the importance of each input feature [32]. By providing interpretable explanations of predictive models, SHAP analysis aids domain experts in making informed decisions, particularly in experimental situations related to the predicted process. Understanding the significance of each input feature empowers experts to discern key factors driving the model’s predictions and to make more informed decisions based on these insights [33]. This not only enhances the

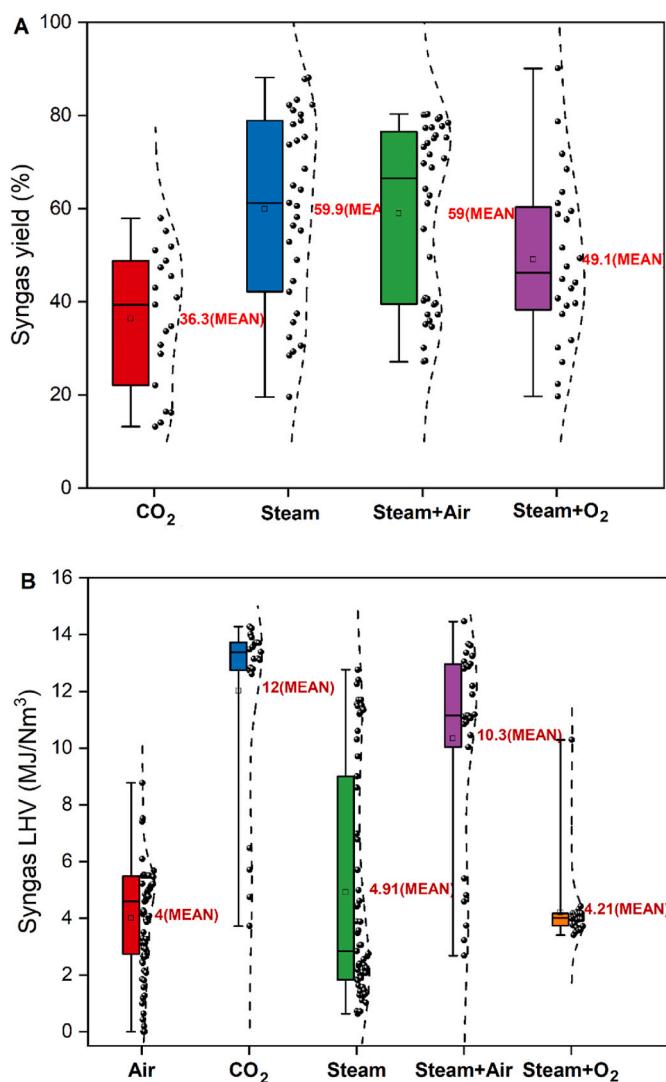


Fig. 4. Effect of gasifying agent on (A) syngas yield and (B) syngas LHV during biomass-coal co-gasification.

trustworthiness of the model but also facilitates better decision-making in real-world applications. It has been successfully employed to interpret ML modeling results in previous studies focused on biomass thermo-chemical conversion [34].

After identifying the best-performing model, it was integrated into several evolutionary algorithms to optimize the biomass-coal co-gasification process. Among these, the genetic algorithm demonstrated superior results. Consequently, the genetic algorithm was selected as the focus for the remainder of this investigation due to its superior performance. To maximize optimization performance, two distinct criteria were considered: maximizing H₂ production and achieving optimum conditions for the maximum Fischer-Tropsch process. For each criterion, specific input and output constraints were defined. The input constraints aimed to ensure that the operational parameters fell within the dataset range while also ensuring that the summation of biomass and coal compositions did not exceed 100 %. The importance of each output for each criterion was also determined. This approach allowed outputs with higher importance degrees to carry more weight during the optimization process. It is important to note that the optimization process was conducted separately for each specific gasifying agent (CO₂, steam, air, N₂, and O₂). Table 2 presents the optimization objectives tailored to meet the criteria for biomass-coal co-gasification.

Table 3
Statistical performance parameters of the developed models.

Model type	Syngas yield (wt %)	H ₂ (mol %)	CO ₂ (mol %)	CH ₄ (mol %)	CO (mol %)	Syngas LHV (MJ/Nm ³)
Training phase						
R ²	KRR	0.69	0.72	0.62	0.57	0.62
	DTR	0.88	0.88	0.86	0.77	0.79
	RFR	0.93	0.92	0.93	0.90	0.87
	GBR	0.98	0.93	0.97	0.94	0.93
MAE	KRR	8.75	6.90	7.62	1.54	6.85
	DTR	4.76	3.46	4.46	0.90	5.12
	RFR	3.97	2.37	2.80	0.54	3.39
	GBR	1.20	2.55	1.27	0.30	2.03
RMSE	KRR	11.89	9.20	11.84	2.44	10.13
	DTR	7.32	6.11	7.02	1.81	7.46
	RFR	5.67	4.85	4.98	1.19	5.98
	GBR	2.81	4.75	2.98	0.94	4.42
Testing phase						
R ²	KRR	0.68	0.70	0.59	0.55	0.52
	DTR	0.84	0.87	0.86	0.72	0.77
	RFR	0.89	0.92	0.93	0.89	0.85
	GBR	0.98	0.93	0.96	0.94	0.92
MAE	KRR	9.12	5.88	5.06	1.50	7.31
	DTR	5.90	3.48	4.26	1.05	5.55
	RFR	6.68	2.53	3.95	0.68	4.05
	GBR	2.06	2.69	2.28	0.43	2.92
RMSE	KRR	10.74	7.35	8.63	2.00	9.80
	DTR	8.02	5.00	6.55	1.51	7.64
	RFR	7.72	3.91	5.67	1.09	6.10
	GBR	3.39	3.82	3.78	0.72	4.45

2.4. Techno-economic analysis

Techno-economic analysis offers valuable insights for decision-makers and stakeholders regarding the economic feasibility and potential commercialization of the optimized biomass-coal co-gasification process. The procedure for conducting techno-economic analysis, based on the optimized conditions provided by the selected ML model for biomass and coal co-gasification, aimed at maximizing H₂ production, involved several key steps. Initially, the optimized conditions, including temperature, pressure, feedstock composition, biomass/coal blending ratio, and reactor design, were identified using the ML model. These optimized parameters served as inputs for process simulation using Aspen Plus software (v14) to model the entire biomass-coal co-gasification process. Subsequently, the simulated process was integrated into an economic model that comprehensively considers factors such as capital costs, operating costs, feedstock costs, energy prices, and product revenues. Finally, the results of the techno-economic analysis were analyzed to determine the economic feasibility and potential profitability of implementing the optimized gasification process on a large scale. This assessment included considerations such as investment payback period, net present value, and internal rate of return.

3. Results and discussion

3.1. Data analysis

Statistical analysis is pivotal in unraveling the complexities of biomass-coal gasification, offering a comprehensive understanding of the factors influencing the process. The collected dataset underwent meticulous examination to extract initial insights into various descriptors and target variables. Basic statistical indicators, such as minimum, maximum, and median values, along with quartile-based box plots, were employed to reveal a plethora of information (Fig. 1). Exploring biomass feedstocks, the dataset reveals a diverse range of ultimate and proximate analyses, encompassing genetic variances, geographical origins, climatic conditions, and harvest seasons. Notably, coal exhibits a high carbon content (43.7–87.9 wt%), while biomass

Table 4

Hyperparameters of the developed GBR models for each output.

Hyperparameter	Syngas yield (wt%)	H ₂ content (mol%)	CO ₂ content (mol%)	CO content (mol%)	CH ₄ content (mol%)	Syngas LHV (MJ/Nm ³)
learning_rate	0.01	0.01	0.01	0.01	0.01	0.01
max_depth	7	4	7	7	4	7
n_estimators	500	500	500	500	500	500
subsample	0.5	0.5	0.5	0.5	0.5	0.5

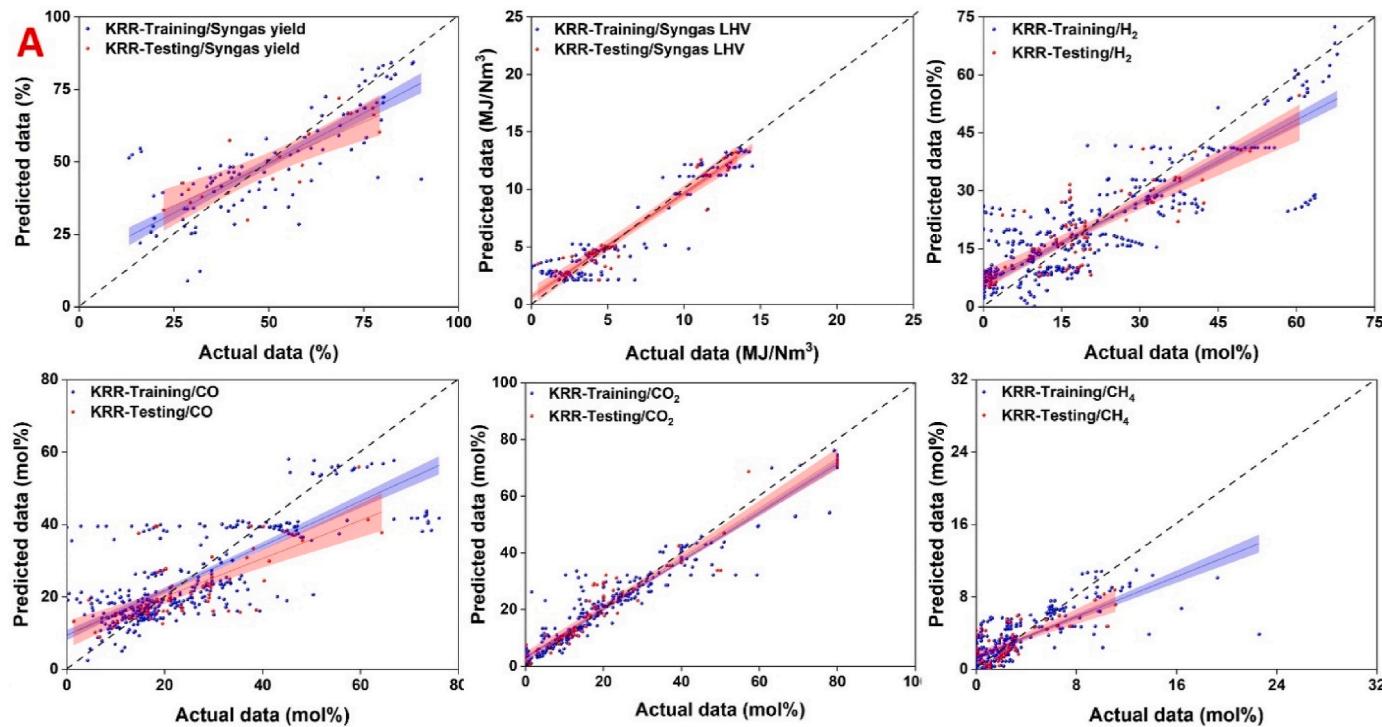


Fig. 5. Scatter plot of predicted data from the developed models (A) KRR, (B) DTR, (C) RFR, and (D) GBR.

displays higher volatile matter (39.23–94.37 wt%), indicating the potential for high-energy-density alternative fuel production.

Operational parameters, such as temperature and equivalent ratio, display a wide range (500–1900 °C and 0.2–9, respectively), providing a fertile ground for developing a robust ML model capable of generalizing across diverse conditions. Output features, notably syngas yield (ranging from 13.2 to 90.1 wt% with a median of 48.2 wt%), highlight significant variations in product distribution. CO emerges as the primary component in syngas composition, with a median value of 18 mol%, followed by H₂ at 14 mol%, CO₂ at 12.1 mol%, and CH₄ at 1.89 mol%. Operational parameters, ranges, and output target distributions further underscore the inclusive nature of the dataset, laying the foundation for a robust ML model. With biomass exhibiting lower carbon content and higher volatile matter compared to coal, alongside distinct temperature and syngas composition ranges, the dataset presents a rich resource for developing an ML model capable of optimizing biomass-coal gasification. Overall, the descriptive analysis of the collected dataset not only elucidates the complexities of biomass-coal gasification but also paves the way for the development of a versatile ML model adept at accommodating the diverse conditions inherent in this field.

In uncovering the intricate relationships within the biomass-coal gasification dataset, Spearman correlation analysis can be a pivotal tool. This non-parametric technique, widely utilized in previous studies, employs rank values to evaluate the strength and direction of associations between variables. Spearman correlation employs a monotonic function, providing a robust means to quantitatively evaluate

interdependence. Derived from rank values, similarity scores delineate the degree of correlation, with +1, 0, and -1 indicating perfect positive, weak, and complete negative correlations, respectively. This approach enables a subtle exploration of the dataset, facilitating the determination of both magnitudes and directions of associations. Spearman correlation matrices distinctly illustrate the interconnectedness between input descriptors and output responses, offering valuable insights into the underlying dynamics of the biomass-coal gasification process. The calculated correlation coefficients reveal influence trends of various parameters, laying the groundwork for a deeper understanding of the intricate interplay within the dataset.

As shown in Fig. 2, the H₂ content of syngas shows a positive correlation with temperature ($r = 0.61, p < 0.001$). Higher temperatures promote endothermic gasification reactions according to Le-Chatelier's principle [35]. These reactions include primary and secondary water gas reactions, secondary cracking, and reforming of heavy hydrocarbons/tars, which lead to the increased formation of H₂ [3]. Similarly, the CO content of syngas indicates a positive correlation with temperature ($r = 0.34, p < 0.01$) since increasing temperature leads to heightened CO production, driven by accelerated heterogeneous and endothermic reactions like water gas and Boudouard reactions [3]. Conversely, the CO₂ content of syngas demonstrates a negative correlation with temperature ($r = -0.25, p < 0.05$), indicating that higher operational temperatures lead to a decrease in the CO₂ content of syngas during the biomass-coal gasification process. Increased temperature prompts carbon to react more with steam and CO₂, yielding higher CO

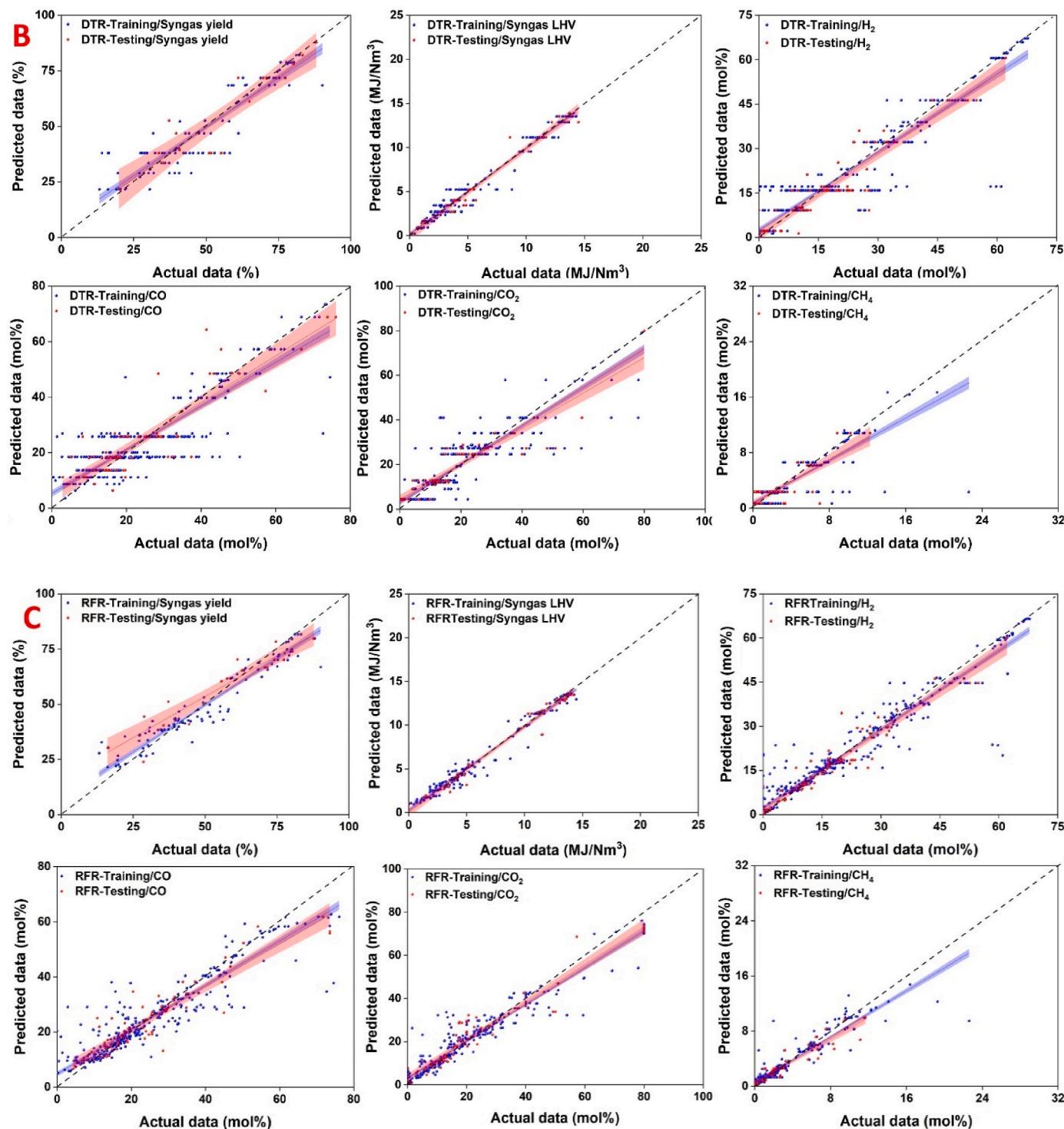


Fig. 5. (continued).

levels. By increasing temperature, the CO₂ level begins to decline due to the heightened occurrence of endothermic reactions, such as the Bou-douard reaction, which entails the consumption of CO₂. Overall, in the chemical reaction-controlled regime, achieving greater co-gasification reactivity is possible by elevating the gasification temperature [9].

The biomass/coal blending ratio demonstrates a positive influence on syngas yield in biomass-coal co-gasification ($r = 0.22$). Elevated concentrations of alkaline and alkaline earth metals (such as K, Al, and Ca) in biomass feedstocks, along with lower carbon structure ordering, contribute to higher reactivity with increasing biomass proportion,

thereby impacting gasification reactions [9]. It is noteworthy that the heightened reactivity of water gas shift, char gasification, and methane reforming reactions is attributed to the presence of alkaline earth metals. Consequently, this results in an increase in H₂ enrichment and a decrease in the CH₄ level [36]. However, the biomass/coal blending ratio exhibits only a weak positive correlation with syngas LHV (Fig. 2).

Employing principal component analysis (PCA), a robust technique for revealing patterns and relationships within datasets, the study investigates the intricacies of biomass-coal co-gasification. PCA efficiently reduces dataset dimensionality, revealing critical components

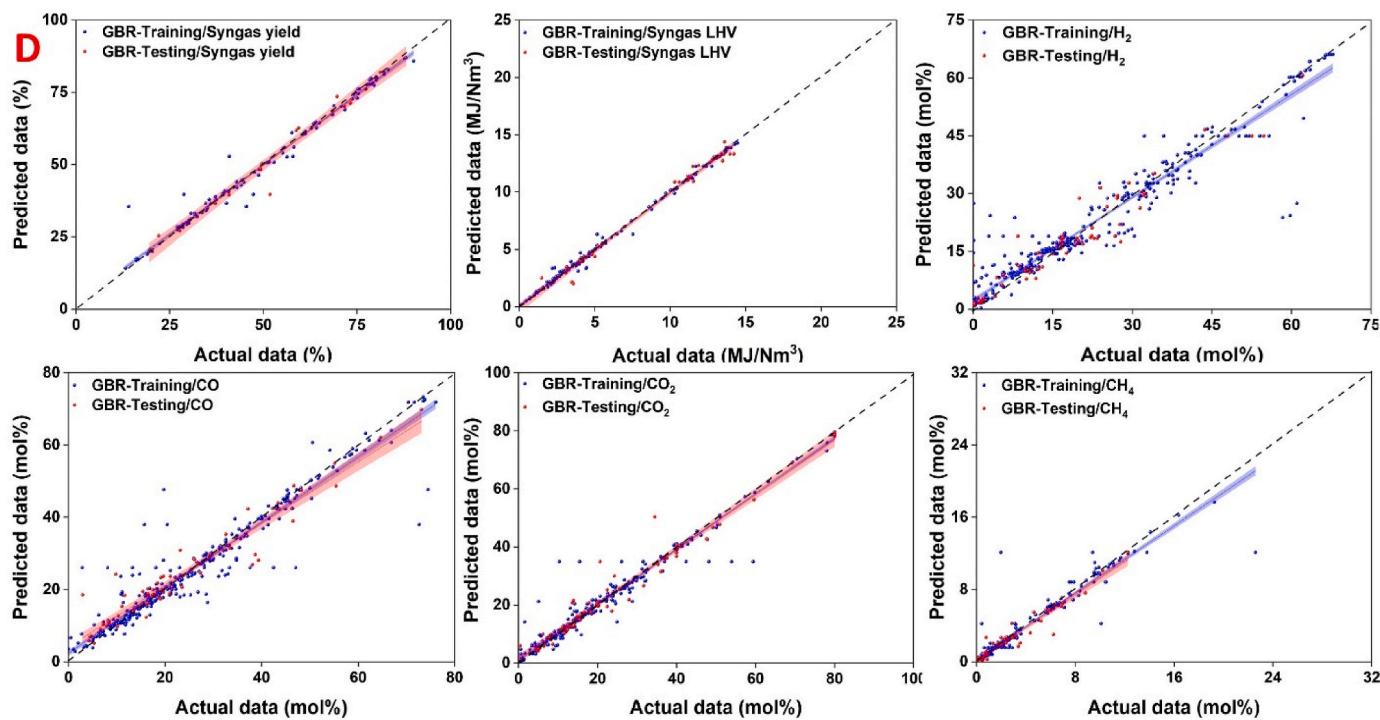


Fig. 5. (continued).

contributing to overall variance. Analysis of the collected dataset indicates that the first three principal components account for approximately 34.9 %, 23.9 %, and 20.9 % of the total variance, respectively (Fig. 3A). Notably, the first principal component predominantly reflects biomass characteristics (e.g., volatile matter), emphasizing its significant role in shaping dataset variability. Conversely, the second principal component encapsulates coal properties, including volatile matter, hydrogen content, and oxygen content. The third principal component revolves around operational parameters, such as equivalent ratio, temperature, and biomass/coal blending ratio. These findings underscore the pivotal role of feedstock properties and operational parameters in shaping the dataset structure. PCA results provide valuable insights into the interplay of variables during biomass-coal co-gasification. Fig. 3B visually depicts the distribution of variables, highlighting the distinct contributions of each principal component to the overall dataset variance.

The equivalent ratio, representing the ratio of actual air/fuel to stoichiometric air/fuel, exhibits a positive correlation with syngas yield. In a more oxidative environment characterized by higher equivalent ratios, increased combustion reactions can lead to a higher production of syngas, primarily composed of carbon dioxide. As the equivalent ratio increases, there is a heightened influx of oxygen into the riser section, resulting in enhanced carbon conversion (oxidation to CO_2) and increased heat generation, consequently raising the temperature. However, it is important to note that elevating the equivalent ratio typically reduces syngas LHV due to heightened levels of combustion byproducts, such as CO_2 . Conversely, in a reducing atmosphere associated with lower equivalent ratios, syngas often achieves a higher LHV, notably enriched in H_2 and CO [37].

In Fig. 4, a box plot illustrates the impact of pure and mixed gasifying agents (such as CO_2 , steam, O_2 , steam + O_2 , and steam + air) on syngas yield and syngas LHV. The choice of gasifying agent significantly influences gasification reactions, affecting both the composition and energy content of the resulting syngas. On average, steam gasification yields a higher syngas yield (59.9 %) compared to other gasifying agents (CO_2 : 36.3 % and steam + O_2 : 49.1 %). In an oxidizing setting facilitated by steam or air, there is a tendency towards higher syngas yield, albeit

with a reduced syngas LHV dominated by combustion byproducts like CO_2 . Conversely, using CO_2 as a gasifying agent in a reducing atmosphere promotes a higher LHV in the syngas enriched in H_2 and CO despite the potential for a lower overall syngas yield. Air-assisted gasification yields syngas with a lower LHV (on average, 4 MJ/Nm^3) due to nitrogen dilution from the air [38]. Despite this drawback, air gasification is valued for its simplicity and cost-effectiveness compared to alternative gasification methods. The strategic selection of the appropriate gasifying agent is crucial for optimizing both the quantity and quality of syngas in biomass-coal co-gasification, highlighting the complex interaction among feedstock composition, operational parameters, and gasifying agents.

3.2. Modeling results

The performance of the developed ML models was evaluated based on three statistical parameters: R^2 , MAE, and RMSE. These parameters were calculated during both the training and testing phases, considering the actual data and the relative predicted values. Table 3 presents the determined statistical parameters for the outputs of the developed models using different ML techniques. The GBR model demonstrates superior performance compared to other techniques, exhibiting the highest R^2 values and the lowest MAE and RMSE values in both the training and testing phases. During the training phase, the R^2 values for the GBR model ranged between 0.93 and 0.99, while the MAE ranged from 0.12 to 2.55, and the RMSE ranged from 0.23 to 4.75. Similarly, in the testing phase, the R^2 values varied from 0.92 to 0.99, the MAE ranged from 0.38 to 2.92, and the RMSE ranged from 0.55 to 4.45. These outcomes indicate that the GBR model provides the most promising results for the prediction of syngas lower heating value, CO_2 content, syngas yield, H_2 content, and CO content, exhibiting the highest R^2 values and the lowest errors. The second-best model is RFR, based on statistical parameters. For further analysis and optimization processes, the GBR model was selected. The hyperparameters of the developed GBR model for each specific output are detailed in Table 4.

A way to visually represent the performance of the developed GBR model for different outputs of the biomass-coal co-gasification process is

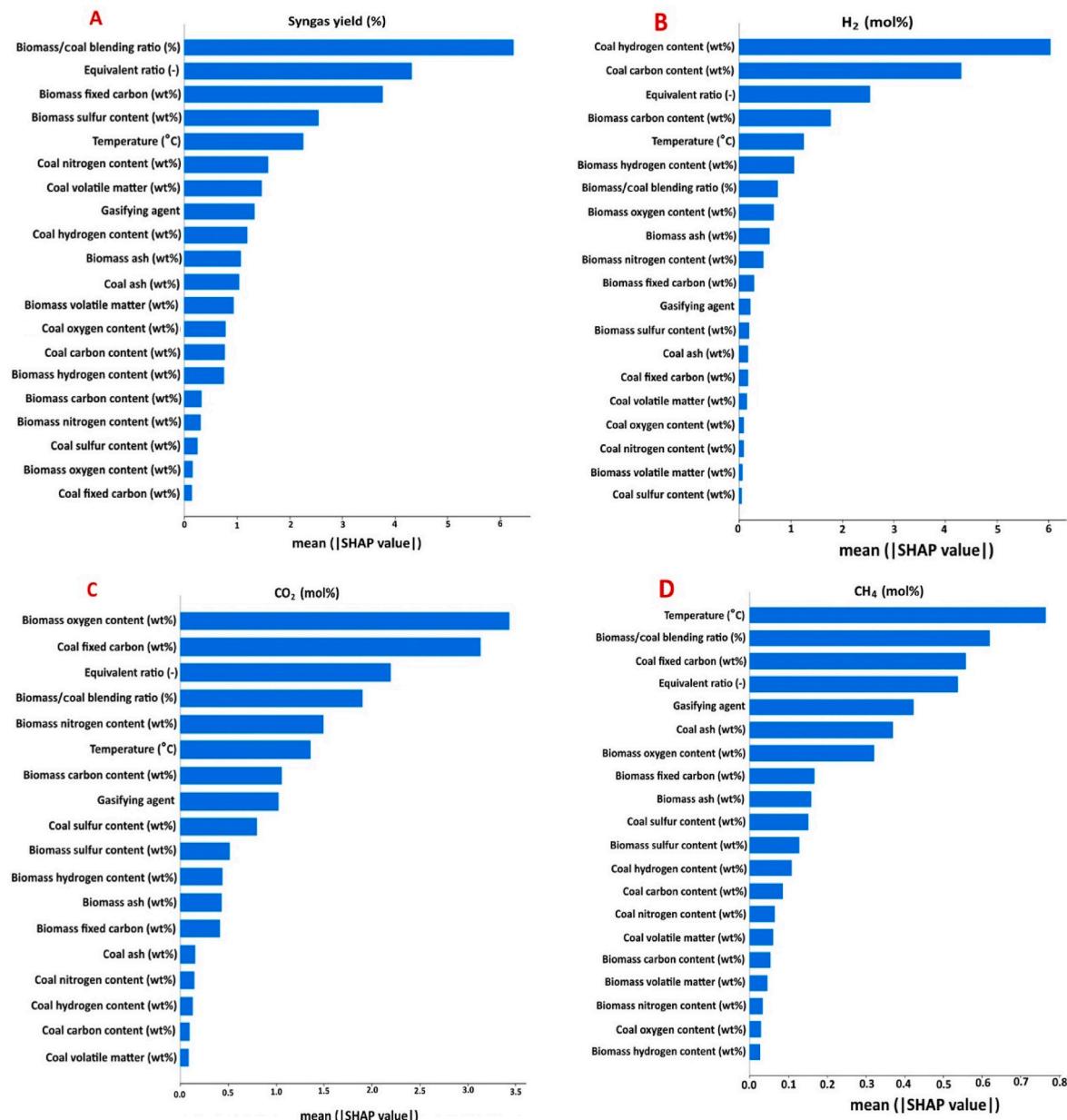


Fig. 6. SHAP results for the independent descriptors and the dependent responses (A) syngas yield, (B) H₂, (C) CO₂, (D) CH₄, (E) CO, (F) syngas LHV.

by creating scatter plots of the predicted data versus the experimental data. In Fig. 5, the scatter plot illustrates the predicted data from the developed models alongside the experimental data for both the training (blue dots) and testing (red dots) phases. Additionally, transparent blue and red bands surrounding the fitted regression lines represent the 95 % confidence intervals for the training and testing phases, respectively. The confidence bands in the scatter plot delineate the region likely to contain the true relationship between the two variables. The width of the confidence band indicates the uncertainty in the estimated relationship: a narrower band suggests a more precise estimation, while a wider band indicates greater uncertainty. The results demonstrate that the predicted data points from the GBR model closely align with the actual data, indicating a strong linear correlation. This visual representation underscores the accuracy and reliability of the GBR model in predicting various outputs of the biomass-coal co-gasification process. The strength and potential of the GBR technique in this context are attributed to the flexibility of the model in handling different types of

data, its robustness, and its insensitivity to outliers [39].

3.3. Feature importance analysis

The SHAP methodology was employed to elucidate feature importance in the best-performing ML model. Grounded in coalitional game theory, SHAP effectively evaluates the significance of each input descriptor on the corresponding output, offering interpretability at both local and global scales. The methodology assigns contribution values to individual data points, enhancing model interpretability. Absolute SHAP values across the dataset provide a comprehensive understanding of the overall importance of input variables. However, it is crucial to note that SHAP analysis solely reveals correlations among the most influential input features during modeling stages, such as training and testing, and does not imply causation among selected descriptors. This approach empowers one to unravel the intricate relationships within ML models and sheds light on the critical factors shaping predictions.

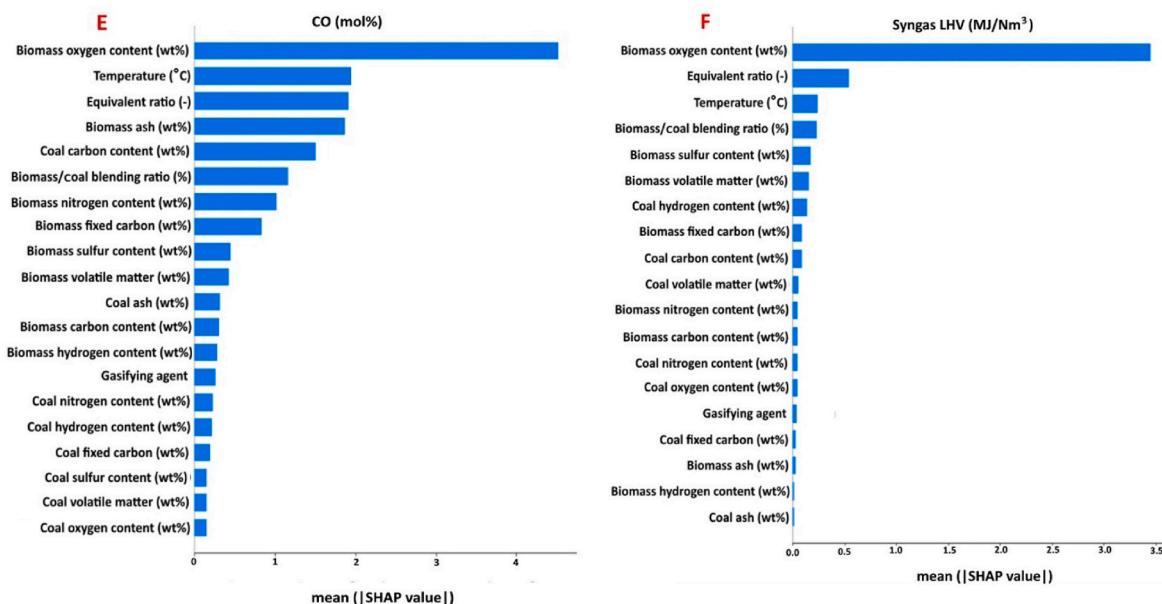


Fig. 6. (continued).

Table 5
Optimization results for biomass-coal co-gasification process targeting H₂ production.

Gasifying agent	Optimum inputs										
	Biomass Coal carbon content (wt%)	Biomass Coal hydrogen content (wt %)	Biomass Coal nitrogen content (wt%)	Biomass Coal oxygen content (wt%)	Biomass Coal sulfur content (wt%)	Biomass Coal volatile matter (wt%)	Biomass Coal fixed carbon content (wt%)	Biomass Coal ash content (wt%)	Biomass / coal blending ratio (%)	Temperature (°C)	Equivalent ratio (-)
CO ₂	33.06 61.10	3.75 6.10	1.18 0.93	51.95 18.66	1.19 3.46	86.85 14.69	11.31 36.27	6.22 41.89	65.21	1869	7.81
Steam	42.31 48.54	6.02 6.46	1.73 0.58	47.66 35.94	1.26 4.70	34.42 30.30	28.16 27.37	30.20 39.71	68.24	1482	8.44
Air	44.29 49.83	3.92 6.24	0.17 0.86	44.14 34.75	1.46 1.11	44.50 45.94	29.57 25.22	16.10 28.21	68.21	1850	8.12
N ₂	42.33 51.42	2.97 6.38	0.92 2.16	43.94 33.25	0.63 4.53	67.58 24.07	21.06 47.08	5.25 31.63	37.61	1820	2.38
O ₂	43.46 51.99	6.10 5.95	1.49 2.21	43.67 29.17	1.66 4.41	61.09 23.54	20.24 49.87	10.36 24.05	58.08	1838	6.08
Gasifying agent	Optimum outputs										
	Syngas yield (%)	H ₂ content (mol%)	CO ₂ content (mol%)	CH ₄ content (mol%)	CO content (mol%)	Syngas LHV (MJ/Nm ³)					
CO ₂	81.04	54.91	17.04	1.72	27.45	9.70					
Steam	81.32	56.59	13.74	2.51	28.79	9.12					
Air	76.80	52.73	14.11	1.93	26.36	9.17					
N ₂	72.53	54.78	27.57	11.18	27.39	10.44					
O ₂	80.50	51.75	13.32	1.58	25.87	10.29					

Fig. 6 portrays the outcomes of the SHAP analysis concerning independent input descriptors and their influence on dependent output features, particularly focusing on syngas properties in biomass-coal co-gasification. The positioning of the input features in the figure corresponds to their SHAP values, where higher values indicate greater importance. This figure delineates the absolute mean effect of each input descriptor on the associated outputs. The biomass/coal blending ratio, equivalent ratio, and temperature emerge as key influencers, ranking among the top five descriptors with the most substantial impact on syngas yield. Notably, the biomass/coal blending ratio plays a pivotal role, indicating that the proportion of biomass in the blend significantly affects gasification reactions. Elevated concentrations of alkaline and alkaline earth metals (e.g., K, Al, and Ca) in biomass feedstocks, coupled with lower carbon structure ordering, contribute to heightened reactivity. This issue, in turn, affects gasification reactions and ultimately

impacts syngas yield [9].

3.4. Optimization results

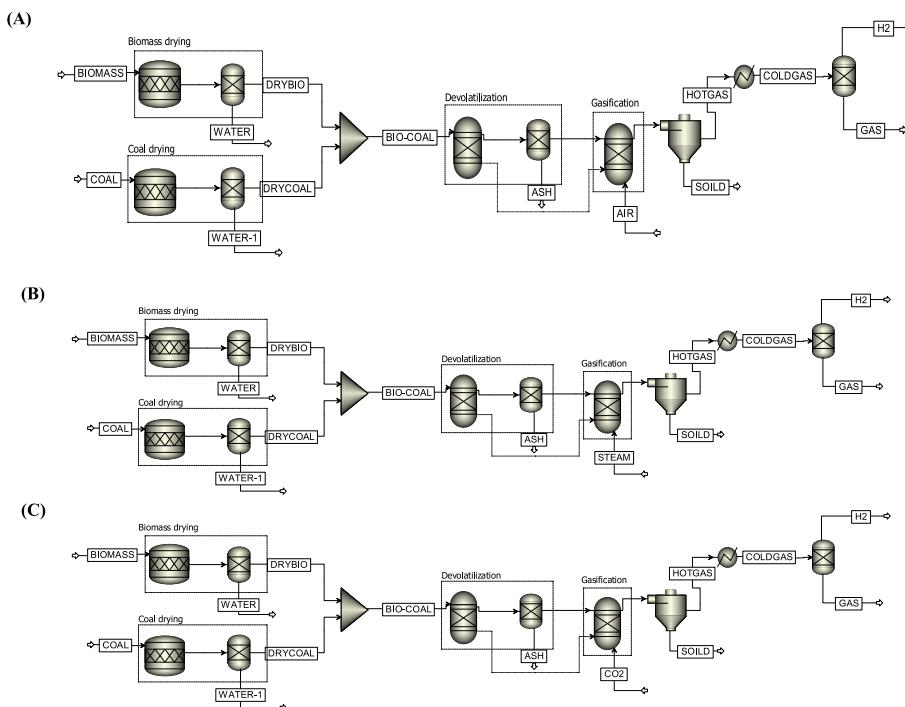
The best-developed GBR model was used as the basis for running a multi-objective optimization process to determine the optimum conditions and suggested outcomes. The genetic algorithm provided superior results compared to other evolutionary techniques. Therefore, to avoid unnecessary and duplicative results, only the optimum conditions identified by the genetic algorithm are reported in this section. Tables 5 and 6 present the optimal biomass composition, coal composition, and operational parameter ranges determined through the genetic algorithm for achieving co-gasification products suitable for H₂ production and the Fischer-Tropsch process, respectively. For H₂ production, the highest syngas yield and H₂ content can be achieved when steam is used as a

Table 6

Optimization results for biomass-coal co-gasification process targeting Fischer-Tropsch criteria.

Gasifying agent	Optimum inputs									Temperature (°C)	Equivalent ratio (-)
	Biomass Coal carbon content (wt%)	Biomass Coal hydrogen content (wt%)	Biomass Coal nitrogen content (wt%)	Biomass Coal oxygen content (wt%)	Biomass Coal sulfur content (wt%)	Biomass Coal volatile matter (wt%)	Biomass Coal fixed carbon content (wt%)	Biomass Coal ash content (wt%)	Biomass / coal blending ratio (%)		
CO ₂	34.34 46.04	6.19 6.33	1.84 0.87	51.61 36.30	0.37 1.93	64.24 44.56	25.88 30.77	0.49 15.36	64.18	1131	3.37
Steam	39.92 63.69	5.72 6.14	0.18 0.71	51.05 19.05	0.47 3.73	55.63 47.63	28.43 24.87	11.74 21.02	57.37	1582	4.96
Air	49.15 55.96	4.24 6.46	0.11 1.86	44.05 22.05	1.66 5.82	51.63 46.72	32.49 42.84	12.41 5.77	58.30	1576	8.59
N ₂	34.61 70.38	5.48 6.54	0.09 1.45	53.84 8.95	0.10 6.57	55.79 23.82	32.93 40.83	10.76 34.28	19.62	1495	8.53
O ₂	47.55 51.68	4.40 6.63	0.58 2.18	43.09 28.32	0.97 2.56	58.88 35.22	19.75 22.92	17.52 40.20	54.77	1405	4.38

Gasifying agent	Optimum outputs						
	Syngas yield (%)	H ₂ content (mol%)	CO ₂ content (mol%)	CH ₄ content (mol%)	CO content (mol%)	Syngas LHV (MJ/Nm ³)	
CO ₂	69.65	55.95	17.68	1.85	26.47	8.52	
Steam	77.16	54.24	17.15	1.89	27.12	9.85	
Air	78.49	56.17	16.27	0.71	28.09	9.77	
N ₂	68.16	50.38	18.59	6.92	24.69	10.85	
O ₂	77.77	53.14	17.11	1.16	26.57	9.47	

**Fig. 7.** Process flow diagrams for (A) air-assisted, (B) steam-assisted, and (C) CO₂-assisted biomass-coal co-gasification processes.

gasifying agent. Steam performs slightly better than CO₂, although all gasifying agents result in similar ranges for syngas yield (81.32–72.53 %) and H₂ production (51.75–56.59 mol%). For the Fischer-Tropsch process, air demonstrates the best results for H₂ production, followed by CO₂. However, in this criterion as well, all gasifying agents produce almost the same results (50.38–56.17 mol%). The syngas LHV in both optimization processes shows similar values, ranging between 9 and 11 MJ/Nm³ under the suggested input conditions.

The optimization outcomes for H₂ production indicate that higher temperatures are required (above 1800 °C), except for steam (above 1400 °C). For the Fischer-Tropsch process, lower temperatures (between 1100 and 1600 °C) are optimal. The results suggest that a biomass/coal

blending ratio between 50 % and 70 % is suitable for almost all gasifying agents in both criteria. However, when using N₂, a lower blending ratio (19 %–40 %) is advised. Different equivalent ratios for each gasifying agent can be observed to meet both goals, with values generally falling between 2 and 9. Additional required inputs for optimal results from the biomass/coal co-gasification process, including biomass and coal composition and proximate analysis, can be found in Tables 5 and 6. The multi-objective optimization results derived from the genetic algorithm can serve as valuable operational guidelines for experimental research or industrial applications, enhancing process efficiency and reducing the need for time-consuming and costly experiments.

Table 7

Criteria, key assumptions, and prices relevant to the techno-economic analysis.

Parameter	Unit	Value
Reference country	–	USA
Currency	–	USD
Biomass feed rate	kg/h	1000
Coal feed rate	kg/h	1000
Plant construction period	month	6
Project economic life	yr	20
Desired rate of return	%	20
Number of working weeks per year	–	52
Interest factor	%	1.2
Tax rate	%	40
Discount rate	%	10
Operating charge	%	25
Plant overhead	%	50
Duration of engineering, procurement, and construction phase	week	16
Length of start-up period	week	26
Electricity price	USD/kWh	0.40548
Hydrogen	USD/t	6000
Biomass	USD/t	50
Coal	USD/t	113
Amortization rate	%	6.7

3.5. Techno-economic analysis of biomass-coal co-gasification

Three distinct scenarios were developed based on the optimum results provided by the genetic algorithm to conduct the techno-economic analysis of biomass-coal co-gasification. Firstly, the air-assisted co-gasification scenario represents a conventional approach where air serves as the gasification agent (Scenario I). Secondly, the steam-assisted co-gasification scenario involves introducing steam into the gasification process to enhance H₂ production and reduce the formation of undesired byproducts (Scenario II). Lastly, the CO₂-assisted co-gasification scenario explores the utilization of carbon dioxide as a gasification agent, presenting an environmentally friendly approach (Scenario III). Fig. 7 illustrates the process flow diagram of these three scenarios. Table 7 also provides a summary of the criteria, key assumptions, and prices relevant to the techno-economic analysis conducted in this study. Through the examination of these scenarios, this study aims to offer insights into the techno-economic feasibility of employing various gasifying agents in biomass-coal co-gasification, thus contributing to the optimization and commercialization of sustainable energy conversion technologies.

Key financial metrics, such as the internal rate of return, net return rate, and profitability index, play a pivotal role in assessing the economic viability of investment projects. The internal rate of return signifies the discount rate at which the net present value of cash flows equals zero, serving as an indicator of the rate of return of the project. Meanwhile, the return rate provides insight into the annualized return on investment, and the profitability index measures the ratio of the present value of cash inflows to outflows. Fig. 8 presents the economic metrics of different scenarios evaluated in this study. Among these scenarios, scenario II—steam-assisted co-gasification of biomass and coal—stands out with significantly superior financial metrics compared to the others. With an exceptionally high internal rate of return of 243 %, return rate of 86.9 %, and profitability index of 1.87, scenario II outperforms scenarios I and III by a considerable margin. This enhanced economic performance is attributed to its capability to produce a higher amount of H₂ as the final product, leading to increased revenue generation. These findings underscore the importance of optimizing biomass-coal co-gasification conditions to maximize valuable product yields, thus enhancing the economic feasibility and profitability of the process.

Table 8 presents the costs, encompassing capital and production costs, alongside their associated annual revenue for different biomass-coal co-gasification scenarios. Among these scenarios, scenario II—steam-assisted co-gasification—demonstrates superior performance in

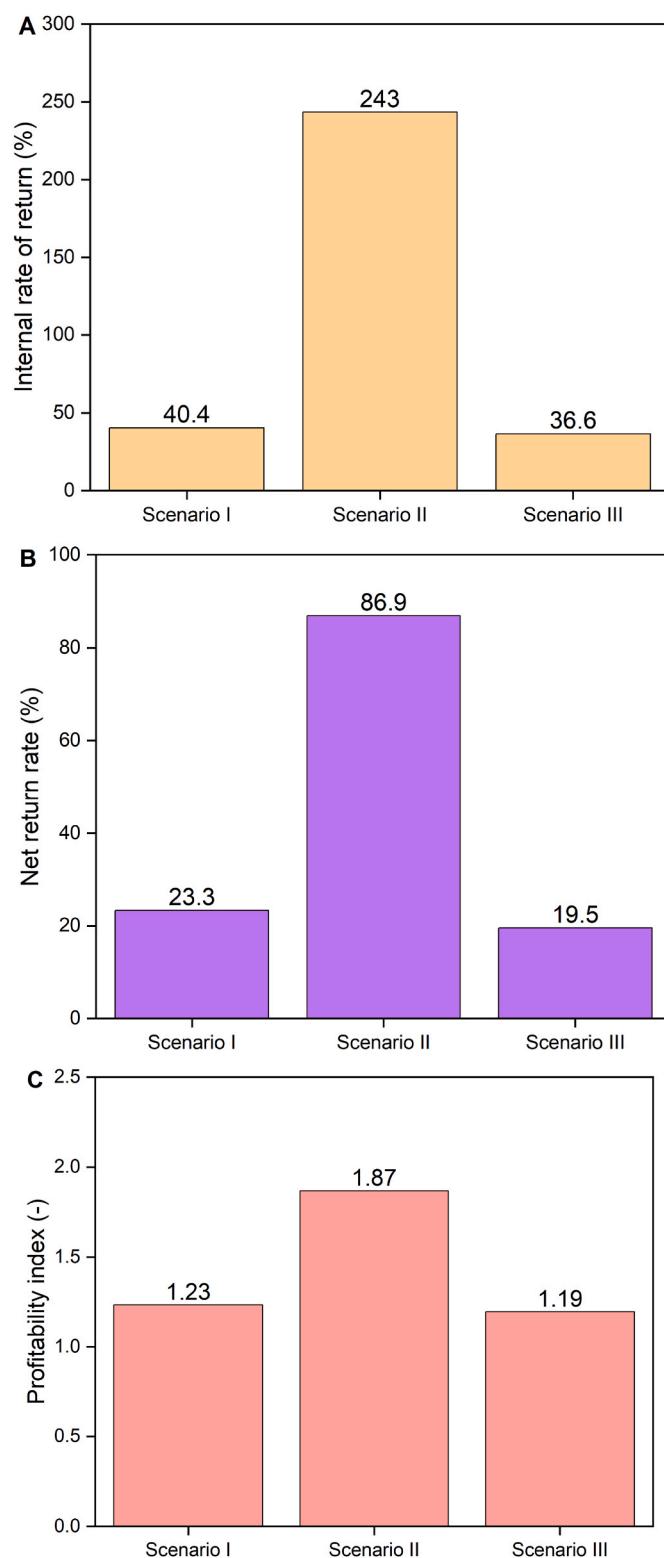


Fig. 8. Economic performance metrics for various scenarios: (A) internal rate of return, (b) net return rate, and (c) profitability index.

terms of total product sales. With a total product sales value of 14,177,900 USD/yr, scenario II significantly outperforms scenario I (4,920,570 USD/yr) and scenario III (4,635,550 USD/yr). The substantial increase in product sales in scenario II can be attributed to its enhanced production of hydrogen as the final product. Steam-assisted co-gasification generally facilitates more efficient H₂ generation

Table 8
Income and expense related to different scenarios.

Item	Scenario I	Scenario II	Scenario III
Total capital cost (USD)	4,066,800	4,061,680	4,056,420
Total operating cost (USD/yr)	2,392,780	2,582,770	2,692,689
Total raw materials cost (USD/yr)	1,332,180	1,332,180	1,332,180
Total product sales (USD/yr)	4,920,570	14,177,900	4,635,550
Payback period (yr)	5.2	1.5	5.9
Equipment cost (USD)	66,900	66,900	66,900
Total installed cost (USD)	510,300	507,500	504,800

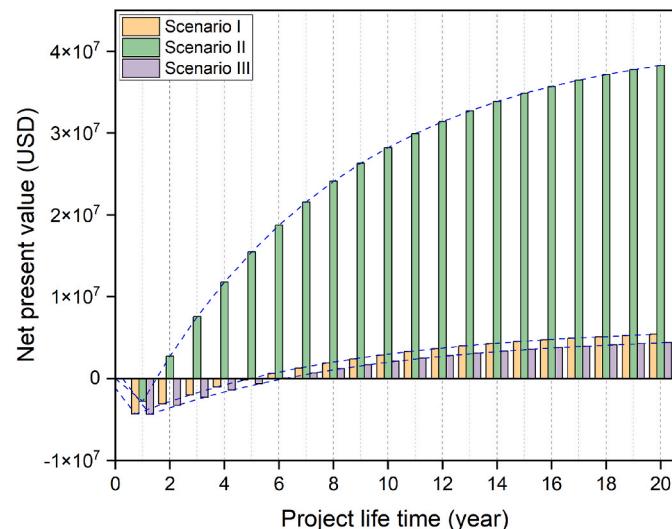


Fig. 9. Annual net present value of different scenarios over the project's lifetime.

compared to air-assisted and CO₂-assisted approaches, owing to the promotion of steam reforming and water gas shift reactions. Therefore, scenario II emerges as the most economically favorable option among the studied scenarios, underscoring the critical role of steam as a gasification agent in enhancing product sales and overall process profitability.

The net present value serves as a key financial metric for assessing investment profitability, comparing the present value of cash inflows to the initial investment. A positive net present value suggests that the project is expected to generate more cash inflows than outflows throughout its lifespan. Conversely, the payback period represents the duration required for the project to recover its initial investment. In this study, scenario II, involving steam-assisted co-gasification of biomass and coal, exhibits superior performance in both net present value and payback period compared to other scenarios. With a net present value of 38,296,700 USD over 20 years and a relatively shorter payback period, scenario II emerges as the most financially viable option among the alternatives. This superiority can be attributed to its capacity for producing a higher volume of H₂ as the final product, leading to increased revenue generation and quicker return on investment. While all scenarios exhibit positive net present values, indicating the financial feasibility of the designed plants, scenario II stands out as the most profitable owing to its optimized co-gasification conditions and enhanced product yields (Fig. 9).

4. Conclusions

In this study, an ML model was developed to predict the biomass-coal co-gasification process, encompassing product yield and syngas quality. A comprehensive database containing a wide array of biomass and coal compositions, as well as biomass/coal blending ratios under various

operational parameters, was assembled from published literature. The dataset underwent preprocessing steps, including imputation of missing data, outlier removal, and normalization, before being introduced to ML modeling techniques such as KRR, DTR, RFR, and GBR. The GBR model showed the best performance, with R² values of 0.98, 0.93, 0.96, 0.94, 0.92, and 0.99 for syngas yield, H₂, CO₂, CH₄, CO content, and syngas LHV, respectively. Additionally, the GBR model exhibited the lowest statistical errors across all outputs, with MAE and RMSE values below 2.92 and 3.39, respectively. SHAP analysis identified the biomass/coal blending ratio and temperature as the most influential parameters in the co-gasification process for syngas production, with coal hydrogen content having the greatest influence on H₂ production.

Multi-objective optimization results indicated that higher temperatures (above 1800 °C) are conducive to H₂ production, while lower temperatures (around 1400 °C) favor the Fischer-Tropsch process. A subsequent techno-economic analysis of the biomass-coal co-gasification process based on the optimum conditions suggested by multi-objective optimization was performed. This analysis provided insights into the commercialization potential of the process. The techno-economic analysis was carried out for three scenarios using air, steam, and CO₂ as gasifying agents. All three scenarios required almost the same fixed costs and total raw material costs. However, the outcomes indicated that using steam as the gasifying agent exhibited superior performance in both net present value and a shorter payback period compared to other scenarios. Using steam as the gasifying agent resulted in an exceptionally high internal rate of return of 243 %, a return rate of 86.9 %, and a profitability index of 1.87.

While the developed model and its optimized conditions provide a promising foundation for industrial biomass-coal co-gasification, there are several avenues for further research. The growing interest in catalytic biomass-coal co-gasification, driven by the beneficial effects of catalysts in lowering gasification temperatures and enhancing syngas quality, suggests that future efforts could focus on modeling and optimizing this advanced process using ML technology. Additionally, if experimental data on the ash composition of both biomass and coal become available, incorporating the effect of ash as an internal catalyst material could lead to more accurate and comprehensive ML models for biomass-coal co-gasification.

Future research should address the impact of alkaline earth metals on biomass gasification. The presence and concentration of these metals (e.g., calcium and magnesium) can significantly influence gasification reactions, affecting the optimization of conditions, such as gasifying agent, temperature, pressure, and feedstock ratios. However, due to insufficient data on alkaline earth metal content in existing literature, these variables were not included as input factors in our ML models. Future studies should prioritize collecting comprehensive data on these metals to enhance the accuracy of predictive models and optimize gasification processes more effectively.

Another potential direction for future ML modeling in this field can focus on reactor types and geometry characteristics, provided this information is readily available in experimental studies. Exploring the effects of biomass constituents (such as cellulose, hemicellulose, and lignin) on biomass-coal co-gasification and developing models based on these constituents can be another interesting area of study. Furthermore, assessing the environmental impact of biomass-coal co-gasification by leveraging the achievements of the current study and conducting a life cycle assessment analysis can be appealing for future studies. This comprehensive approach can provide valuable insights into the environmental implications of biomass-coal co-gasification, contributing to the development of more sustainable energy conversion technologies.

CRediT authorship contribution statement

Junting Pan: Writing – original draft, Investigation. **Hossein Shahbeik:** Writing – original draft, Formal analysis. **Alireza Shafizadeh:** Writing – original draft, Methodology. **Shahin Rafiee:** Validation,

Software. Milad Golvirdizadeh: Investigation, Data curation. Seyyed Alireza Ghafarian Nia: Visualization, Software. Hossein Mobli: Validation, Resources. Yadong Yang: Visualization, Formal analysis. Guilong Zhang: Writing – review & editing, Funding acquisition, Conceptualization. Meisam Tabatabaei: Writing – review & editing, Project administration, Conceptualization. Mortaza Aghbashlo: Writing – review & editing, Supervision, Conceptualization.

Declaration of competing interest

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

Acknowledgments

Guilong Zhang would like to thank the financial support of “National Key R&D Program of China (2023YFD1701800)”. Junting Pan would like to acknowledge the Earmarked fund for CARS (CARS-06-14.5-A31), Agricultural Science and Technology Innovation Program and Youth Talent Scholar of Chinese Academy of Agricultural Sciences. This work is also supported by the Ministry of Higher Education, Malaysia, under the Higher Institution Centre of Excellence (HICoE), Institute of Tropical Aquaculture and Fisheries (AKUATROP) program. The authors would also like to extend their sincere appreciation to the University of Tehran for the support provided throughout this project.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.renene.2024.120772>.

References

- [1] A. Alegre-Bravo, C.L. Anderson, Exploring the influence of multidimensional variables on access to electricity in rural areas of the Global South, *Appl. Energy* 333 (2023) 120509, <https://doi.org/10.1016/j.apenergy.2022.120509>.
- [2] K. Islam, R. Yokoi, M. Motoshita, S. Murakami, Ecological footprint accounting of mining areas and metal production of the world, *Resour. Conserv. Recycl.* 183 (2022) 106384, <https://doi.org/10.1016/j.resconec.2022.106384>.
- [3] L.E. Taba, M.F. Irfan, W.A.M.W. Daud, M.H. Chakrabarti, The effect of temperature on various parameters in coal, biomass and CO₂-gasification: a review, *Renew. Sustain. Energy Rev.* 16 (2012) 5584–5596, <https://doi.org/10.1016/j.rser.2012.06.015>.
- [4] J. Rizkiana, G. Guan, W.B. Widayatno, X. Hao, W. Huang, A. Tsutsumi, A. Abudula, Effect of biomass type on the performance of cogenasification of low rank coal with biomass at relatively low temperatures, *Fuel* 134 (2014) 414–419, <https://doi.org/10.1016/j.fuel.2014.06.008>.
- [5] M. Bharath, V. Raghavan, B. Prasad, S.R. Chakravarthy, Co-gasification of Indian rice husk and Indian coal with high-ash in bubbling fluidized bed gasification reactor, *Appl. Therm. Eng.* 137 (2018) 608–615, <https://doi.org/10.1016/j.aplthermaleng.2018.04.035>.
- [6] M. Shahabuddin, S. Bhattacharya, Enhancement of performance and emission characteristics by co-gasification of biomass and coal using an entrained flow gasifier, *J. Energy Inst.* 95 (2021) 166–178, <https://doi.org/10.1016/j.joei.2021.01.012>.
- [7] I. Rosyadi, S. Suyitno, A.X. Ilyas, A. Faishal, A. Budiono, M. Yusuf, Producing hydrogen-rich syngas via microwave heating and co-gasification: a systematic review, *Biofuel Res. J.* 9 (2022) 1573–1591, <https://doi.org/10.18331/BRJ2022.9.1.4>.
- [8] H. Shahbeik, W. Peng, H. Kazemi Shariat Panahi, M. Dehhaghi, G.J. Guillemin, A. Fallahi, H. Amiri, M. Rehan, D. Raikwar, H. Latine, B. Pandalone, B. Khoshnevisan, C. Sonne, L. Vaccaro, A.-S. Nizami, V.K. Gupta, S.S. Lam, J. Pan, R. Luque, B. Sels, M. Tabatabaei, M. Aghbashlo, Synthesis of liquid biofuels from biomass by hydrothermal gasification: a critical review, *Renew. Sustain. Energy Rev.* 167 (2022) 112833, <https://doi.org/10.1016/j.rser.2022.112833>.
- [9] J. Wei, M. Wang, F. Wang, X. Song, G. Yu, Y. Liu, H. Vuthaluru, J. Xu, Y. Xu, H. Zhang, A review on reactivity characteristics and synergy behavior of biomass and coal Co-gasification, *Int. J. Hydrogen Energy* 46 (2021) 17116–17132, <https://doi.org/10.1016/j.ijhydene.2021.02.162>.
- [10] X. Yang, X. Liu, R. Li, C. Liu, T. Qing, X. Yue, S. Zhang, Co-gasification of thermally pretreated wheat straw with Shengli lignite for hydrogen production, *Renew. Energy* 117 (2018) 501–508, <https://doi.org/10.1016/j.renene.2017.10.055>.
- [11] B. Cabuk, G. Duman, J. Yanik, H. Olgun, Effect of fuel blend composition on hydrogen yield in co-gasification of coal and non-woody biomass, *Int. J. Hydrogen Energy* 45 (2020) 3435–3443, <https://doi.org/10.1016/j.ijhydene.2019.02.130>.
- [12] Y.-C. Song, M.-S. Ji, J. Feng, W.-Y. Li, Product distribution from Co-gasification of coal and biomass in a fluidized-bed reactor, *Energy Sources, Part A Recover. Util. Environ. Eff.* 37 (2015) 2550–2558, <https://doi.org/10.1080/15567036.2012.711798>.
- [13] S. Chang, Z. Zhang, L. Cao, L. Ma, S. You, W. Li, Co-gasification of digestate and lignite in a downdraft fixed bed gasifier: effect of temperature, *Energy Convers. Manag.* 213 (2020) 112798, <https://doi.org/10.1016/j.enconman.2020.112798>.
- [14] L. Yan, Y. Cao, X. Li, B. He, Characterization of a dual fluidized bed gasifier with blended biomass/coal as feedstock, *Bioresour. Technol.* 254 (2018) 97–106, <https://doi.org/10.1016/j.biotech.2018.01.067>.
- [15] H. Li, J. Chen, W. Zhang, H. Zhan, C. He, Z. Yang, H. Peng, L. Leng, Machine-learning-aided thermochemical treatment of biomass: a review, *Biofuel Res. J.* 10 (2023) 1786–1809, <https://doi.org/10.18331/BRJ2023.10.1.4>.
- [16] Q. Yang, J. Zhang, J. Zhou, L. Zhao, D. Zhang, A hybrid data-driven machine learning framework for predicting the performance of coal and biomass gasification processes, *Fuel* 346 (2023) 128338, <https://doi.org/10.1016/j.fuel.2023.128338>.
- [17] S. Tiwary, S.B. Ghugare, P.D. Chavan, S. Saha, S. Datta, G. Sahu, S.S. Tambe, Co-Gasification of high ash coal-biomass blends in a fluidized bed gasifier: experimental study and computational intelligence-based modeling, *Waste and Biomass Valorization* 11 (2020) 323–341, <https://doi.org/10.1007/s12649-018-0378-7>.
- [18] A. Bahadar, R. Kanthasamy, H.H. Sait, M. Zwawi, M. Algarni, B.V. Ayodele, C. K. Cheng, L.J. Wei, Elucidating the effect of process parameters on the production of hydrogen-rich syngas by biomass and coal Co-gasification techniques: a multi-criteria modeling approach, *Chemosphere* 287 (2022) 132052, <https://doi.org/10.1016/j.chemosphere.2021.132052>.
- [19] J. Qi, Y. Wang, P. Xu, M. Hu, T. Huhe, X. Ling, H. Yuan, Y. Chen, Study on the Co-gasification characteristics of biomass and municipal solid waste based on machine learning, *Energy* 290 (2024) 130178, <https://doi.org/10.1016/j.energy.2023.130178>.
- [20] K. Rakesh, P.N. Suganthan, An ensemble of Kernel Ridge Regression for multi-class classification, *Procedia Comput. Sci.* 108 (2017) 375–383, <https://doi.org/10.1016/j.procs.2017.05.109>.
- [21] M.J.S. Safari, S. Rahimzadeh Arashloo, Kernel ridge regression model for sediment transport in open channel flow, *Neural Comput. Appl.* 33 (2021) 11255–11271, <https://doi.org/10.1007/s00521-020-05571-6>.
- [22] B. An, M. Liang, T. Chang, X. Duan, L. Du, L. Xu, L. Zhang, X. Gao, J. Li, H. Gao, KCRR: a nonlinear machine learning with a modified genomic similarity matrix improved the genomic prediction efficiency, *Briefings Bioinf.* 22 (2021) bbab132, <https://doi.org/10.1093/bib/bbab132>.
- [23] M.S. Chowdhury, M.N. Rahman, M.S. Sheikh, M.A. Sayeid, K.H. Mahmud, B. Hafsa, GIS-based landslide susceptibility mapping using logistic regression, random forest and decision and regression tree models in Chatogram District, Bangladesh, *Heliyon* 10 (2024) e23424, <https://doi.org/10.1016/j.heliyon.2023.e23424>.
- [24] N.M. Kassim, S. Santhiran, A.A. Alkahtani, M.A. Islam, S.K. Tiong, M.Y. Mohd Yusof, N. Amin, An adaptive decision tree regression modeling for the output power of large-scale solar (LSS) farm forecasting, *Sustainability* 15 (2023) 13521, <https://doi.org/10.3390/su151813521>.
- [25] Z. Zhou, C. Qiu, Y. Zhang, A comparative analysis of linear regression, neural networks and random forest regression for predicting air ozone employing soft sensor models, *Sci. Rep.* 13 (2023) 22420, <https://doi.org/10.1038/s41598-023-49899-0>.
- [26] V. Rodriguez-Galiano, M. Sanchez-Castillo, M. Chica-Olmo, M. Chica-Rivas, Machine learning predictive models for mineral prospectivity: an evaluation of neural networks, random forest, regression trees and support vector machines, *Ore Geol. Rev.* 71 (2015) 804–818, <https://doi.org/10.1016/j.oregeorev.2015.01.001>.
- [27] K. Were, D.T. Bui, Ø.B. Dick, B.R. Singh, A comparative assessment of support vector regression, artificial neural networks, and random forests for predicting and mapping soil organic carbon stocks across an Afromontane landscape, *Ecol. Indicat.* 52 (2015) 394–403, <https://doi.org/10.1016/j.ecolind.2014.12.028>.
- [28] R.L. Bailey, M.A. Reiss, C.N. Arge, C. Möstl, C.J. Henney, M.J. Owens, U. V. Amerstorfer, T. Amerstorfer, A.J. Weiss, J. Hinterreiter, Using gradient boosting regression to improve ambient solar wind model predictions, *Space Weather* 19 (2021) e2020SW002673, <https://doi.org/10.1029/2020SW002673>.
- [29] A.M. Peco Chacón, I. Segovia Ramírez, F.P. García Márquez, K-nearest neighbour and K-fold cross-validation used in wind turbines for false alarm detection, *Sustain. Futur.* 6 (2023) 100132, <https://doi.org/10.1016/j.sfr.2023.100132>.
- [30] P.R. Jeon, J.-H. Moon, N.O. Ogunsola, S.H. Lee, J.L.J. Ling, S. You, Y.-K. Park, Recent advances and future prospects of thermochemical biofuel conversion processes with machine learning, *Chem. Eng. J.* 471 (2023) 144503, <https://doi.org/10.1016/j.cej.2023.144503>.
- [31] S.M. Lundberg, S.-I. Lee, A unified approach to interpreting model predictions, in: *NIPS'17 Proc. 31st Int. Conf. Neural Inf. Process. Syst., California, 2017*, pp. 4768–4777.
- [32] I.U. Ekanayake, D.P.P. Meddage, U. Rathnayake, A novel approach to explain the black-box nature of machine learning in compressive strength predictions of concrete using Shapley additive explanations (SHAP), *Case Stud. Constr. Mater.* 16 (2022) e01059, <https://doi.org/10.1016/j.cscm.2022.e01059>.
- [33] L. Corbucci, R. Guidotti, A. Monreale, Explaining black-boxes in federated learning, in: *World Conf. Explain. Artif. Intell.*, 2023, pp. 151–163, https://doi.org/10.1007/978-3-031-44067-0_8.
- [34] Y. Yang, H. Shahbeik, A. Shafizadeh, S. Rafiee, A. Hafezi, X. Du, J. Pan, M. Tabatabaei, M. Aghbashlo, Predicting municipal solid waste gasification using machine learning: a step toward sustainable regional planning, *Energy* 278 (2023) 127881, <https://doi.org/10.1016/j.energy.2023.127881>.

- [35] A. Shehzad, M.J.K. Bashir, S. Sethupathi, System analysis for synthesis gas (syngas) production in Pakistan from municipal solid waste gasification using a circulating fluidized bed gasifier, *Renew. Sustain. Energy Rev.* 60 (2016) 1302–1311, <https://doi.org/10.1016/j.rser.2016.03.042>.
- [36] M. Shahbaz, T. Al-Ansari, M. Inayat, S.A. Sulaiman, P. Parthasarathy, G. McKay, A critical review on the influence of process parameters in catalytic co-gasification: current performance and challenges for a future prospectus, *Renew. Sustain. Energy Rev.* 134 (2020) 110382, <https://doi.org/10.1016/j.rser.2020.110382>.
- [37] D. Mallick, P. Mahanta, V.S. Moholkar, Co-gasification of biomass blends: performance evaluation in circulating fluidized bed gasifier, *Energy* 192 (2020) 116682, <https://doi.org/10.1016/j.energy.2019.116682>.
- [38] J. Watson, Y. Zhang, B. Si, W.-T. Chen, R. de Souza, Gasification of biowaste: a critical review and outlooks, *Renew. Sustain. Energy Rev.* 83 (2018) 1–17, <https://doi.org/10.1016/j.rser.2017.10.003>.
- [39] X. Li, J. Wu, J. Yu, Z. Zhou, Q. Wang, W. Zhao, L. Hu, Inversion of soil moisture content in cotton fields using GBR-RF algorithm combined with sentinel-2 satellite spectral data, *Agronomy* 14 (2024) 784, <https://doi.org/10.3390/agronomy14040784>.