



Predicting and optimizing syngas production from fluidized bed biomass gasifiers: A machine learning approach

Jun Young Kim^{a,b,*¹}, Dongjae Kim^{c,d,1}, Zezhong John Li^e, Claudio Dariva^f, Yankai Cao^g, Naoko Ellis^g

^a Institute of Convergent Chemical Engineering and Technology, Sungkyunkwan University (SKKU), Suwon, Gyeonggi-do 16419, Republic of Korea

^b School of Chemical Engineering, Sungkyunkwan University, 2066 Seobu-Ro, Jangan, Suwon 16419, Republic of Korea

^c Department of Chemical Engineering, Soochunhyang University, 22 Soochunhyang-ro, Shinchang-myeon, Asan-si 31538, Chungcheongnam-do, Republic of Korea

^d Department of Electronic Materials and Devices Engineering, Soochunhyang University, 22 Soochunhyang-ro, Shinchang-myeon, Asan-si 31538, Chungcheongnam-do, Republic of Korea

^e Laboratory of Sustainable and Catalytic Processing, Institute of Chemical Sciences and Engineering, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne 1015, Switzerland

^f Center for Study on Colloidal Systems (NUESC), Institute of Technology and Research (ITP), PEP/PBI, Tiradentes University (UNIT), Aracaju-SE, CEP 49032-490, Brazil

^g Department of Chemical & Biological Engineering, University of British Columbia, 2360 East Mall, Vancouver, B.C. V6T 1Z3, Canada

ARTICLE INFO

Keywords:

Random forest
Artificial neural network
Support vector machine
Monte Carlo Filtering
Biomass gasification
Machine learning

ABSTRACT

Biomass gasification is one of the primary thermal conversion processes where fluidized bed reactors are often used to produce syngas with low heating values. However, there has not yet been an effective model to predict gasification yield with broad applicability. In this study, machine learning was adopted to realize the prediction of syngas compositions and lower heating values (LHV) using various lignocellulosic biomass feedstocks at a wide range of operating conditions. Three machine learning techniques, i.e., Random Forest (RF), Support Vector Machine (SVM) and Artificial Neural Network (ANN) were adopted after determining hyperparameters optimization. Pearson correlation and permutation importance were used for the sensitivity analysis. RF and ANN were found to have high prediction accuracy with R^2 and RMSE results (RF: $R^2=0.809\text{--}0.946$, RMSE=1.39–11.54%; ANN: $R^2=0.565\text{--}0.924$, RMSE=1.46–10.56%). Monte Carlo filtering (MCF) was integrated into the three machine learning algorithms to forecast the desired products by predicting the important features of the operating conditions and biomass characteristics. Considering the desired $\text{H}_2/\text{CO} > 1.1$ and $LHV > 5.86 \text{ MJ/m}^3$, the RF-MCF was a more suitable approach with $R^2=0.791\text{--}0.902$ for H_2 , CO and LHV features. The machine learning approach can be widely adapted in various scenarios predicting output features as well as MCF for finding the significant variables for optimization.

1. Introduction

As a result of increasing energy demand caused by economic and population growth as well as increasing attention on environmental issues, efforts in replacing fossil fuels with renewable sources are gaining more attention. Biomass is one of the most abundant raw materials on the Earth, a promising alternative to conventional fossil fuels [1]. Lignocellulosic biomass can be considered as one of the renewable carbonaceous resources with huge reserves, capable of producing heat

and electricity via combustion and gaseous/liquid biofuels with thermochemical conversion approaches such as gasification or pyrolysis [2]. Amongst these conversion technologies, biomass gasification is a highly efficient thermochemical process that converts biomass feedstocks to syngas, mainly composed of H_2 , CO, CO_2 and CH_4 , which can be further used in fuel or chemical production [1].

There are several types of gasification reactors such as fixed bed, fluidized bed and entrained flow gasifier [3–5], among which fluidized bed biomass gasifier became one of the most mature technologies after being commercialized over 50 years [6]. A fluidized gasifier has the

* Corresponding author. Institute of Convergent Chemical Engineering and Technology, Sungkyunkwan University (SKKU), Suwon, Gyeonggi-do 16419, Republic of Korea

E-mail address: june0112@g.skku.edu (J.Y. Kim).

¹ first authors.

Abbreviations

ANN	Artificial neural network
CDF	Cumulative probability distribution function
ER	Equivalence ratio
LHV	Lower heating value
MCF	Monte Carlo filtering
RF	Random forest
RMSE	Root mean square error
RBF	Radial based function
SBR	Steam to biomass ratio
SVM	Support vector machine

Nomenclature

d	Kolmogorov-Smirnov statistic
$F_n(x)$	Empirical distribution function
l	Cross-validation loss
N	Number of datasets
R^2	Coefficient of determination
α	Significance level

advantages of high mass production capability, good solids mixing and fast heat transfer compared to fixed bed or moving bed reactors [7–11].

The syngas composition, lower heating value (*LHV*) and char/tar yield are key performance metrics that heavily depend on biomass compositions and reactor feeding gas, as well as operating conditions. Comprehensive understandings of those effects are of great significance for designing a fluidized bed biomass gasifier. There are many experimental attempts to elucidate these effects. McKendry [12] summarized the ranges of lignocellulose contents with four types of biomass (soft-wood, hardwood, wheat straw and switchgrass) and claimed that the high lignin content is not prone to convert to syngas. Different lignocellulose components can vary the kinetic characteristics, thus influencing the biomass gasification behaviour [13,14]. Molino et al. [15] summarized the effect of feedstock characteristics and operating conditions on the performance of the biomass gasification process. By increasing the equivalence ratio (*ER*) ranges from 0.2 to 0.3, H₂ and CO concentrations decrease while CO₂ concentration increases and a reduction of syngas heating value was observed [16]. Also, tar cracking can be promoted by higher *ER* ranges from 0.6 to 0.64, due to more O₂ available for tar reforming reactions [17]. In terms of steam to biomass ratio (*SBR*) from 0 to 1.5 at 800 °C, increasing *SBR* increases the concentration of H₂ and CO₂ as well as increases the heating value of syngas [18]. On the other hand, under the *SBR* ranges of 0.3–1.0, increasing *SBR* decreases CO and tar concentration because of water-gas shift, reforming and cracking reactions, which are promoted by steam [19, 20]. Although many experimental findings showed the qualitative trend of how operating conditions and feedstock compositions affect syngas characteristics, these experiments were done in distinct and narrow operating windows. Besides, some results from the previous research did not distinguish the type of reactors (i.e., fixed bed and fluidized bed) which was significantly affected by the operating conditions such as superficial gas velocity. This work will be limited to the bubbling fluidized bed system to reduce inhomogeneity in reactor hydrodynamics.

Machine learning is one of the advanced approaches dealing with nonlinear and multivariable systems that have shown excellent performance across various fields. Generally, artificial neural network (ANN), random forest (RF), and support vector machine (SVM) are reported having a supervised learning concept, able to predict syngas compositions by classifying unlabeled data as a certain class using related inputs (features) [21,22]. These machine learning algorithms have been successfully applied in many related fields, including biomass devolatilization kinetics [13], hydrodynamics of circulating fluidized bed riser

Table 1

Statistical information of the input and output variables of the constructed database.

Variables	Ranges	
Input variables		
Lignocellulose composition [wt.%]	Cellulose (<i>Cell.</i>)	0.20–0.58
	Hemicellulose (<i>Hem.</i>)	0.08–0.63
	Lignin	0.10–0.49
Temperature (<i>T</i>) [°C]		600–900
Pressure (<i>P</i>) [abar]		1–10
Equivalence ratio (<i>ER</i>) [–]		0–0.86
Steam to biomass ratio (<i>SBR</i>) [–]		0–8.03
Superficial gas velocity (<i>U_g</i>) [m/s]		0.02–9.59
Output variables		
Syngas composition [vol%]	H ₂	5.39–66.03
	CO	5.01–55.44
	CO ₂	6.78–62.56
	CH ₄	1.31–20.1
Lower heating value (<i>LHV</i>) [MJ/Nm ³]		1.74–15.0
Char yield [wt.%]		0–45
Tar yield [g/Nm ³]		0–134.1

with cold-flow model [23,24], and parameter tuning for CFD simulations [25]. Xing et al. compared empirical correlation, ANN and RF to predict the kinetic parameters of biomass devolatilization [13]. Souza et al. developed an ANN model to correlate operating conditions of the biomass gasifier (steam to biomass ratio, temperature and ultimate analysis of biomass) with the produced gas compositions, including syngas and hydrocarbons [26]. They used lab and pilot-scale experiments tested at atmospheric pressure. Serrano et al. applied ANN to predict the effect of bed materials in bubbling fluidized bed gasification [27]. The input variables were the ultimate analysis results of biomass, equivalent ratio, temperature, steam to biomass ratio and 4 different types of bed materials, whereas the output features were gas yield and syngas composition.

On the basis of the above background, the objective of this research is to find the complex and nonlinear correlations between the input parameters (i.e., lignocellulosic compositions and operating conditions) and output variables (i.e., syngas compositions and byproduct properties) through accurate and efficient machine learning models, including RF, SVM and ANN methods. One emphasis is to allow a relatively wide range of the input parameters while keeping the inherent inhomogeneity low. Such a model can be useful, for example, when screening operating conditions for a certain feedstock to yield expected product features. The performances of these models were compared to highlight the optimal input parameters affecting H₂/CO ratio in syngas at the same time maintaining low tar/char formation level, using sequential Monte Carlo filtering method as sensitivity analysis. Also, Monte Carlo filtering (MCF) was integrated into the machine learning algorithm to forecast the desired products via predicting the important features of the operating conditions and biomass characteristics. To the best of our knowledge, this is the first study that uses Monte Carlo Filtering as a way to predict operating conditions for producing the desired syngas composition with high LHV (lower heating value) in a fluidized bed.

2. Materials and methods

2.1. Dataset collection

Since biomass gasification in a fluidized bed has a significant difference from the one in a fixed bed in terms of temperature distribution, particle size range and tar formation [28], datasets were carefully collected from various reliable studies in the literature, focusing on the experimental results with fluidized beds. Overall, the dataset used in this study has 336 samples [29–67]. Here, all data given from the references were conducted in steady-state operation. All data contain lignocellulosic compositions, operating conditions (temperature, pressure and

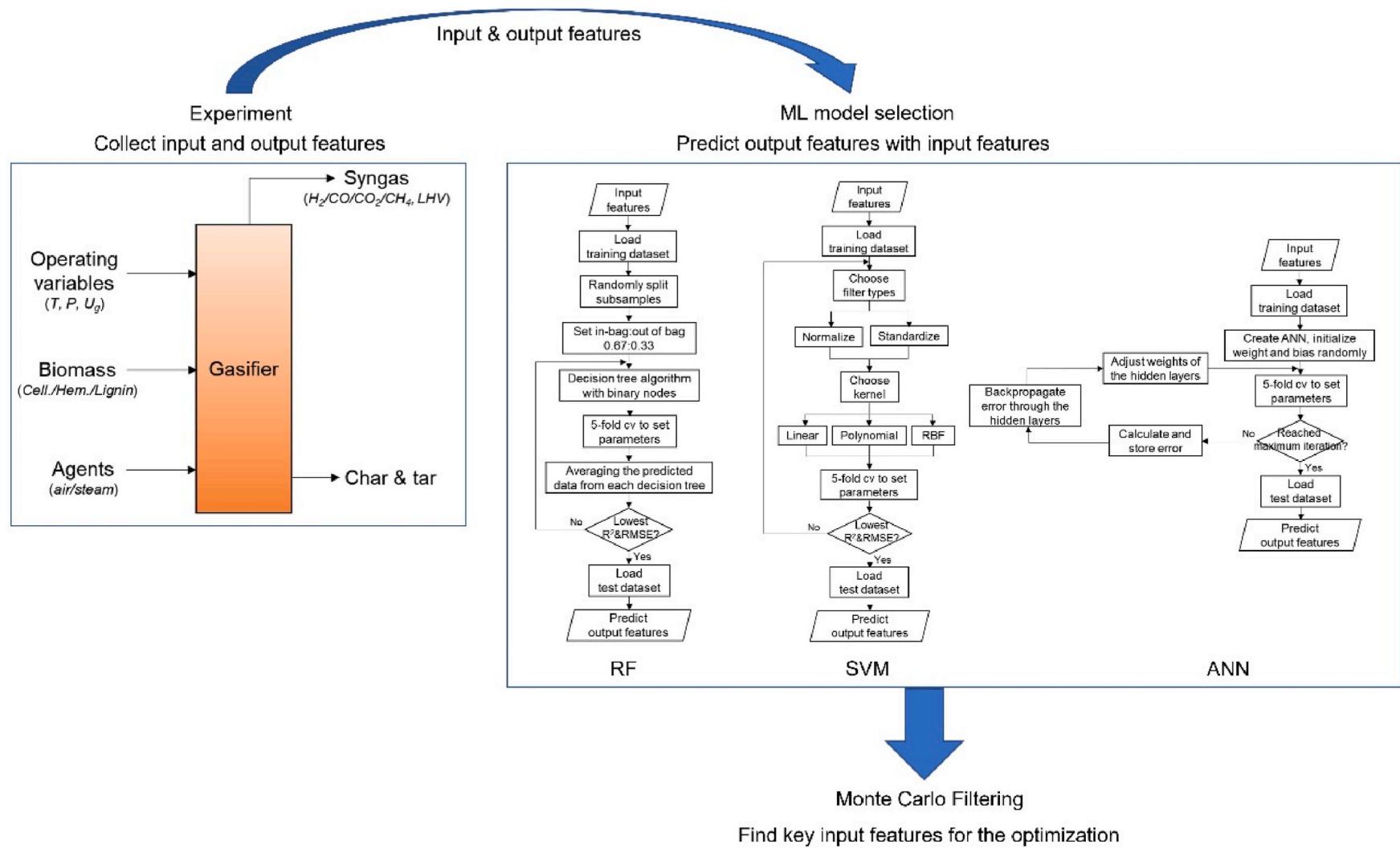


Fig. 1. Schematic diagram of adapting machine learning algorithm from the experimental data, and Monte Carlo Filtering for finding key input features for the optimization.

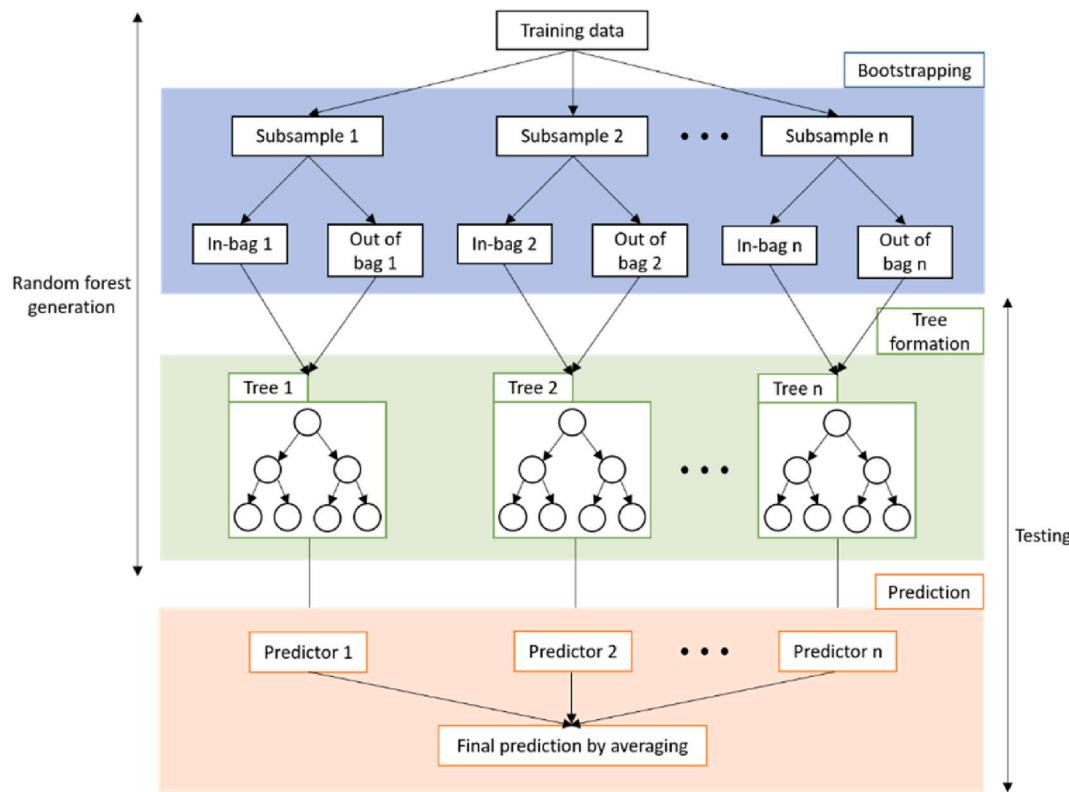


Fig. 2. Schematic structure of random forest algorithm.

superficial gas velocity), equivalence ratio and steam to biomass ratio as input variables, and syngas compositions (H_2 , CO , CO_2 and CH_4 without N_2 and dry basis), lower heating value (LHV), char and tar yield as the output variables whose ranges are shown in Table 1. Instead of the ultimate analysis of the biomass, the proximate analysis dataset was chosen because the lignocellulosic components are more likely to affect the syngas composition. The elemental compositions of cellulose, hemicellulose and lignin are similar across different woody biomass species. For example, lignin-rich biomass is known to have aromatic polymers with a higher C/O ratio that favours tar formation [68]. Note that the biomass fuels include both woody biomass and agricultural residues which have distinctive lignocellulose compositions. Details of the datasets are provided in **Supplementary information 1 (SI 1)**. These operating conditions as input variables can represent the complex variables that need to be considered for the fluidized bed gasifier, such as hydrodynamic performance and reactions. For example, the residence time can be modulated with the gas properties and particle properties which are designated to superficial gas velocity, temperature, pressure, *ER* and *SBR* as an element. The catalysts used in the dataset are limited to tar removal during biomass gasification in the bubbling fluidized bed [69]. Although the details of these tar removal catalyst mechanisms may be different, we have focused on the role of tar removal during the gasification in the fluidized bed system, not the specific reaction kinetics or catalytic sites on tar removal since these are outside of the scope.

A schematic diagram on the details of the machine learning approaches is given in Fig. 1. With the input features for the biomass composition, operating conditions and experimental results from the given dataset, three perspective machine learning models (RF, SVM and ANN) were carefully selected with 5-fold cross-validation and optimization. Details of each machine learning algorithm are given in each section below, respectively. Also, the Monte Carlo Filtering (MCF) was adapted to find the key input features for predicting the optimized H_2/CO ratio and LHV of the products.

2.2. Random forest

Random forest (RF) is an ensemble type of machine learning method built upon the multitude of decision trees, especially suitable for classification and regression [70]. A brief structure of this algorithm is presented in Fig. 2. The random forest can control the variance with the bootstrap aggregating method (bootstrapping), as suggested by Breiman [70]. For example, with n samples and m features, bootstrapping randomly generates n subsamples for each tree which are randomly split into in-bag or out-of-bag data with approximately 0.67:0.33 ratio [71, 72]. Here, the in-bag data are used to derive the regression function. The out-of-bag data provides new training samples when the decision tree is reiterated until the trees grow without pruning. The best trees with an adequate split can be determined by the number of m features which are randomly selected during the training. m is used as a tuning parameter and is kept less than the number of decision trees. The ability to generalize the random forests using independent test data is enhanced with these built-in validation features.

The RF calculation was performed on the MATLAB software (R2020b) with '*fittrensemble*' code for optimizing regression ensembles, using Bayesian Optimization. The optimization was explored over two methods: *Bag* (Bootstrap aggregating) and *LSBoost* (Least square boosting). The hyperparameters such as the number of learning cycles, learning rates (for *LSBoost*), maximum and minimum leaf size were automatically optimized via minimizing five-fold cross-validation loss. Herein, the data were divided into 5 arbitrary sets among which one was selected for testing with the remaining for training. Note that the five-fold cross-validation is performed only on the training data. Such training-testing process was repeated such that each set was taken as the testing set once and all the data were utilized adequately. The objective function to be minimized was $\log(1+l)$ where l is a cross-validation loss. The maximum number of evaluations of the objective function was 30. The ensemble random forest was finally determined by choosing the hyperparameters resulting in a minimum objective function value.

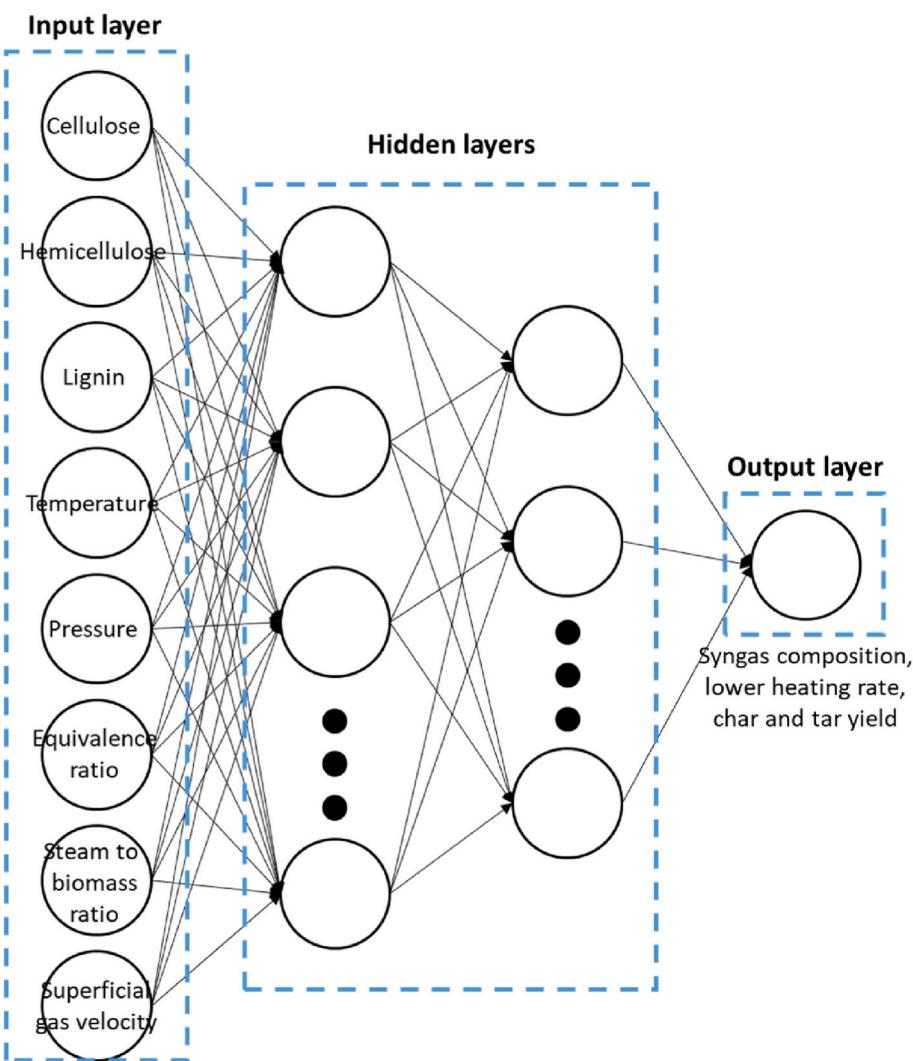


Fig. 3. Schematic ANN architecture with 8 inputs and 1 output. The numbers of hidden layers are illustrative only.

Randomly selected 70% of the data were used for training and the other 30% were used for testing.

2.3. Support vector machine

Support Vector Machine (SVM) regression is another supervised machine learning algorithm with low prediction bias and high variance. For simplicity, the Support Vector Machine regression is referred to as ‘SVM’ in this work. The SVM exploits inequality type constraints to optimize the quadratic function of variables [22]. Similar to the RF, SVM can also be used in both classification and regression. Moreover, SVM generally has an advantage of reducing overfitting over the artificial neural network (ANN) because SVM minimizes structural risk while ANN minimizes the empirical risk or training error by tuning more hyperparameters [73]. In SVM, the error function to be minimized is convex, meaning that the global optimum is *always* reached [74].

Similar to the RF, SVM was also optimized with the built-in Matlab code ‘*fittsrm*’. Generally, kernel functions are used for mapping the input space to a high-dimensional space to enhance the learning performance. There are various kernel functions: linear, polynomial, sigmoid, and radial based function (*RBF*). Based on empirical knowledge and preliminary tests, RBF showed the best performance for the given dataset. Hyperparameters such as kernel scale, half width of the ϵ -insensitivity (within the region where the error is disregarded), and *BoxConstraint* (the higher value resulting in the higher cost for data with large error)

were found by minimizing the five-fold cross-validation loss. Similar to the RF calculation, a sequential minimal optimization was used, and the objective function was $\log(1+l)$ where l is a cross-validation loss. The randomly selected 70% of data were used for training and the other 30% were used for tests. Details of this optimization method can be found elsewhere [75,76].

2.4. Artificial neural network

A schematic topological architecture of ANN model is described in Fig. 3. ANN is composed of an input layer, a number of hidden layers and an output layer, with a modifiable number of ‘neurons’ per layer. Throughout the training process, the weights and biases in the individual neurons are iteratively adjusted so that the overall network reduces the sum of the squared differences between predicted outputs and observed outputs given a known set of training data (observation). Since ANN can discover relationships between inputs and outputs of a system without the need for a detailed mechanistic understanding, it can be effectively applied in systems where such relationships are not clearly understood or very complex, such as biomass gasification [77].

The ANN was trained by MATLAB. Note that the number of hidden layers and neurons in each layer is a crucial hyperparameter for optimizing ANN. To set the ANN architectures, 5-fold cross-validation was used to set the optimum numbers of the node in the hidden layers to avoid overfitting. Also, the Bayesian backpropagation algorithm was

Table 2

Best-selected topologies including different layers and neurons in ANN.

Output parameters	Number of nodes in hidden layer 1	Number of nodes in hidden layer 2	RSME
H ₂	8	2	3.56
CO	4	4	3.88
CO ₂	8	4	3.69
CH ₄	4	4	1.61
LHV	2	4	1.46
Char	4	4	5.41
Tar	2	8	10.6

used for optimizing the weights of the ANN. The best-selected topologies including different layers and neurons in ANN are provided in [Table 2](#). We chose the ANN with the smallest root mean square error for test set data. Here, randomly selected 70% of the data were used for training, and 30% were used for the tests, same as the other two machine learning algorithms.

2.5. Model assessment and feature evaluation

To assess the prediction capability of these models, many datasets obtained from previously published literature were randomly selected for training and testing. The combination set that provided the optimal model performance was considered as the best hyperparameters for the model. Each of these machine learning models was statistically assessed with regression coefficient (R^2) and root mean square error (RMSE) [2]:

$$R^2 = 1 - \sum_{i=1}^N (y_{i,prediction} - y_{i,data})^2 / \sum_{i=1}^N (y_{i,data} - \bar{y}_{data})^2 \quad (1)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (y_{i,data} - \bar{y}_{data})^2}{N}} \times 100\% \quad (2)$$

where $y_{i,data}$ and \bar{y}_{data} are the individual observed and the average observed values of the data and N is the number of measurements. $y_{i,model}$ is the prediction by machine learning models of the i th sample in the collected datasets.

Although these models can predict output variables with input parameters, these are black-box type models that cannot show which input features are more significant to the output parameters. In order to quantify the contribution of input feature information on the model prediction accuracy, two analyses were applied: 1) Pearson correlation and 2) feature importance. Pearson correlation analysis is one of the most common methods in statistics to measure the strength of the correlation between two variables [21,78,79]. Although the Pearson correlation shows a linear relationship between two variables, it can provide intuitive data visualization. On the other hand, the permutation feature importance can show the significance of each feature in the predictive performance of the models, regardless of linearity and the direction of the feature effectiveness [80]. Especially, the permutation feature importance is useful for non-linear or opaque estimators. Permutation feature importance is defined to be the decrease in a model score when a single feature value is randomly shuffled [70]. This procedure breaks down the relationship between features and the target, and thus the drop in the model score is indicative of the dependency of the model on this feature. This technique benefits from being model agnostic and can be calculated with different permutations of the feature.

Here, the MATLAB code '*predictorImportance*' is used to compute the importance of input features affecting each output feature, by summing changes in the node impurity (such as Gini index and deviance) due to splits on every predictor, and then dividing the sum by the total number of branch nodes. Here, the *Gini index* of a node was used for the importance computation:

$$Gini\ index = 1 - \sum_i p^2(i) \quad (3)$$

where the sum is over the class i at the node and $p(i)$ is the observed fraction of classes with class i that reach the node. The *Gini index* is 0 when all the elements belong to a specified class; otherwise, the *Gini index* increases with increasing impurities.

2.6. Monte Carlo filtering for finding key parameters predicting desired output parameters

These machine learning models predict output features as a function of biomass compositions and operating conditions. One of the most powerful aspects of the machine learning model is that one can explore the whole input features space (parameters set with arbitrary numbers) and predicts the outputs (performance). Even though such a dataset was not considered in the training and/or even unfeasible features set in reality due to safety, economics, or other potential problems, the models can predict the output, following the supervised trained set. By doing so, one can figure out which input features are the most effective to affect the output value, especially when many parameters are involved in. Note that the accuracy of the extrapolation estimated by the machine learning model is generally not reasonable, hence the exploration of the parameter should be in the range of the training set.

The permutation importance in each algorithm shows which parameters are more significant to predict the output features [81]. Similar to the permutation importance, Monte Carlo filtering (MCF) can also evaluate the operating conditions and forecast the primary operating conditions by filtering the product ranges (i.e., setting the output feature ranges) [82–84]. While machine learning is a supervised model, providing the known input and output features for the model development, the results can be interpreted with respect to the feature importance, explaining which features have higher priorities for the general biomass gasification process in fluidized beds. The MCF (or sequential Monte Carlo), however, filters the input features according to a set range of the desired outputs [85]. In other words, the MCF shows which features have higher priorities in achieving the desired ranges. By setting certain desired characteristics for the product, MCF can filter the inputs to satisfy the goal. Here, the goal could be a specific value referring to the literature or user-defined values which is generally decided as the value near the middle of the output range in the measured data. Also, the MCF can be used in non-linear and non-Gaussian data distribution [84]. Thus, integrating both machine learning and MCF results can show the primary operating conditions and input features leading to the desired syngas compositions and their LHV in fluidized beds biomass gasification.

The sensitivity analysis was performed with the two-sample Kolmogorov-Smirnov test applied to the training data. In detail, data from MCF were divided into two groups: set 1 for the data that passed the filter and set 2 for the data that failed the filter. Cumulative probability distribution function (CDF) graphs from each group were compared to quantify the distance, representing the sensitivity. The distance, called the Kolmogorov-Smirnov statistic (or statistic value) is [86]:

$$d_{n,m} = \sup |F_{1,n}(x) - F_{2,m}(x)| \quad (4)$$

where $F_{1,n}$ and $F_{2,m}$ are the empirical CDF of the set 1 with n observation and the set 2 with m observation, respectively, and sup is the supremum function. The empirical distribution function $F_n(x)$ for n independent and identically distributed ordered observations is defined as [86]:

$$F_n(x) = \frac{\text{number of elements in the sample} \leq x}{n} \quad (5)$$

with $d_{n,m}$, the null hypothesis is determined whether these two groups (set 1 and set 2) are drawn from the same distribution or not [82,86]. The null hypothesis is rejected at a certain significance level α , where:

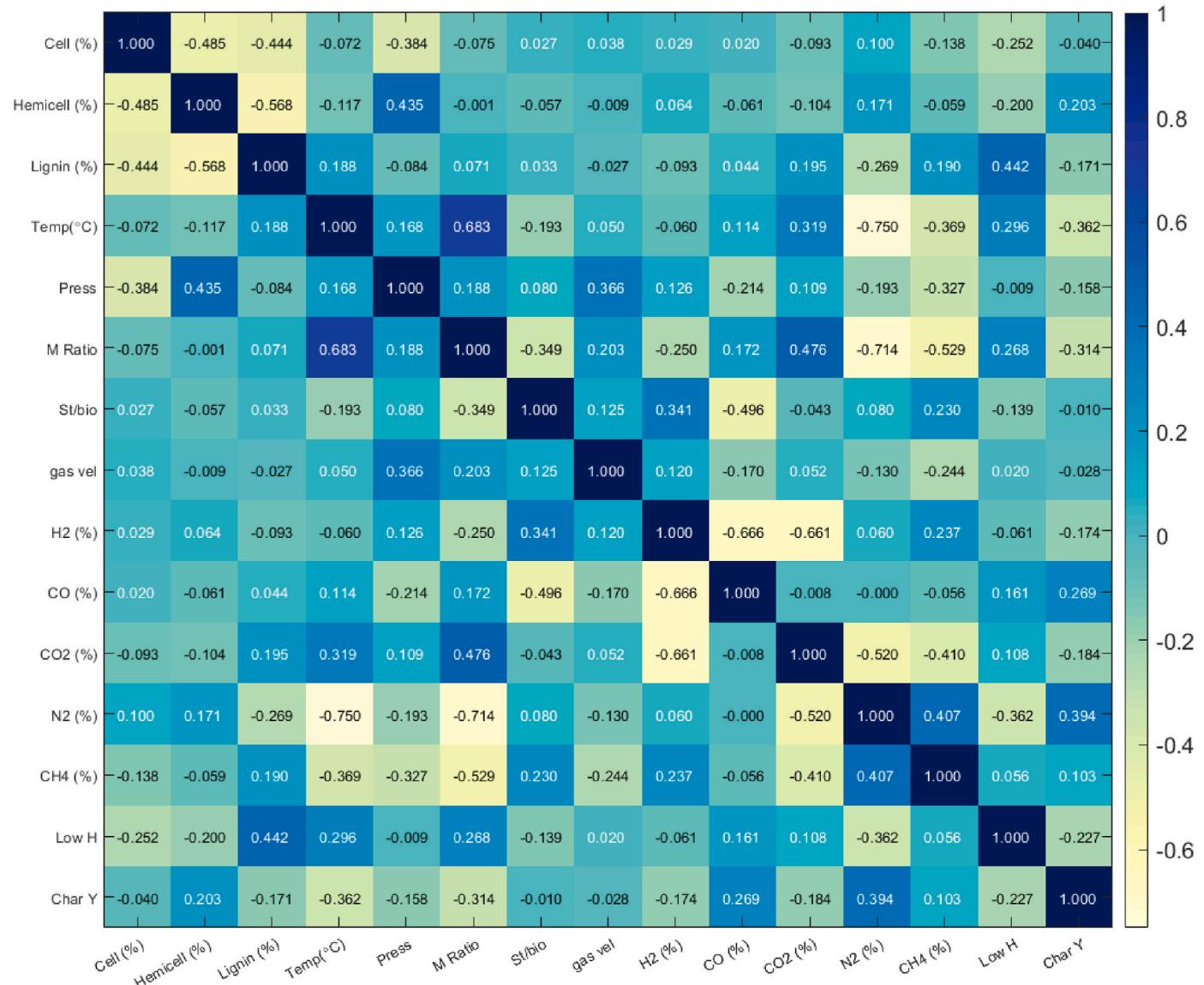


Fig. 4. Pearson correlation matrix between any two variables.

$$d_{n,m} > \sqrt{-\frac{n+m}{2nm} \ln\left(\frac{\alpha}{2}\right)} \quad (6)$$

Note that the significance level, α , is the probability of rejecting the null hypothesis when it is true [86]. Here, α was set to 0.05 for the statistical significance test.

3. Results and discussion

3.1. Pearson correlation and permutation importance analysis

Pearson correlation is a correlation coefficient, commonly used in linear regression. The range of the Pearson coefficient is between -1 and 1 , where 1 indicates a strong positive relationship; 0 indicates no relationship; and -1 indicates a strong negative relationship. Note that the results should be compared with the previous literature with numerical quantities because some data may have strong correlations in a certain range by coincidence, distorting the direct meaning of the matrix values. Fig. 4 shows the Pearson correlation results for the input and output variables. Regarding the relationship between the lignocellulosic composition and output variables, lignin has a strong positive linear relationship with char yield, showing 0.313 correlation coefficients, in

good agreement with previous work [42,45,87,88]. Temperature shows a negative relationship with CH_4 (-0.474) and tar yield (-0.353), indicating that the tar is more prone to decompose under high temperature, as well as methane combustion [36,45,47]. Equivalence ratio (ER) shows strong negative relationships with lower heating value (LHV) because higher ER produces more CO_2 which reduces the LHV [89,90]. Moreover, ER has a negative relationship with tar yield and CH_4 since increasing ER improves oxidation and eventually enhances CO_2 formation [91,77].

Although the Pearson correlation can simply understand the relationship between input and output variables, prediction accuracy improvement is necessary because 1) the Pearson correlation only represents the strength of the linear relationship between two features and 2) the correlation coefficient cannot quantify such correlations, as explained above.

Another useful method for finding the important features in black-box models is permutation importance. The permutation importance is a common and reliable technique that measures variable importance by observing the effect on model accuracy by arbitrarily shuffling each predictor variable [70]. Since the permutation importance is applicable to any black-box models as well as to non-linear regression, this method

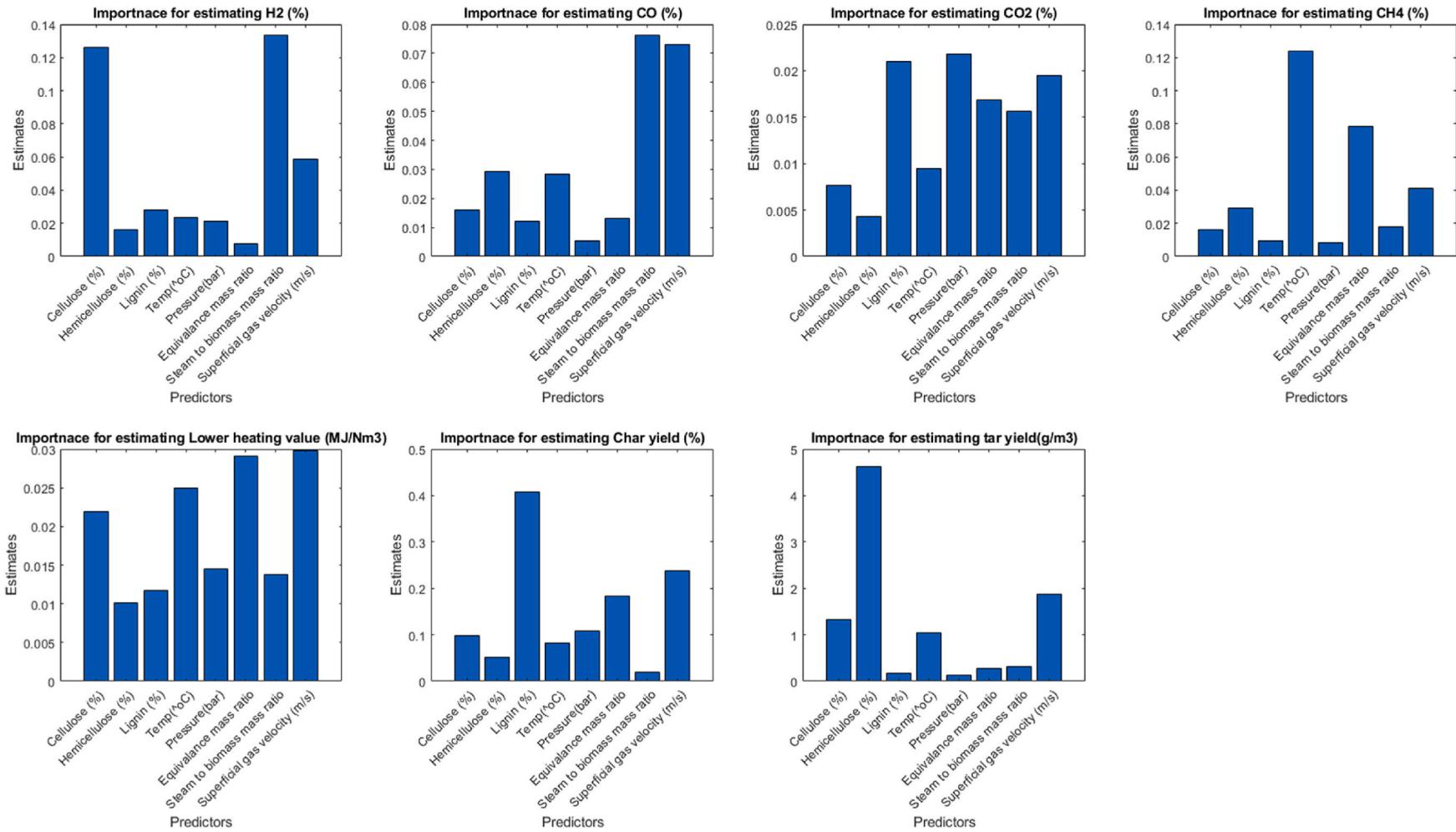


Fig. 5. Permutation importance of each input on each biomass gasification product.

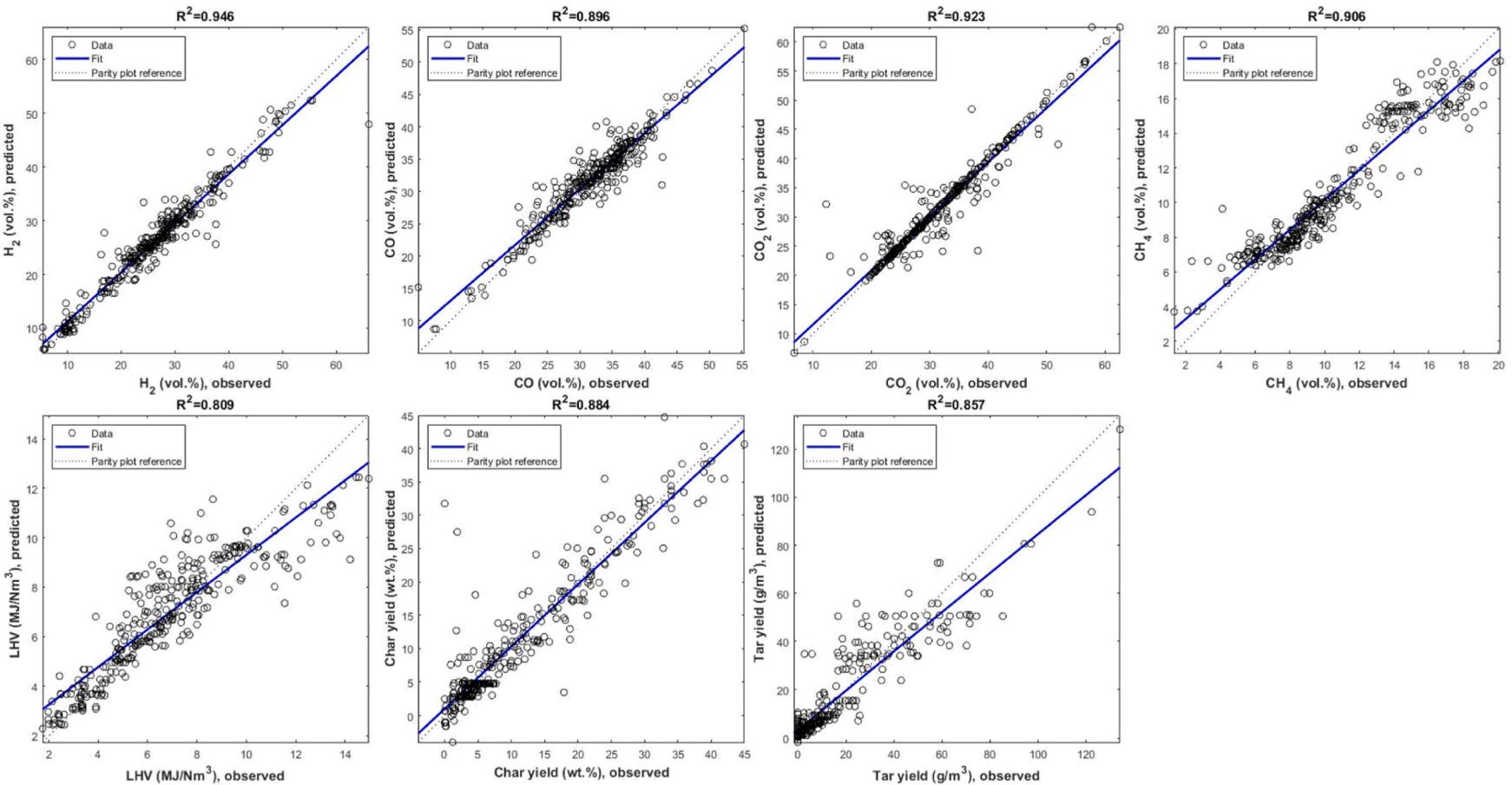


Fig. 6. Prediction results of random forest (RF) with each output parameter.

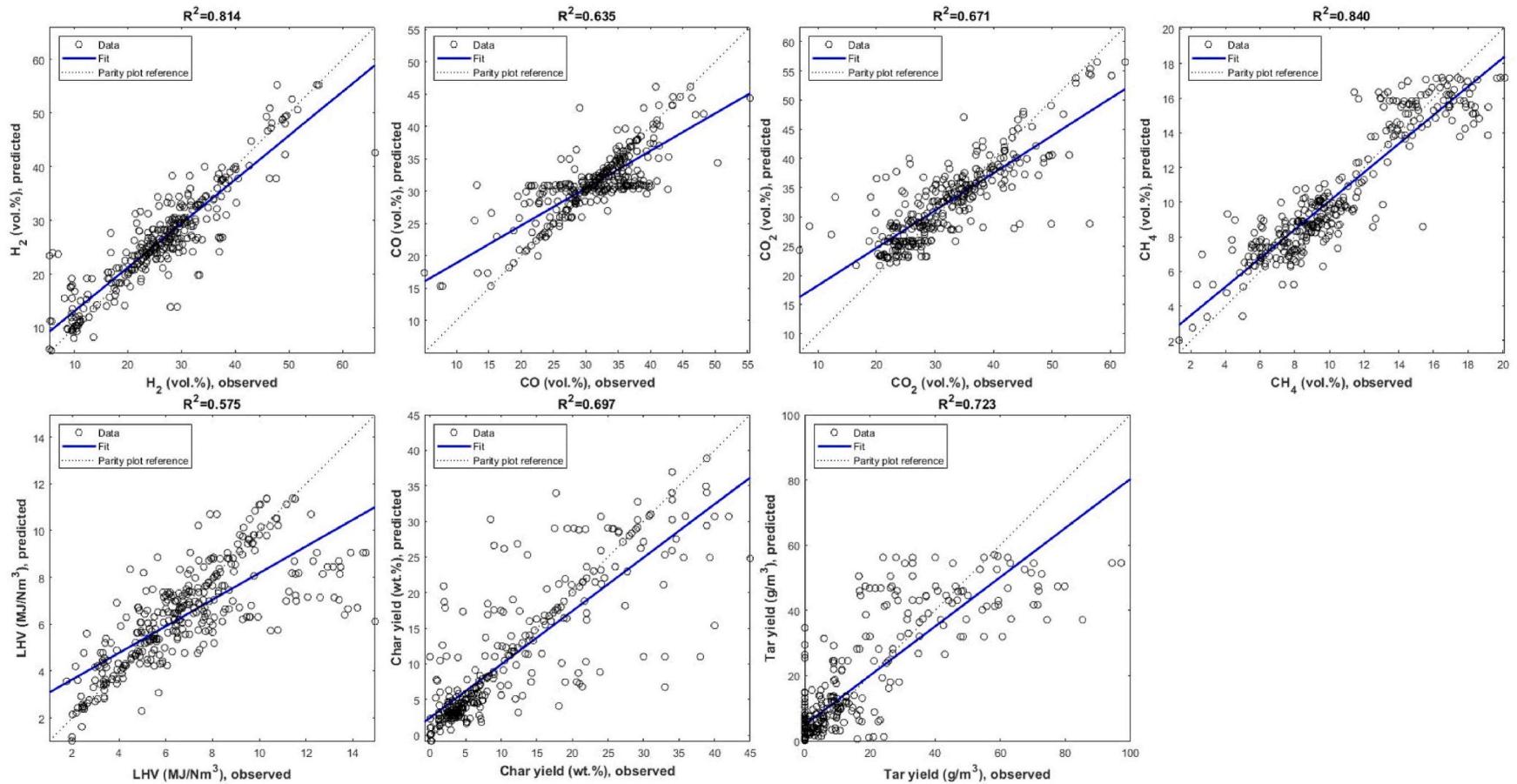


Fig. 7. Prediction results of support vector machine (SVM) with each output parameter.

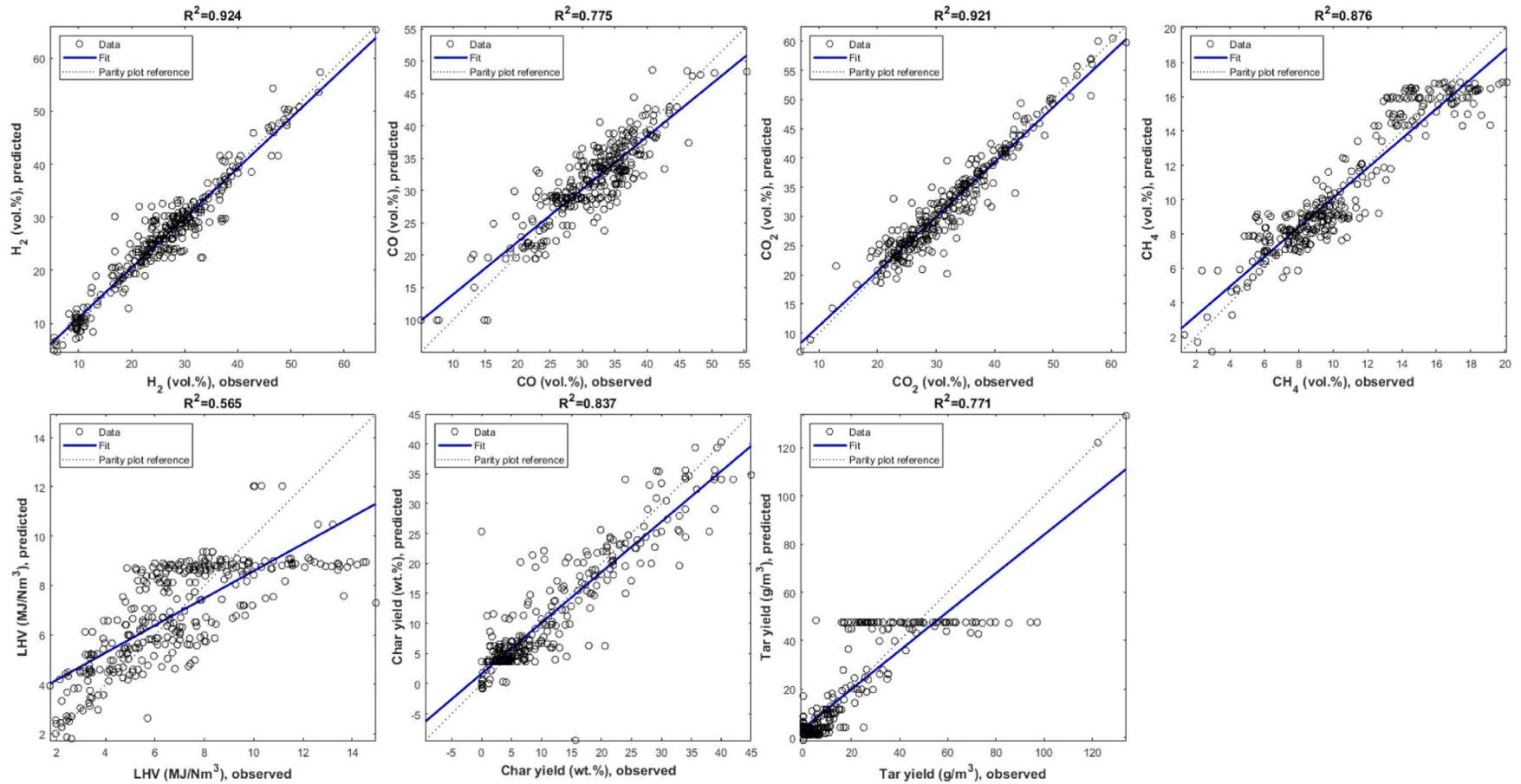


Fig. 8. Prediction results of artificial neural network (ANN) with each output parameter.

Table 3

Statistical indicators of prediction results with each machine learning algorithm.

Machine learning algorithm	Statistical indicator	H ₂	CO	CO ₂	CH ₄	LHV	Char	Tar
RF	R ²	0.946	0.896	0.923	0.906	0.809	0.884	0.857
	RMSE %	3.85	3.54	4.38	1.50	1.39	4.70	11.54
SVM	R ²	0.814	0.635	0.671	0.840	0.575	0.697	0.723
	RMSE %	4.94	4.77	5.65	1.74	1.97	6.25	13.07
ANN	R ²	0.924	0.775	0.921	0.876	0.565	0.837	0.771
	RMSE %	3.55	3.88	3.69	1.61	1.46	5.41	10.56

was also used to compensate for the feature importance. In detail, when a certain feature is not available, the permutation importance evaluates its importance by evaluating the increase in prediction error. As mentioned above, the built-in MATLAB code, ‘*predictorImportance*’, was used for the regression models. These sensitivity analysis results were plotted in Fig. 5 to find the detailed roles of the diverse input variables on each output feature. The results showed good accordance with the prior experimental work: steam to biomass mass ratio affects the H₂ composition in the syngas [3,91,78,79]. The equivalence mass ratio affects the CO₂ composition in syngas, as proven by experimental results from previous research [17,18,50,92,93]. For the CH₄ composition, since temperature affects the carbon conversion efficiency, the CH₄ composition is also influenced by temperature [87]. Cellulose plays an important role affecting H₂ composition in machine learning (see Fig. 5), while no significant relationship is not observed in the Pearson correlation. Since H₂ as an output feature shows a multivariable and non-linear relationship, the permutation importance results reveal that cellulose is an essential feature for the H₂ composition, and the other features (i.e., SBR and U_g) may offset its significance. The positive effect of the cellulose composition on H₂ production, while fixing the SBR and U_g, is experimentally proved by Tian et al. [88].

3.2. Training performance with different machine learning approaches

After the hyperparameter optimization for each machine learning model, as explained in Sections 2.2–2.5, the comparison between the observed (trained) and the target (predicted) data with each machine learning technique is plotted in Figs. 6–8. The R² and RMSE results for the target variables are given in Table 3. Note that the R² and RMSE of training performances of each machine learning model are also given in Table S1, SI 2. Although the SVM showed a good training performance, comparable to the other two machine learning algorithms, the target (prediction) performance of the SVM was poor for all output R² parameters. The RF model predictions generally show a great agreement with the observed data with R² higher than that of SVM and ANN. The observed versus predicted value plots closely follow the diagonal line for the RF algorithm. Some of the predicted data were deviated from the one observed data point. This may result from the lack of orthogonality in the selected datapoint for machine learning modeling as they were mostly collected from varying “one factor at a time” experiments. Note that each tree in the RF model represents a learning process, i.e., an ensemble learning method, while the ANN model features combined two single correlations or learning processes as the ANN has two hidden layers (see Table 2). The above discussion indicates that both ANN and RF models can well represent the complex nonlinear correlations between raw material biomass, operating conditions and biomass gasification products.

3.3. Machine learning adopted Monte Carlo filtering

Monte Carlo filtering (MCF) integrated with machine learning algorithms was tested to describe how well the predictions on the highest priority amongst input features driven by the machine learning model fit the experimental results. The input features were arbitrarily selected in specific ranges as follows. Since the summation of the lignocellulosic

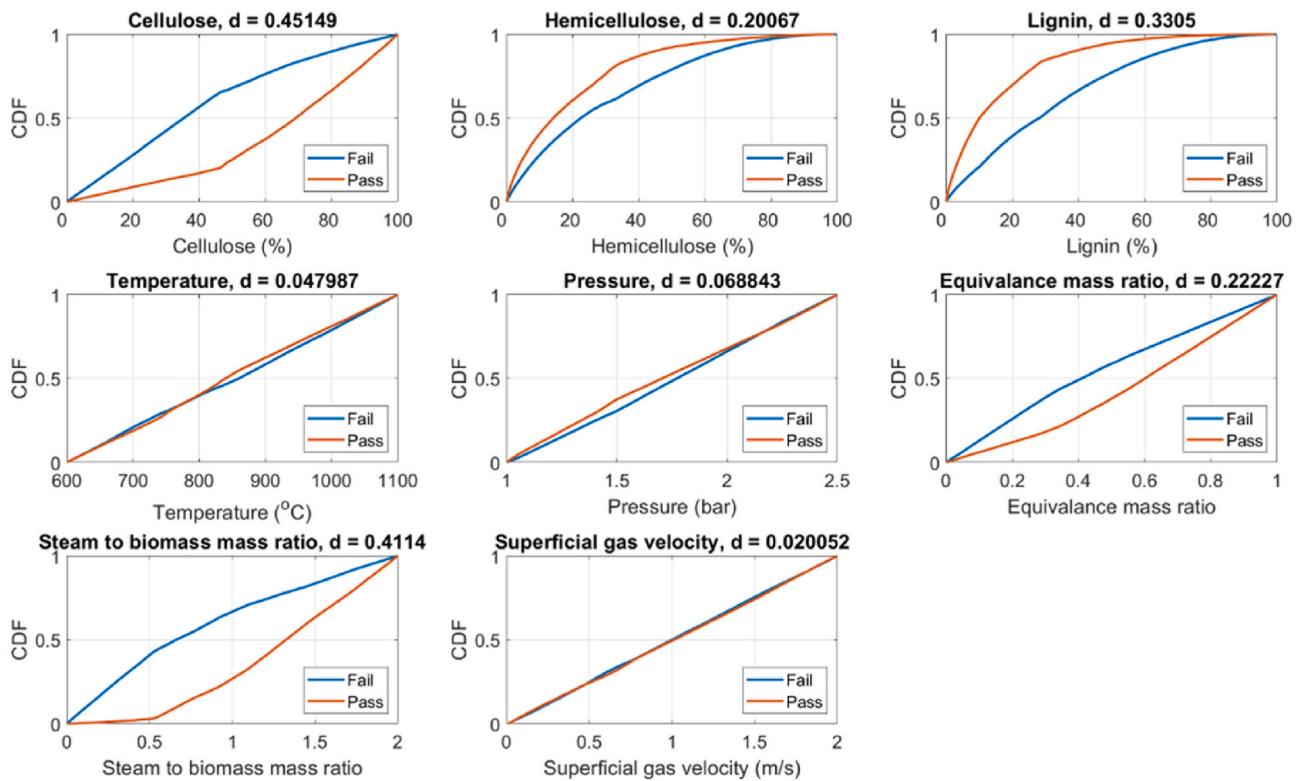
composition must be equal to 1 as unity, the cellulose was randomly selected in [0, 1]. Then, the hemicellulose was randomly selected in the ranges of [0, 1 – Cell.]. Lastly, the composition of lignin was determined as (1 – Cell. – Hem.). The ranges of the other input parameters were based on Table 1. After randomly selecting the random numbers from each range, the output value was estimated by each machine learning model. Here, the pass and fail curves of the MCF were decided with the specific value as defined: 25.5, 32.08, 32.49, 8.812, 5.86, 8.5, 5.57, 1.1 for the model of H₂ (vol%), CO (vol%), CO₂ (vol%), CH₄ (vol%), LHV (MJ/Nm³), char yield (wt%), tar yield (g/Nm³), and H₂/CO ratio, respectively. Note that the LHV and H₂/CO were based on the values suggested by AlNouss et al. [94], the optimum H₂/CO ratio of the syngas produced from the fluidized bed is higher than 1.1, as well as LHV > 5.86 MJ/m³ such that the syngas readily acts as an intermediate to produce chemicals, including aldehydes. The other values were obtained from the median value of the whole data set. The MCF was performed 100,000 times for each MCF integrated machine learning model to find which features driven by the machine learning model are more significant to meet these optimization conditions.

The RF and ANN integrated MCF sensitivity analysis of (a) H₂/CO ratio and (b) LHV for finding significant input features are plotted in Figs. 9 and 10, respectively. Similar to the results above, SVM showed the lowest prediction accuracy. Thus, only RF- and ANN-MCF results were considered here for finding a suitable model prediction. Pass/failure of the Kolmogorov-Smirnov test for all input features are given in SI 2. The Kolmogorov-Smirnov statistic between RF and ANN algorithms integrated with Monte Carlo filtering for predicting optimum syngas ranges (H₂/CO > 1.1 and LHV > 5.86 MJ/m³) is addressed in Table 4. To meet the H₂/CO ratio higher than 1.1, the three input parameters of the highest priority predicted from the RF are steam to biomass ratio, cellulose and lignin. Compared to the results from the RF-MCF, the two highest priorities predicted from the ANN-MCF are cellulose and hemicellulose. The statistical values for cellulose and hemicellulose are 0.396 and 0.438, respectively. Since the lignocellulosic compositions are interdependent (i.e., the sum of these three components is added up to 1 for all biomass), the biomass composition itself may not fully represent the importance of the required H₂/CO ratio for the syngas. For the LHV > 5.86 MJ/m³, biomass compositions, ER, lignin and U_g showed high statistic values tested with RF integrated MCF. Similar to the results obtained from the H₂/CO ratio, lignocellulosic compositions showed the highest importance with the ANN integrated MCF. However, only ER showed distinctive significance, whereas the other showed very little importance with the ANN-MCF analysis. As the ER significantly affects the H₂ composition in the syngas both shown in the permutation importance (see Fig. 5) and experimentally [69], RF-MCF is determined to be more suitable rather than the ANN-MCF.

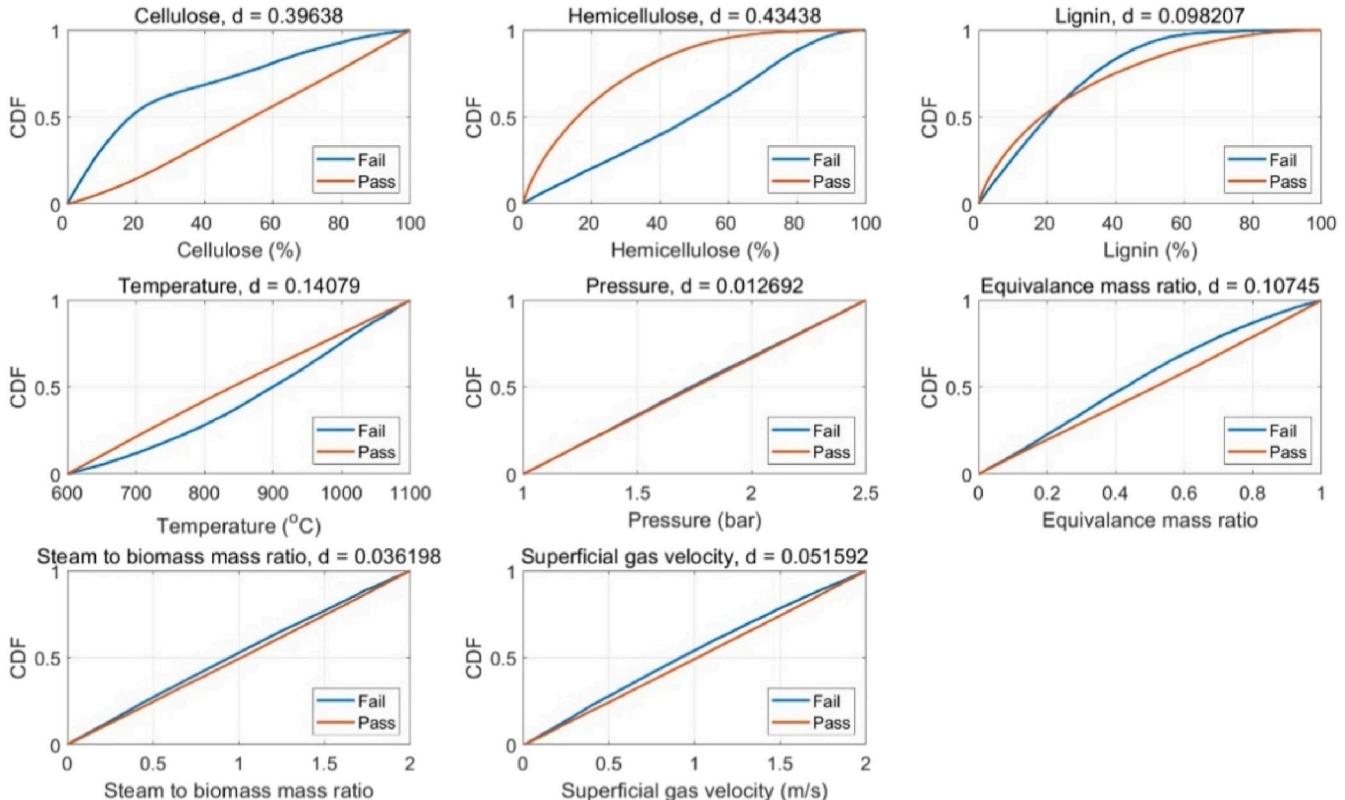
4. Conclusions

Three machine learning algorithms, Random Forest, Support Vector Machine and Artificial Neural Network, were applied to predict the syngas products, LHV and tar/char yield with lignocellulosic information and operating conditions from 336 literature data points. Their performances were compared to highlight the optimal model. For the sensitivity analysis, Pearson correlation and permutation importance

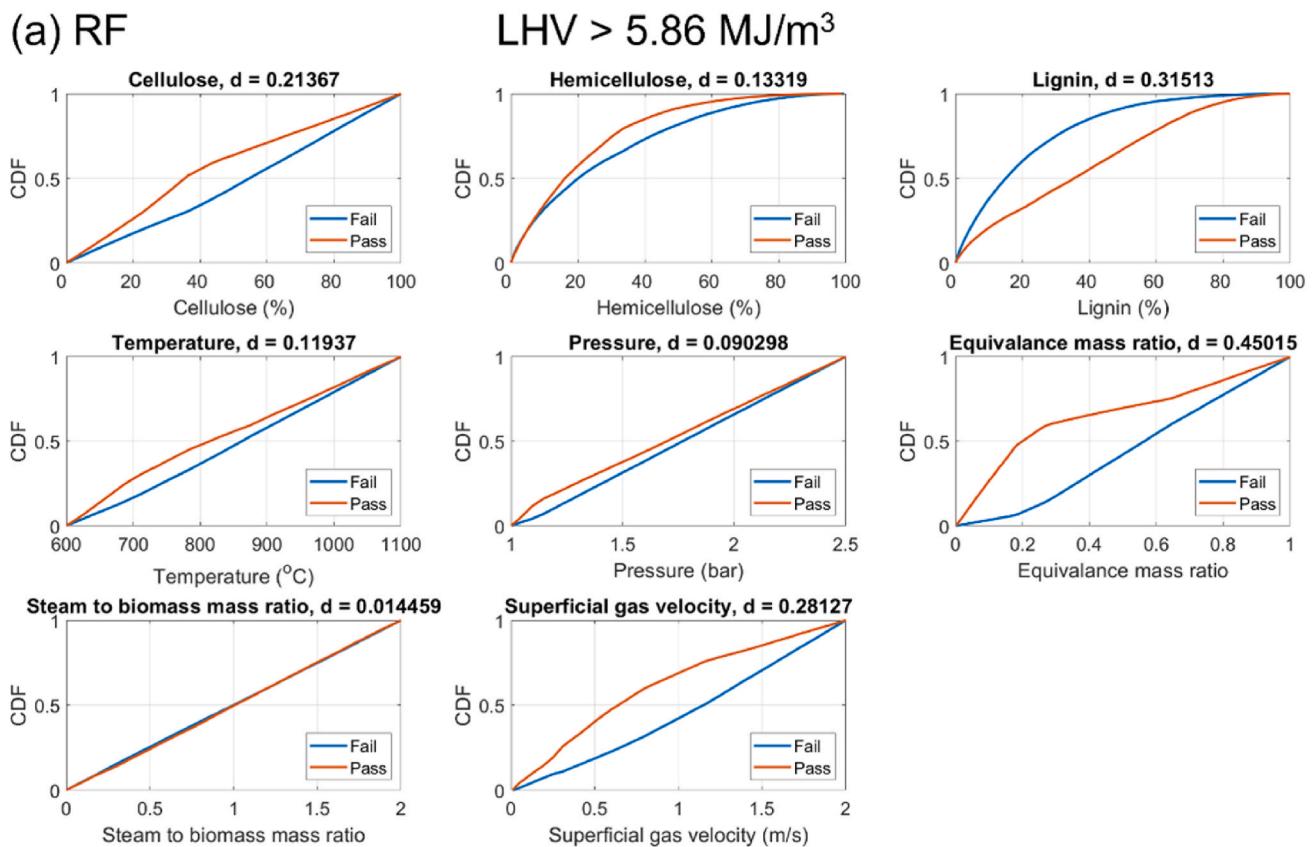
(a) RF

 $H_2/CO > 1.1$ 

(b) ANN

 $H_2/CO > 1.1$ Fig. 9. Cumulative probability function of each input feature on H_2/CO ratio predicted with (a) RF and (b) ANN integrated MCF.

(a) RF



(b) ANN

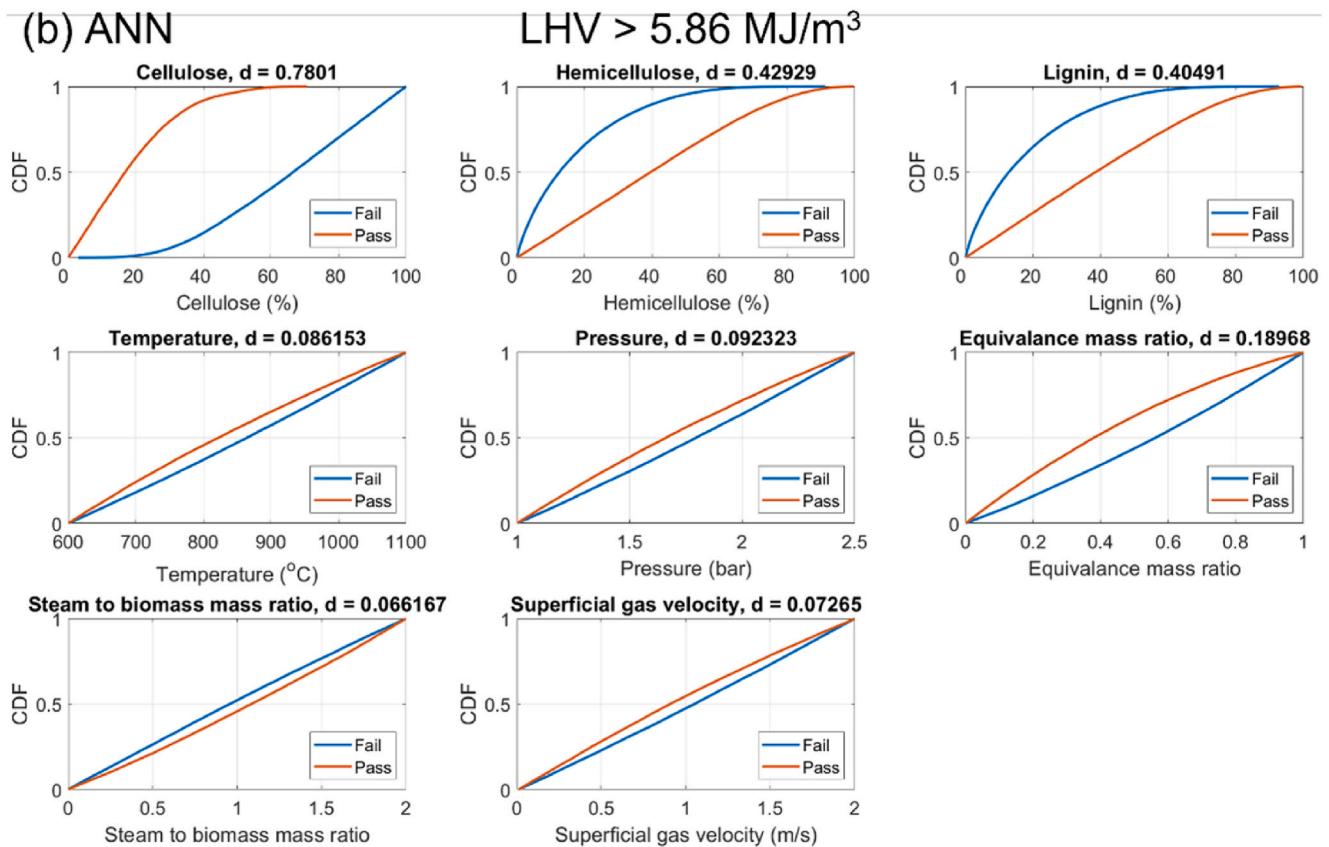


Fig. 10. Cumulative probability function of each input feature on LHV predicted with (a) RF and (b) ANN integrated with MCF.

Table 4

Kolmogorov-Smirnov statistic (d) comparison between two machine learning algorithms integrated with Monte Carlo filtering with each output features.

		Cell.	Hem.	Lignin	Temp.	Pressure	ER	SBR	U_g
$H_2/CO > 1.1$	RF	0.451	0.201	0.331	0.048	0.069	0.222	0.411	0.020
	ANN	0.396	0.434	0.098	0.141	0.013	0.107	0.036	0.052
$LHV > 5.84 \text{ MJ/m}^3$	RF	0.214	0.133	0.315	0.119	0.090	0.450	0.014	0.281
	ANN	0.780	0.423	0.405	0.086	0.092	0.190	0.066	0.073

were used to evaluate the algorithms considering the physical meaning. The RF and ANN showed good prediction with high R^2 and lower $RSME$, whereas the SVM failed to reach satisfactory prediction accuracy of R^2 . Monte Carlo Filtering integrated with machine learning algorithms was adopted to find the feature importance of the desired syngas composition and LHV . To meet the H_2/CO ratio higher than 1.1, steam to biomass ratio, cellulose and lignin are the three most important input features forecasted by RF integrated with MCF, whereas only cellulose and hemicellulose are the most important factors identified by ANN integrated with MCF. Regarding the $LHV > 5.86 \text{ MJ/m}^3$, equivalence ratio, lignin and superficial gas velocity high statistic values tested with RF-MCF. Similar to the results obtained from the H_2/CO ratio, lignocellulosic compositions showed the highest importance with the ANN-MCF. The knowledge of this data is decisive in setting the operating conditions of the fluidized bed gasifier with biomass as an energy source, as well as finding key operating parameters affecting the optimum syngas composition.

Credit author statement

Jun Young Kim: conceptualization, methodology, formal analysis, writing-original draft, writing-review & editing, project administration. **Dongjae Kim:** software, validation, visualization, writing-original draft, writing-review & editing. **Zezhong John Li:** writing-review & editing. **Claudio Dariva:** supervision. **Yankai Cao:** supervision, writing-review & editing. **Naoko Ellis:** supervision, funding acquisition.

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

I have shared my datafile in the attached file

Acknowledgement

The authors would like to acknowledge the financial support from the National Research Foundation of Korea (NRF) grant funded by Korea (Project#: NRF-2018M3D3A1A01018009).

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.energy.2022.125900>.

References

- [1] Abdelouahed L, Authier O, Mauviel G, Corriou JP, Verdier G, Dufour A. Detailed modeling of biomass gasification in dual fluidized bed reactors under Aspen Plus. *Fuels & Fuels*; 2012. p. 3840–55.
- [2] Xing J, Luo K, Wang H, Jin T, Fan J. Novel sensitivity study for biomass directional devolatilization by random forest models. *Energy Fuel* 2020;34:8414–23. <https://doi.org/10.1021/acs.energyfuels.0c00822>.
- [3] Deraman MR, Abdul Rasid R, Othman MR, Suli LNM. Co-gasification of coal and empty fruit bunch in an entrained flow gasifier: a process simulation study. In: IOP conf ser mater sci eng. IOP Publishing Ltd; 2019, 012005. <https://doi.org/10.1088/1757-899X/702/1/012005>.
- [4] Kim JY, Li ZJ, Ellis N, Lim CJ, Grace JR. Dynamic Monte Carlo reactor modeling of calcium looping with sorbent purge and utilization decay. *Chem Eng J* 2022;435: 134954. <https://doi.org/10.1016/J.CEJ.2022.134954>.
- [5] Ebneyamini A, Li ZJ, Kim JY, Grace JR, Lim CJ, Ellis N. Effect of calcination temperature and extent on the multi-cycle CO₂ carrying capacity of lime-based sorbents. *J CO₂ Util* 2021;49:101546. <https://doi.org/10.1016/j.jcou.2021.101546>.
- [6] Siedlecki M, de Jong W, Verkooijen AHM. Fluidized bed gasification as a mature and reliable technology for the production of bio-syngas and applied in the production of liquid transportation fuels-a review. *Energies* 2011;4:389–434. <https://doi.org/10.3390/en4030389>.
- [7] Saayman J, Xu M, Lim CJ, Ellis N. Gas leakage between reactors in a dual fluidized bed system. *Powder Technol* 2014;266:196–202. <https://doi.org/10.1016/j.powtec.2014.06.012>.
- [8] Kim JY, Bae JW, Grace JR, Epstein N, Lee DH. Hydrodynamic characteristics at the layer inversion point in three-phase fluidized beds with binary solids. *Chem Eng Sci* 2017;157:99–106. <https://doi.org/10.1016/j.ces.2015.11.021>.
- [9] Chen Z, Grace JR, Lim CJ. Development of particle size distribution during limestone impact attrition. *Powder Technol* 2011;207:55–64. <https://doi.org/10.1016/J.POWTEC.2010.10.010>.
- [10] Kim JY, Bae JW, Grace JR, Epstein N, Lee DH. Horizontal immersed heater-to-bed heat transfer with layer inversion in gas-liquid-solid fluidized beds of binary solids. *Chem Eng Sci* 2017;170:501–7. <https://doi.org/10.1016/j.ces.2017.01.007>.
- [11] Couto N, Rouboa A, Silva V, Monteiro E, Bouziâne K. Influence of the biomass gasification processes on the final composition of syngas. In: Energy procedia. Elsevier Ltd; 2013. p. 596–606. <https://doi.org/10.1016/j.egypro.2013.07.068>.
- [12] McKendry P. Energy production from biomass (part 1): overview of biomass. *Bioresour Technol* 2002;83:37–46. [https://doi.org/10.1016/S0960-8524\(01\)00118-3](https://doi.org/10.1016/S0960-8524(01)00118-3).
- [13] Xing J, Wang H, Luo K, Wang S, Bai Y, Fan J. Predictive single-step kinetic model of biomass devolatilization for CFD applications: a comparison study of empirical correlations (EC), artificial neural networks (ANN) and random forest (RF). *Renew Energy* 2019;136:104–14. <https://doi.org/10.1016/j.renene.2018.12.088>.
- [14] Sunphorkot S, Chalermsinsuwan B, Plumsomboon P. Artificial neural network model for the prediction of kinetic parameters of biomass pyrolysis from its constituents. *Fuel* 2017;193:142–58. <https://doi.org/10.1016/j.fuel.2016.12.046>.
- [15] Molino A, Larocca V, Chianese S, Musmarra D. Biofuels production by biomass gasification: a review. *Energy* 2018;11. <https://doi.org/10.3390/en11040811>.
- [16] Sikarwar VS, Zhao M, Clough P, Yao J, Zhong X, Memon MZ, Shah N, Anthony EJ, Fennell PS. An overview of advances in biomass gasification. *Energy Environ Sci* 2016;9:2939–77. <https://doi.org/10.1039/c6ee00935b>.
- [17] Sansanwilai SK, Pal K, Rosen MA, Tyagi SK. Recent advances in the development of biomass gasification technology: a comprehensive review. *Renew Sustain Energy Rev* 2017;72:363–84. <https://doi.org/10.1016/j.rser.2017.01.038>.
- [18] Ahmad AA, Zawawi NA, Kasim FH, Inayat A, Khasri A. Assessing the gasification performance of biomass: a review on biomass gasification process conditions, optimization and economic evaluation. *Renew Sustain Energy Rev* 2016;53: 1333–47. <https://doi.org/10.1016/j.rser.2015.09.030>.
- [19] Ramos A, Monteiro E, Silva V, Rouboa A. Co-gasification and recent developments on waste-to-energy conversion: a review. *Renew Sustain Energy Rev* 2018;81: 380–98. <https://doi.org/10.1016/j.rser.2017.07.025>.
- [20] Parthasarathy P, Narayanan KS. Hydrogen production from steam gasification of biomass: influence of process parameters on hydrogen yield - a review. *Renew Energy* 2014;66:570–9. <https://doi.org/10.1016/j.renene.2013.12.025>.
- [21] Elmaz F, Büyükkırkör B, Yıldız Ö, Mutlu AY. Classification of solid fuels with machine learning. *Fuel* 2020;266:117066. <https://doi.org/10.1016/j.fuel.2020.117066>.
- [22] Mutlu AY, Yıldız Ö. An artificial intelligence based approach to predicting syngas composition for downdraft biomass gasification. *Energy* 2018;165:895–901. <https://doi.org/10.1016/j.energy.2018.09.131>.
- [23] Chew JW, Cocco RA. Application of machine learning methods to understand and predict circulating fluidized bed riser flow characteristics. *Chem Eng Sci* 2020;217: 115503. <https://doi.org/10.1016/j.ces.2020.115503>.
- [24] Zhong H, Sun Z, Zhu J, Zhang C. Prediction of solid holdup in a gas-solid circulating fluidized bed riser by artificial neural networks. *Ind Eng Chem Res* 2021;60:3452–62. <https://doi.org/10.1021/ACS.IECR.0C05474>.
- [25] Zhu L-T, Tang J-X, Luo Z-H. Machine learning to assist filtered two-fluid model development for dense gas-particle flows. *AIChE J* 2020;66:e16973. <https://doi.org/10.1002/AIC.16973>.
- [26] de Souza MB, Couceiro L, Barreto AG, Quiteste CPB. Neural network based modeling and operational optimization of biomass gasification processes. In: Yun Y, editor. Gasification for practical applications. Rijeka: InTech; 2012. p. 297–312. <https://doi.org/10.5772/48516>.

- [27] Serrano D, Golpour I, Sánchez-Delgado S. Predicting the effect of bed materials in bubbling fluidized bed gasification using artificial neural networks (ANNs) modeling approach. *Fuel* 2020;266:117021. <https://doi.org/10.1016/j.fuel.2020.117021>.
- [28] Warnecke R. Gasification of biomass: comparison of fixed bed and fluidized bed gasifier. *Biomass Bioenergy* 2000;18:489–97. [https://doi.org/10.1016/S0961-9534\(00\)00009-X](https://doi.org/10.1016/S0961-9534(00)00009-X).
- [29] Czernik S, Koeberle PG, Jollez P, Bilodeau JF, Chornet E. Gasification of residual biomass via the biosyn fluidized bed technology. In: Advances in thermochemical biomass conversion. Springer Netherlands; 1993. p. 423–37. https://doi.org/10.1007/978-94-011-1336-6_33.
- [30] Corella J, Herguido J, Gonzalez-Saiz J, Alday FJ, Rodriguez-Trujillo JL. Fluidized bed steam gasification of biomass with dolomite and with a commercial FCC catalyst. In: Research in thermochemical biomass conversion. Springer Netherlands; 1988. p. 754–65. https://doi.org/10.1007/978-94-009-2737-7_57.
- [31] Siedlecki M, de Jong W. Biomass gasification as the first hot step in clean syngas production process - gas quality optimization and primary tar reduction measures in a 100 kW thermal input steam-oxygen blown CFB gasifier. S40–S62 *Biomass Bioenergy* 2011;35. <https://doi.org/10.1016/j.biombioe.2011.05.033>.
- [32] Mayerhofer M, Mitsakis P, Meng X, de Jong W, Spliethoff H, Gaderer M. Influence of pressure, temperature and steam on tar and gas in all thermal fluidized bed gasification. *Fuel* 2012;99:204–9. <https://doi.org/10.1016/j.fuel.2012.04.022>.
- [33] Caballero MA, Corella J, Aznar MP, Gil J. Biomass gasification with air in fluidized bed. Hot gas cleanup with selected commercial and full-size nickel-based catalysts. *Ind Eng Chem Res* 2000;39:1143–54. <https://doi.org/10.1021/ie990738t>.
- [34] Cherney JH, Johnson KD, Volenec JJ, Anliker KS. Chemical composition of herbaceous grass and legume species grown for maximum biomass production. *Biomass* 1988;17:215–38. [https://doi.org/10.1016/0144-4565\(88\)90105-9](https://doi.org/10.1016/0144-4565(88)90105-9).
- [35] Christodoulou C, Grimekis D, Panopoulos KD, Pachatouridou EP, Iliopoulos EF, Kakaras E. Comparing calcined and un-treated olivine as bed materials for tar reduction in fluidized bed gasification. *Fuel Process Technol* 2014;124:275–85. <https://doi.org/10.1016/j.fuproc.2014.03.012>.
- [36] Horvat A, Pandey DS, Kwapinska M, Mello BB, Gómez-Barea A, Fryda LE, Rabou LPLM, Kwapinski W, Leahy JJ. Tar yield and composition from poultry litter gasification in a fluidised bed reactor: effects of equivalence ratio, temperature and limestone addition. *RSC Adv* 2019;9:13283–96. <https://doi.org/10.1039/c9ra02548k>.
- [37] Carpenter DL, Bain RL, Davis RE, Dutta A, Feik CJ, Gaston KR, Jablonski W, Phillips SD, Nimlos MR. Pilot-scale gasification of corn stover, switchgrass, wheat straw, and wood: 1. Parametric study and comparison with literature. *Ind Eng Chem Res* 2010;49:1859–71. <https://doi.org/10.1021/ie900595m>.
- [38] e Silva CFL, Schirmer MA, Maeda RN, Barcelos CA, Pereira N. Potential of giant reed (*Arundo donax* L.) for second generation ethanol production. *Electron J Biotechnol* 2015;18:10–5. <https://doi.org/10.1016/j.ejbt.2014.11.002>.
- [39] Christodoulou C, Tsekos C, Tsalidis G, Fantini M, Panopoulos KD, de Jong W, Kakaras E. Attempts on cardoon gasification in two different circulating fluidized beds. *Case Stud Therm Eng* 2014;4:42–52. <https://doi.org/10.1016/j.csite.2014.06.004>.
- [40] Serrano D, Sánchez-Delgado S, Horvat A. Effect of sepiolite bed material on gas composition and tar mitigation during *C. cardunculus* L. gasification. *Chem Eng J* 2017;317:1037–46. <https://doi.org/10.1016/j.cej.2017.02.106>.
- [41] Dwivedi G, Sharma MP. Impact of cold flow properties of biodiesel on engine performance. *Renew Sustain Energy Rev* 2014;31:650–6. <https://doi.org/10.1016/j.rser.2013.12.035>.
- [42] Nilsson S, Gómez-Barea A, Fuentes-Cano D, Haro P, Pinna-Hernández G. Gasification of olive tree pruning in fluidized bed: experiments in a laboratory-scale plant and scale-up to industrial operation. *Energy Fuel* 2017;31:542–54. <https://doi.org/10.1021/acs.energyfuels.6b02039>.
- [43] Fernández JL, Sáez F, Castro E, Manzanares P, Ballesteros M, Negro MJ. Determination of the lignocellulosic components of olive tree pruning biomass by near infrared spectroscopy. *Energies* 2019;12:2497. <https://doi.org/10.3390/en12132497>.
- [44] Kwapinska M, Xue G, Horvat A, Rabou LPLM, Dooley S, Kwapinski W, Leahy JJ. Fluidized bed gasification of torrefied and raw grassy biomass (*miscanthus* × *giganteus*). The effect of operating conditions on process performance. *Energy Fuel* 2015;29:7290–300. <https://doi.org/10.1021/acs.energyfuels.5b01144>.
- [45] Campoy M, Gómez-Barea A, Fuentes-Cano D, Ollero P. Tar reduction by primary measures in an autothermal air-blown fluidized bed biomass gasifier. *Ind Eng Chem Res* 2010;49:11294–301. <https://doi.org/10.1021/ie101267c>.
- [46] Khoja AH, Tahir M, Saidina Amin NA. Evaluating the performance of a Ni catalyst supported on La₂O₃-MgAl₂O₄ for dry reforming of methane in a packed bed dielectric barrier discharge plasma reactor. *Energy Fuel* 2019;33:11630–47. <https://doi.org/10.1021/acs.energyfuels.9b02236>.
- [47] Virginie M, Adánez J, Courson C, de Diego LF, García-Labiano F, Niznansky D, Kienemann A, Gayán P, Abad A. Effect of Fe-olivine on the tar content during biomass gasification in a dual fluidized bed. 121–Appl Catal, B 2012;122:214–22. <https://doi.org/10.1016/j.apcatb.2012.04.005>.
- [48] Siedlecki M, Nieuwstraten R, Simeone E, de Jong W, Verkooijen AHM. Effect of magnesite as bed material in a 100 kWth steam-oxygen blown circulating fluidized-bed biomass gasifier on gas composition and tar formation. *Energy Fuel* 2009;23: 5643–54. <https://doi.org/10.1021/ef900420c>.
- [49] Serrano D, Kwapinska M, Horvat A, Sánchez-Delgado S, Leahy JJ. Cynara cardunculus L. gasification in a bubbling fluidized bed: the effect of magnesite and olivine on product gas, tar and gasification performance. *Fuel* 2016;173:247–59. <https://doi.org/10.1016/j.fuel.2016.01.051>.
- [50] Kurkela E, Ståhlberg P. Air gasification of peat, wood and brown coal in a pressurized fluidized-bed reactor. I. Carbon conversion, gas yields and tar formation. *Fuel Process Technol* 1992;31:1–21. [https://doi.org/10.1016/0378-3820\(92\)90038-R](https://doi.org/10.1016/0378-3820(92)90038-R).
- [51] Arpiainen V, Lappi M. Products from the flash pyrolysis of peat and pine bark. *J Anal Appl Pyrolysis* 1989;16:355–76. [https://doi.org/10.1016/0165-2370\(89\)80018-X](https://doi.org/10.1016/0165-2370(89)80018-X).
- [52] Soria-Verdugo A, von Berg L, Serrano D, Hochenauer C, Scharler R, Anca-Couce A. Effect of bed material density on the performance of steam gasification of biomass in bubbling fluidized beds. *Fuel* 2019;257:116118. <https://doi.org/10.1016/j.fuel.2019.116118>.
- [53] Pasangulapati V, Ramachandriya KD, Kumar A, Wilkins MR, Jones CL, Huhnke RL. Effects of cellulose, hemicellulose and lignin on thermochemical conversion characteristics of the selected biomass. *Bioresour Technol* 2012;114:663–9. <https://doi.org/10.1016/j.biotech.2012.03.036>.
- [54] Li XT, Grace JR, Lim CJ, Watkinson AP, Chen HP, Kim JR. Biomass gasification in a circulating fluidized bed. *Biomass Bioenergy* 2004;26:171–93. [https://doi.org/10.1016/S0961-9534\(03\)00084-9](https://doi.org/10.1016/S0961-9534(03)00084-9).
- [55] Mimunin J, Limpitipanich P, Promwungkwa A. Delignification of bana grass using sodium hydroxide and ozone. *Waste Biomass Valorization* 2018;9:2099. <https://doi.org/10.1007/s12649-017-0002-2>. 2105.
- [56] Skoulou V, Koufodimos G, Samaras Z, Zabaniotou A. Low temperature gasification of olive kernels in a 5-kW fluidized bed reactor for H₂-rich producer gas. *Int J Hydrogen Energy* 2008;33:6515–24. <https://doi.org/10.1016/j.ijhydene.2008.07.074>.
- [57] Motauang TE, Anandjiwala RD. Effect of alkali and acid treatment on thermal degradation kinetics of sugar cane bagasse. *Ind Crop Prod* 2015;74:472–7. <https://doi.org/10.1016/j.indcrop.2015.05.062>.
- [58] Weerachanchai P, Horio M, Tangsathithkulchai C. Effects of gasifying conditions and bed materials on fluidized bed steam gasification of wood biomass. *Bioresour Technol* 2009;100:1419–27. <https://doi.org/10.1016/j.biotech.2008.08.002>.
- [59] Wang Z, Yang X, Sun B, Chai Y, Liu J, Cao J. Effect of vacuum heat treatment on the chemical composition of larch wood. *Bioresources* 2016;11. <https://doi.org/10.1537/biores.11.3.5743-5750>.
- [60] Miccio F, Piriou B, Ruoppolo G, Chirone R. Biomass gasification in a catalytic fluidized reactor with beds of different materials. *Chem Eng J* 2009;154:369–74. <https://doi.org/10.1016/j.cej.2009.04.002>.
- [61] Campoy M, Gómez-Barea A, Villanueva AL, Ollero P. Air-steam gasification of biomass in a fluidized bed under simulated autothermal and adiabatic conditions. *Ind Eng Chem Res* 2008;47:5957–65. <https://doi.org/10.1021/ie800220t>.
- [62] Campoy M, Gómez-Barea A, Vidal FB, Ollero P. Air-steam gasification of biomass in a fluidised bed: process optimisation by enriched air. *Fuel Process Technol* 2009; 90:677–85. <https://doi.org/10.1016/j.fuproc.2008.12.007>.
- [63] Lv P, Chang J, Xiong Z, Huang H, Wu C, Chen Y, Zhu J. Biomass air-steam gasification in a fluidized bed to produce hydrogen-rich gas. *Energy Fuel* 2003;17: 677–82. <https://doi.org/10.1021/ef0201811>.
- [64] Bhaskar T, Sera A, Muto A, Sakata Y. Hydrothermal upgrading of wood biomass: influence of the addition of K₂CO₃ and cellulose/lignin ratio. *Fuel* 2008;87: 2236–42. <https://doi.org/10.1016/j.fuel.2007.10.018>.
- [65] Li XT, Grace JR, Lim CJ, Watkinson AP, Chen HP, Kim JR. Biomass gasification in a circulating fluidized bed. *Biomass Bioenergy* 2004;26:171–93. [https://doi.org/10.1016/S0961-9534\(03\)00084-9](https://doi.org/10.1016/S0961-9534(03)00084-9).
- [66] Mansaray KG, Ghaly AE, Al-Tawee AM, Hamdullahpur F, Ugursal VI. Air gasification of rice husk in a dual distributor type fluidized bed gasifier. *Biomass Bioenergy* 1999;17:315–32. [https://doi.org/10.1016/S0961-9534\(99\)00046-X](https://doi.org/10.1016/S0961-9534(99)00046-X).
- [67] Abbas A, Ansumali S. Global potential of rice husk as a renewable feedstock for ethanol biofuel production. *Bioenergy Res* 2010;3:328–34. <https://doi.org/10.1007/s12155-010-9088-0>.
- [68] Liakakou ET, Vreugdenhil BJ, Cerone N, Zimbardi F, Pinto F, André R, Marques P, Mata R, Girio F. Gasification of lignin-rich residues for the production of biofuels via syngas fermentation: comparison of gasification technologies. *Fuel* 2019;251: 580–92. <https://doi.org/10.1016/j.fuel.2019.04.081>.
- [69] Horvat A, Pandey DS, Kwapinska M, Mello BB, Gómez-Barea A, Fryda LE, Rabou LPLM, Kwapinski W, Leahy JJ. Tar yield and composition from poultry litter gasification in a fluidised bed reactor: effects of equivalence ratio, temperature and limestone addition. *RSC Adv* 2019;9:13283–96. <https://doi.org/10.1039/CRA02548K>.
- [70] Breiman L. Random forests. *Mach Learn* 2001;45:5–32. <https://doi.org/10.1023/A:1010933404324>.
- [71] Lei C, Deng J, Cao K, Ma L, Xiao Y, Ren L. A random forest approach for predicting coal spontaneous combustion. *Fuel* 2018;223:63–73. <https://doi.org/10.1016/j.fuel.2018.03.005>.
- [72] Iskenderoğlu FC, Baltacıoğlu MK, Demir MH, Baldinelli A, Barelli L, Bidini G. Comparison of support vector regression and random forest algorithms for estimating the SOFC output voltage by considering hydrogen flow rates. *Int J Hydrogen Energy* 2020;45:35023–38. <https://doi.org/10.1016/j.ijhydene.2020.07.265>.
- [73] Wang L. Support vector machines: theory and applications. Berlin, Heidelberg: Springer Berlin Heidelberg; 2005. <https://doi.org/10.1007/b95439>.
- [74] Bansal S, Roy S, Larachi F. Support vector regression models for trickle bed reactors. 207–Chem Eng J 2012;208:822–31. <https://doi.org/10.1016/j.cej.2012.07.081>.
- [75] Ye J, Xu Y, Song X, Yu J. Numerical modelling and multi-objective optimization of the novel hydrocyclone for ultra-fine particles classification. *Chem Eng Sci* 2019; 207:1072–84. <https://doi.org/10.1016/j.ces.2019.07.031>.

- [76] Jia S, Qian X, Yuan X. Optimal design for dividing wall column using support vector machine and particle swarm optimization. *Chem Eng Res Des* 2017;125: 422–32. <https://doi.org/10.1016/J.CHERD.2017.07.028>.
- [77] Xiao G, Ni M, Chi Y, Jin B, Xiao R, Zhong Z, Huang Y. Gasification characteristics of MSW and an ANN prediction model. *Waste Manag* 2009;29:240–4. <https://doi.org/10.1016/j.wasman.2008.02.022>.
- [78] Xing J, Luo K, Wang H, Gao Z, Fan J. A comprehensive study on estimating higher heating value of biomass from proximate and ultimate analysis with machine learning approaches. *Energy* 2019;188:116077. <https://doi.org/10.1016/j.energy.2019.116077>.
- [79] Tavares R, Monteiro E, Tabet F, Rouboa A. Numerical investigation of optimum operating conditions for syngas and hydrogen production from biomass gasification using Aspen Plus. *Renew Energy* 2020;146:1309–14. <https://doi.org/10.1016/J.RENENE.2019.07.051>.
- [80] Casalicchio G, Molnar C, Bischl B. Visualizing the feature importance for black box models, lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics). 11051 LNAI 2018: 655–70. https://doi.org/10.1007/978-3-030-10925-7_40.
- [81] Antoniadis A, Lambert-Lacroix S, Poggi JM. Random forests for global sensitivity analysis: a selective review. *Reliab Eng Syst Saf* 2021;206:107312. <https://doi.org/10.1016/j.ress.2020.107312>.
- [82] Brockmann D, Morgenroth E. Evaluating operating conditions for outcompeting nitrite oxidizers and maintaining partial nitrification in biofilm systems using biofilm modeling and Monte Carlo filtering. *Water Res* 2010;44:1995. <https://doi.org/10.1016/J.WATRES.2009.12.010>. –2009.
- [83] Rose KA, Smith EP, Gardner RH, Brenkert AL, Bartell SM. Parameter sensitivities, Monte Carlo filtering, and model forecasting under uncertainty. *J Forecast* 1991; 10:117–33. <https://doi.org/10.1002/FOR.3980100108>.
- [84] Cadini F, Zio E, Avram D. Monte Carlo-based filtering for fatigue crack growth estimation. *Probabilist Eng Mech* 2009;24:367–73. <https://doi.org/10.1016/J.PROBENGMECH.2008.10.002>.
- [85] Jha MS, Dauphin-Tanguy G, Ould-Bouamama B. Particle filter based hybrid prognostics for health monitoring of uncertain systems in bond graph framework. *Mech Syst Signal Process* 2016;75:301–29. <https://doi.org/10.1016/J.YMSSP.2016.01.010>.
- [86] Razavi S, Jakeman A, Saltelli A, Prieur C, Iooss B, Borgonovo E, Plischke E, Io Piano S, Iwanaga T, Becker W, Tarantola S, Guillaume JHA, Jakeman J, Gupta H, Melillo N, Rabitti G, Chabridon V, Duan Q, Sun X, Smith S, Sheikholeslami R, Hosseini N, Asadzadeh M, Puy A, Kucherenko S, Maier HR. The Future of Sensitivity Analysis: an essential discipline for systems modeling and policy support. *104954 Environ Model Software* 2021;137. <https://doi.org/10.1016/j.envsoft.2020.104954>.
- [87] Guangul FM, Sulaiman SA, Raghavan VR. Gasification and effect of gasifying temperature on syngas quality and tar generation: a short review. *AIP Conf Proc* 2012;1440:491. <https://doi.org/10.1063/1.4704254>.
- [88] Tian T, Li Q, He R, Tan Z, Zhang Y. Effects of biochemical composition on hydrogen production by biomass gasification. *Int J Hydrogen Energy* 2017;42:19723–32. <https://doi.org/10.1016/J.IJHYDENE.2017.06.174>.
- [89] Mohammed MAA, Salmiaton A, Wan Azlina WAKG, Mohammad Amran MS, Fakhrul-Razi A. Air gasification of empty fruit bunch for hydrogen-rich gas production in a fluidized-bed reactor. *Energy Convers Manag* 2011;52:1555–61. <https://doi.org/10.1016/j.enconman.2010.10.023>.
- [90] Dai X, Wu C, Li H, Chen Y. The fast pyrolysis of biomass in CFB reactor. *Energy Fuel* 2000;14:552–7. <https://doi.org/10.1021/ef9901645>.
- [91] Guo F, Dong Y, Dong L, Guo C. Effect of design and operating parameters on the gasification process of biomass in a downdraft fixed bed: an experimental study. *Int J Hydrogen Energy* 2014;39:5625–33. <https://doi.org/10.1016/j.ijhydene.2014.01.130>.
- [92] van der Drift A, van Doorn J, Vermeulen JW. Ten residual biomass fuels for circulating fluidized-bed gasification. *Biomass Bioenergy* 2001;20:45–56. [https://doi.org/10.1016/S0961-9534\(00\)00045-3](https://doi.org/10.1016/S0961-9534(00)00045-3).
- [93] Tsekos C, Tandurella S, de Jong W. Estimation of lignocellulosic biomass pyrolysis product yields using artificial neural networks. *J Anal Appl Pyrolysis* 2021;157: 105180. <https://doi.org/10.1016/j.jaatp.2021.105180>.
- [94] AlNouss A, McKay G, Al-Ansari T. Production of syngas via gasification using optimum blends of biomass. *J Clean Prod* 2020;242:118499. <https://doi.org/10.1016/J.JCLEPRO.2019.118499>.