



# Predicting biomass composition and operating conditions in fluidized bed biomass gasifiers: An automated machine learning approach combined with cooperative game theory

Jun Young Kim <sup>a,\*</sup>, Ui Hyeon Shin <sup>b</sup>, Kwangsu Kim <sup>c</sup>

<sup>a</sup> School of Chemical Engineering, Sungkyunkwan University, 2066 Seobu-Ro, Jangan, Suwon, 16419, Republic of Korea

<sup>b</sup> Department of Artificial Intelligence, Sungkyunkwan University, 2066 Seobu-Ro, Jangan, Suwon, 16419, Republic of Korea

<sup>c</sup> College of Computing and Informatics, Sungkyunkwan University, 2066 Seobu-Ro, Jangan, Suwon, 16419, Republic of Korea

## ARTICLE INFO

Handling Editor: Krzysztof (K.J.) Ptasinski

**Keywords:**

Machine learning  
Automated machine learning  
SHAP analysis  
Cooperative game theory  
Biomass gasification  
Fluidized bed

## ABSTRACT

A fluidized bed system for biomass gasification is considered a useful implementation with advantages of good heat and mass transfer and uniform heat across the bed that can reduce local hot spot which produces heavy hydrocarbons including tar. Such a system has been studied for decades for product conversion and selectivity optimization with empirical correlations, mathematical modelling, and machine learning. Although machine learning has been recently adopted for predicting biomass composition and operating conditions, the number of algorithms chosen from previous research is limited due to various types of models and complexity. An automated machine learning (AutoML) was adopted here to select the best machine learning algorithm amongst various types of models including tree ensembles and neural networks. Using AutoML, operating conditions and lignocellulosic compositions were predicted with output features from the system, including syngas composition, LHV, char yield, and tar yield. Generally, CatBoost (gradient boosting on decision trees) algorithm showed a good match with experimental data results/test data with high  $R^2$  (0.689) and low RMSE (0.220). Combined cooperative game theory (Shapely additive explanation, SHAP) was also applied to develop an interpretable model. AutoML combined with the SHAP algorithm for explainable machine learning was first tried in the field of fluidized bed systems to find a suitable machine learning model for each feature with hyperparameters optimization and expected to help in interpreting the results with limited results that require exhaustive experimentation. The results can be widely adapted in various scenarios, such as monitoring the process as a soft sensor.

## 1. Introduction

After the European Union and the United Nations Framework Convention on Climate Change (UNFCCC) participating countries ratified the Paris agreement, there has been active carbon footprint reduction in many fields, including food production, transportation, and energy production with political regulation. Especially, to reduce carbon emissions in energy production, power generation systems integrated with carbon capture and storage (CCS) technologies or replacing energy sources with renewable energy sources are generally considered [1,2].

Amongst renewable energy sources to replace fossil fuels, biomass is one of the prospective energy sources due to its abundant resources and carbon-neutral characteristics [3]. Besides, power generation facilities

using biomass have been studied for over 50 years. Now they can be applied at the industrial level that can contribute to  $\text{CO}_2$  emission reduction [4]. Thermal conversion processes can also convert biomass into valuable liquid and/or gas products which can be further used as a fuel source for power generation. Although biomass pyrolysis has advantages of producing gas (e.g., CO, H<sub>2</sub>, CH<sub>4</sub>) and liquid products (e.g., oil and naphtha), residual carbons such as tar and char which may reduce the performance of the desired product are considerably produced as well [5]. Gasification, on the other hand, produces mainly on gas products (*i.e.*, syngas) with a small amount of char and tar [6].

Three biomass gasification types are generally categorized: fixed bed, fluidized bed, and entrained bed gasifiers [7–9]. As particles are fluidized in both fluidized bed and entrained bed reactor, these reactor

\* Corresponding author.

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

types are capable of controlling uniform temperature across the entire bed and decreasing the amount of tar formation [10–15]. Especially, the gasification temperature of the entrained bed gasifier could reach 1400 °C with a pressure of up to 70 bar which could accelerate tar cracking [16]. However, gasification operating under high temperatures requires a finely fed biomass material with a diameter less than 0.1–0.4 mm, which is unsuitable for most biomass feedstock such as bark or sawdust [17]. Thus, only a fluidized bed gasifier was considered in this study.

In a fluidized bed gasifier, compositions of syngas and char/tar yield are the key factors depending on biomass compositions, gasifying agent, and operating conditions of the fluidized bed reactor. Previously, these features have been estimated and/or predicted as key factors in biomass gasifiers by empirical correlations and modellings. Marda et al. proposed a correlation between the methanol conversion and the ratio of CO to CO<sub>2</sub> for the bio-oil gasification to syngas at 650 °C [18]. Sansaniwal et al. developed a mathematical model for the syngas composition with mass balance and heat balance [19]. Mansaray and Tawee developed mathematical modeling of a fluidized bed rice husk gasifier with bed height, fluidization velocity and equivalent ratios to estimate mole fractions of gas components and higher heating value [20]. Fatoni et al. developed a biomass fluidized gasifier considering hydrodynamics and reaction kinetics and compared it with a thermodynamic simulation using ASPEN [21]. Natale et al. developed a population balance model for the particle size distribution in wood gasification considering both mechanical breakage and reaction [22]. However, the models and empirical correlations which are based on the experimental findings require a comprehensive understanding of the fluidized bed gasifier, especially hydrodynamics of bubble/particle mixing and reaction kinetics of biomass conversion and/or devolatilization [23,24]. In order to understand the interactions between the biomass compositions, operating conditions, and product while circumventing the need for various understandings of the reactor, machine learning can be useful to intuitively find these relationships.

Several attempts for key feature optimization with machine learning (ML) have been tried by previous researchers. Souza et al. proposed an artificial neural network (ANN) to correlate the operating conditions of the biomass gasifier (steam to biomass ratio and temperature) with syngas and hydrocarbon gases [25]. Kardani et al. used five different machine learning models applied with an optimized ensemble model for predicting lower heating value (*LHV*) and syngas yield [26]. Mutlu and Yucel used machine learning classification instead of regression, and postulated a suitable syngas composition prediction with classification accuracies of 96% and 89% for random forest and support vector machine, respectively [27]. Previous research indicates that machine learning can be used as a data-driven soft sensor replacing traditional hard sensors and serving as backup units [28]. However, predicting the operating conditions (e.g., steam to biomass ratio, biomass composition, and temperature) via outcome products (e.g., syngas composition) with machine learning approaches often may not be very useful since some products are highly influenced by multivariable input features. For example, the H<sub>2</sub> composition in the syngas is mainly influenced by the steam to biomass ratio, lignocellulosic biomass composition, and equivalence ratio. Thus, developing machine learning models with an appropriate strategy is needed for high accuracy.

Applying an automated machine learning (AutoML) framework instead of trying several machine learning model algorithms is gaining much attention as state-of-the-art ML techniques grow in sophistication. It is becoming more difficult even for an ML expert to incorporate all recent best practices into their modelling [29]. AutoML frameworks can remove many barriers to deploying high-performance ML models. They also offer the best ML algorithm only once, including model selection strategies, hyperparameter tuning, feature engineering, and data splitting/processing [30]. Several AutoML frameworks have been developed in past decades, including auto-sklearn [31], TPOT [32], Auto-WEKA [33], and AutoGluon [29]. Surprisingly, although AutoML has been

studied in past decades and developed for both novices and experts with excellent performance, machine learning modeling in the field of fluidized bed systems and biomass gasification tried AutoML has not been reported yet. Thus, we chose AutoGluon as an AutoML framework known to be faster, more robust, and much more accurate than previously proposed AutoML (i.e., auto-sklearn, TPOT and Auto-WEKA) for tabular data for machine learning modelling in this study.

Another consideration for machine learning is to understand and explain the optimized machine learning results because machine learning is usually a ‘black box’ model. Thus, explanation of modeling results is merely comparing the statistic indicators among different machine learning algorithms. For explainable machine learning, Shapley additive explanation (SHAP) based on a combined cooperative game theory is a powerful mathematical tool. The huge advantage of using SHAP is that it highlights essential feature importance analysis. Rather than just simply highlighting essential features affecting output features, SHAP can analyze interactions for any pair of all features considered in machine learning [34].

Based on the above background, the objective of this research is to predict input features from the produced output features including syngas composition from a biomass fluidized gasifier to monitor and optimize the biomass types and operating conditions as a soft sensor in the fluidized bed biomass gasifier. First, types of biomass feedstocks are classified by their lignocellulosic compositions. To decipher subtle patterns in multivariable and complex datasets, machine learning regression models were adopted for the prediction. To the best of our knowledge, this is the first paper adopting the AutoML framework combined with the SHAP method to find a suitable and explainable machine learning approach from tabular data much faster and more accurately in the field of biomass gasifiers. The optimized modeling can be directly used as a soft sensor serving as a backup unit in the fluidized bed biomass gasifier to predict the operating conditions and types of biomass from the produced syngas.

## 2. Materials and methods

### 2.1. Dataset collection

Biomass gasifier performances with different reactor types (i.e., fixed, fluidized, or entrained bed) show significant differences, affecting the temperature gradient, particle size for the feedstock, and tar formation [35]. Therefore, datasets were carefully collected from various reliable studies in the literature, focusing on experimental results tested in fluidized beds. At least five datasets with the same biomass types operated in steady-state conditions were also collected for biomass type prediction by lignocellulosic composition.

The lignocellulosic composition (*Cell.:Hem.:Lignin* by wt. percentage), gasification temperature, pressure, equivalence ratio (*ER*), steam to biomass ratio (*SBR*), and superficial gas velocity (*U<sub>g</sub>*) significantly affect syngas composition and tar/char yield. Data collected from references were conducted in a steady-state operation. Note that the geometries of the experimental unit from the datasets collected are different. However, all these data were commonly obtained from fluidized bed systems which can explain the various phenomena such as hydrodynamics and biomass conversion despite the difference in reactor size. Besides, normalized bed height and normalized radial position of the particles are affecting hydrodynamics rather than the reactor height and diameter [36]. The selected data features and their ranges are provided in Table 1. Since the objective of the machine learning approach is to estimate biomass composition and operating conditions with gas product and tar/char yield, input features and output features are set for the machine learning model. Here, biomass fuels include both woody biomass and agricultural residues. Furthermore, syngas compositions are collected without N<sub>2</sub> and dry basis. Details of the datasets are provided in Supplementary Information 1. Overall, 353 datasets were prepared for the machine learning modelling. Catalysts used in the

**Table 1**

Input and output feature ranges of the constructed database from previous literature for machine learning.

Input features	Ranges
Syngas composition [vol%]	
H <sub>2</sub>	8.21–61.70
CO	5.16–48.24
CO <sub>2</sub>	12.9–65.0
CH <sub>4</sub>	3.18–60.75
Lower heating value [MJ/m <sup>3</sup> ]	1.16–14.55
Char yield [wt.%]	0–50.9
Tar yield [g/m <sup>3</sup> ]	0–97.1
Output features	Ranges
Biomass composition [wt.%]	
Cellulose ( <i>Cell.</i> )	0.20–0.58
Hemicellulose ( <i>Hem.</i> )	0.08–0.74
Lignin	0.03–0.49
Temperature ( <i>T</i> ) [°C]	600–1042
Pressure ( <i>P</i> ) [bar]	1–10
Equivalence ratio ( <i>ER</i> ) [-]	0–0.52
Steam to biomass ratio ( <i>SBR</i> ) [-]	0–4.04
Superficial gas velocity ( <i>U<sub>g</sub></i> ) [m/s]	0.02–2.35

dataset are limited to tar removal during biomass gasification in the bubbling fluidized bed [37]. Although details of these tar removal catalyst mechanisms may be different, we focused on the role of tar removal during gasification in the fluidized bed system, not specific reaction kinetics or catalytic sites on tar removal since these were outside the scope of this study.

A schematic diagram on the methodological flowchart of machine learning is given in Fig. 1. To predict the biomass compositions and operating conditions in biomass fluidized gasifiers via machine learning modelling, the input features, including syngas composition and char/tar yield were used from the given experimental dataset. Six different classical machine learning models, including random forests (RandomForest) weighted ensemble (WeightedEnsemble), extremely randomized trees (ExtraTree), k-nearest neighbors (KNeighbors), boosted trees (CatBoost, lightGBM, XGBoost), and deep neural networks (NeuralNetFastAI, NeuralNetTorch), were adopted with different impurity index (i.e., Entropy (referred to as Entr) and Gini). Considering the main scope and importance on this research is focused on AutoML combined

with the SHAP method, mathematical explanations and architectural details on each machine learning model adopted here are not explained in this paper. These classical algorithm explanations are already sufficiently explained elsewhere [34,38]. The hyperparameter was tuned with k-fold cross-validation for optimization. After the optimization, the machine learning models were ranked with the highest prediction accuracy ( $R^2$  and RMSE) using Python with AutoGluon. Details of the mechanism that AutoGluon finds optimized machine learning algorithm are given in Section 2.2 below. Here, a machine learning algorithm with the regression method was used for predicting output features (biomass composition, pressure, equivalence ratio, steam to biomass ratio, and superficial gas velocity). Types of biomass feedstocks were also predicted based on their lignocellulosic compositions. Similar to the regression mechanism, machine learning classifiers were optimized, and each classification model was validated and ranked based on precision accuracy for each feature.

## 2.2. AutoGluon for machine learning prediction process

As briefly explained in the Introduction, the AutoML framework was applied for machine learning modelling due to its simplicity and high accuracy. Among the AutoML frameworks developed previously, most of them exclusively focused on the task of combined algorithm selection and hyperparameter optimization (CASH), selecting the best models and their optimized hyperparameters from the given dataset [39]. However, the CASH algorithm is generally known as computationally expensive as this algorithm search is intractable (i.e., nonconvex optimization). It optimizes the best model from huge possibilities, even evaluating poor models and hyperparameters. The AutoGluon, however, can automatically recognize the data type in each feature for robust data preprocessing without classical hyperparameter optimization and shows excellent performance amongst AutoML approaches [29]. Here, the Python software (ver. 3.10.4) and Autogluon tabular models were used to fit tabular data frames using ‘autogluon.tabular.TabularPredictor’.

A simplified diagram for the automated ML for predicting data and the AutoGluon framework for preprocessing is given in Fig. 2. First, data preprocessing was applied. There are two sequential preprocessing steps in the AutoGluon – model agnostic and model specific. Model-agnostic

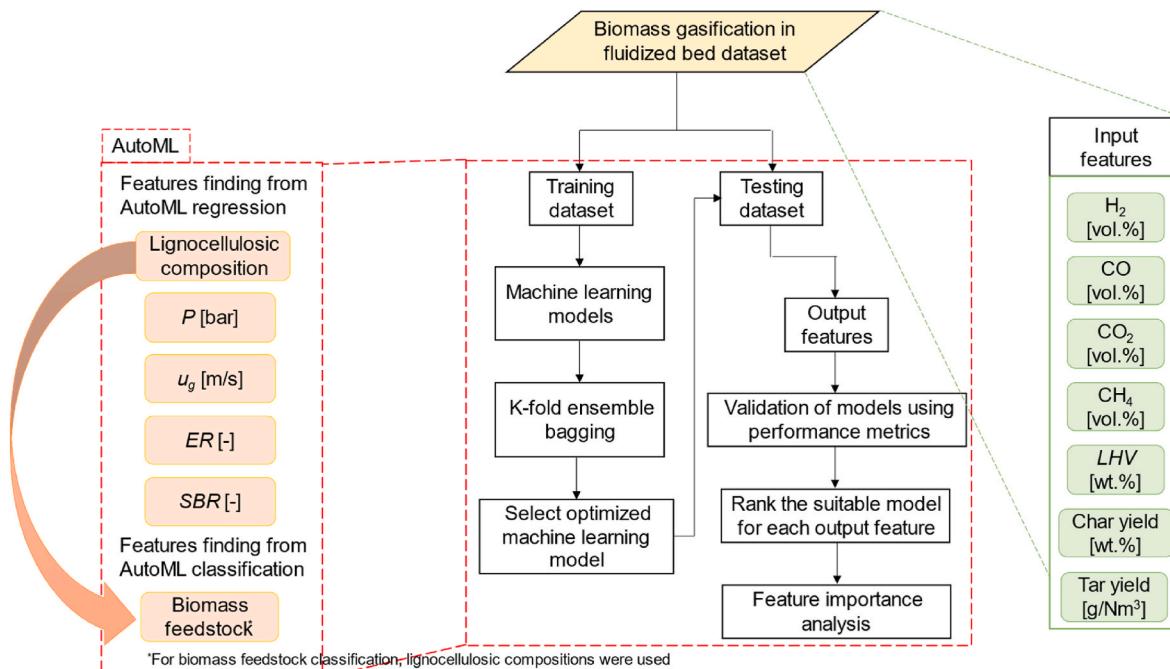
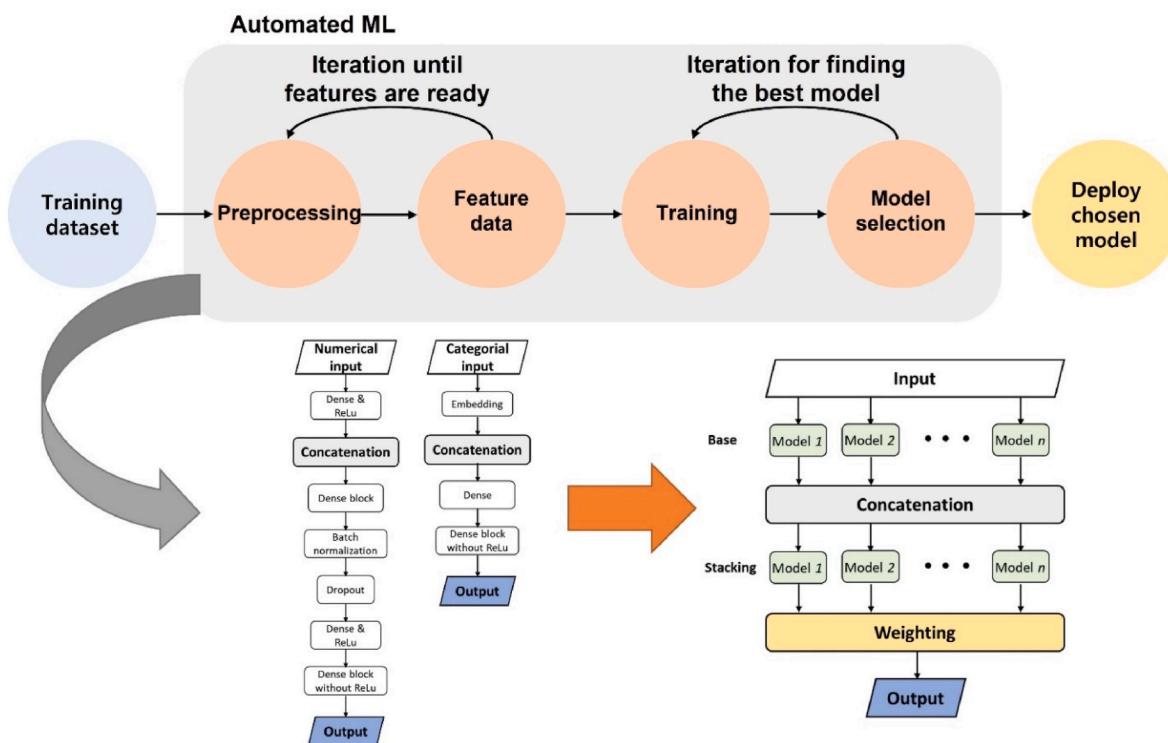


Fig. 1. Methodological Flowchart of this study.



**Fig. 2.** Schematic diagram of automated ML for predicting data and preprocessing algorithm from AutoGluon. Processing algorithms were adopted from Erikson et al. [29].

preprocessing transforms the input features to all models and automatically categorizes numerical or categorical data with a neural network architecture as shown in Fig. 2. The neural network for the model-agnostic preprocessing applies a separate embedding layer to each categorical feature from the multivariate data before data are concatenated with other variables. Here, different models such as KNN (K-Nearest Neighbor), ensemble models, and complex neural network are trained on the same dataset and stacked. The outputs of these models are then entered into a linear model to obtain the output which is the weighted sum of the outputs where the weights are obtained by training. Secondly, repeated k-fold bagging originated from the k-fold cross-validation is used. Multiple weights and data blocks with different initial values may be used. The outputs of these models are then averaged to reduce the variance of the prediction. After the data inputs are transformed into all models, the model-specific preprocessing is applied to a copy of the data used to train a particular model. A multi-layer stack ensemble is used here, combining the outputs and data of multiple models for another stacking. In other words, multiple models are trained initially, and then a linear model is used to produce the output. In order to prevent overfitting, k-fold bagging can be also used here.

As both regression and classification are used here, the rationale for the difference between the machine learning regressor and the classifier needs to be explained. In brief, classification predicts a category whereas regression predicts the quantity in machine learning. In other words, the machine learning classifier approximates the input features to the discrete output categories for a given observation. Here, the probability or accuracy can be interpreted as the likelihood or confidence of the target belonging to each category. Conversely, the machine learning regressor approximates the input features to continuous output features. Thus, the machine learning regressor can predict the quantity. Statistical indicators showing errors of the model from the prediction such as root mean squared error (RMSE) and  $R^2$  are generally used:

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

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

where  $y_{i,data}$  and  $\bar{y}_{data}$  are the individual observed value and the average observed value 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.

### 2.3. SHAP for explainable machine learning

To assess the prediction capability of the machine learning models, several simple statistical methods such as Pearson correlation have been used for the algorithm interpretation. However, these methods cannot interpret the complex relationships in the data and sometimes may distort the results. Instead of using these simple statistical methods for indirect interpretation on machine learning results, model-agnostic interpretability tools are implemented and developed, which can quantify the feature importance [40]. Permutation importance is also a common and reliable technique that can measure variable importance by observing the effect on model accuracy by arbitrarily shuffling each predictor variable [41]. However, the permutation importance only describes the importance of each feature without providing more explanations because the permutation importance calculation is based on the decrease in each input feature performance. On the other hand, SHAP is based on the magnitude of feature attributes. It combines feature importance with feature effects. Thus, SHAP was used to interpret and explain the ML model predictions.

As briefly explained above, SHAP is based on the combined cooperative game theory approach, which assigns each importance value to each feature at a particular prediction, along with local accuracy, missingness, and consistency as three parameters [42]. Lundberg and Lee proposed SHAP from the additive feature attribution method considering three properties: local accuracy, missingness, and consistency [42]. In detail, the additive feature attribution method has an

explanation model denoted as  $g$ , which explains the complex original model,  $f(x)$ , based on a single input  $x$ . Here, a linear function of binary variables is used for  $g$ :

$$g(x') = \varphi_0 + \sum_{i=1}^M \varphi_i x'_i = f(x) \quad (3)$$

where  $x'$  is a simplified input for an original model that maps original inputs through a mapping function,  $x = h_x(x')$ .  $M$  is the number of simplified input features.  $\varphi_0 = f(h_x(0))$  represents the model output with all simplified inputs missing. Through Eq. (3),  $g(x')$  can represent the local accuracy of the original model at datapoint of  $x$ . To meet the missingness property,  $x'_i = 0$  has no attributed impact, which refers to  $\varphi_i = 0$ . For the consistency property,  $z' \setminus \{i\}$  is defined when  $z'_i = 0$  and  $f_x(z') = f(h_x(z'))$  for any two models  $f$  and  $f'$ , if:

$$f_x(z') - f_x(z' \setminus \{i\}) \geq f_x(z') - f_x(z' \setminus \{i\}) \quad (4)$$

With all inputs being  $z' \in \{0, 1\}^M$ , then Eq. (4) implies  $\varphi_i(f', x) \geq \varphi_i(f, x)$ . Particularly, if a model changes with a simplified input by either increasing or staying the same regardless of other inputs, the attribution of that input should not be lower than the value with the other inputs [43]. Following the definition of additive feature attribution with three properties, the SHAP (Shapely additive explanation) value for the explanation model  $g$  is defined as:

$$\varphi_i(f, x) = \sum_{z' \subseteq x' \setminus \{i\}} \frac{|z'| (M - |z'| - 1)!}{M!} (f_x(z' \cup \{i\}) - f_x(z')) \quad (5)$$

where  $|z'|$  is the number of non-zero entries in  $z'$ .  $z' \subseteq x' \setminus \{i\}$  represents all the non-zero entries  $z'$  which are a subset of the non-zero entries in  $x' \setminus \{i\}$ .  $|z'| (M - |z'| - 1)! / M!$  is the fraction of permutations with the feature  $\{i\}$ .  $(f_x(z' \cup \{i\}) - f_x(z'))$  is the marginal contribution of the feature  $\{i\}$  to coalition  $z'$ . Eq. (5) follows combined cooperative game theory results [43]. For the model agnostic feature importance test, the kernel SHAP was used in this work.

#### 2.4. 'Uncertainty' in machine learning

Many researchers who are not familiar with machine learning studies often confuse the 'uncertainty' of the machine learning approach. Moreover, some researchers undervalue machine learning research because the majority of results and discussions are explained with statistical analysis. However, in order to understand machine learning modeling, then statistics are crucial in understanding the relevance between the input and output features. Again, this 'underestimation' or 'misunderstanding' is from the lack of knowledge on machine learning, especially what is the uncertainty in this kind of research dealing with and evaluating algorithms. However, the explanation on this is usually dismissed in the research literature because this is textbook knowledge in the field of machine learning. Here, it is noteworthy to explain the types of uncertainty corresponding to the machine learning algorithm, because this manuscript is focused on predicting biomass composition and operating conditions using datasets from the fluidized bed biomass gasifiers using the AutoML with kernel SHAP method, thus some readers may have no or little background on the machine learning. To the best of our knowledge, this is the very first paper in the fields of biomass gasifier and fluidization engineering that fully explains these main sources of uncertainty in machine learning.

There are three factors of uncertainty from the machine learning modeling: 1) noise in the collected datasets, 2) incomplete prediction range in the dataset, and 3) valid model selection [44]. Type 1 is originated from the dataset collection, mostly attributed to a lack of sufficient data. We have mentioned above (see Section 2.1) how the datasets were carefully collected to explain the type 1 uncertainty. For type 2, the

repeated k-fold bagging originated from the k-fold cross-validation is used (see Section 2.2). In detail, we aim to collect suitably representative random samples of observations to train and evaluate a machine learning model. However, as we are training and evaluating the machine learning model based on the collected dataset, the data must be split into training and validation sets, thus the whole dataset for training is not allowed. In other words, there will always be some unobserved cases no matter how well the models are generalized because some data must be remained for predicting the trained machine learning model. Thus, the resampling method such as k-fold cross-validation is used so that the one-fold is treated as validation once and then used for training k-1 times. Such a method is to handle the uncertainty in the representativeness of the prediction range setup. For type 3 uncertainty, statistical indicators including  $R^2$  and RMSE (see above) are used to choose the suitable model from each optimized machine learning model in AutoGluon. Specifically,  $R^2$  and RMSE were used in this study to investigate the suitable machine learning algorithm optimized from AutoGluon. Furthermore, we used the kernel SHAP to improve the credibility of the automatically optimized models from AutoML. Furthermore, the kernel SHAP can evaluate the machine learning model algorithm considering the physical meaning.

### 3. Results and discussion

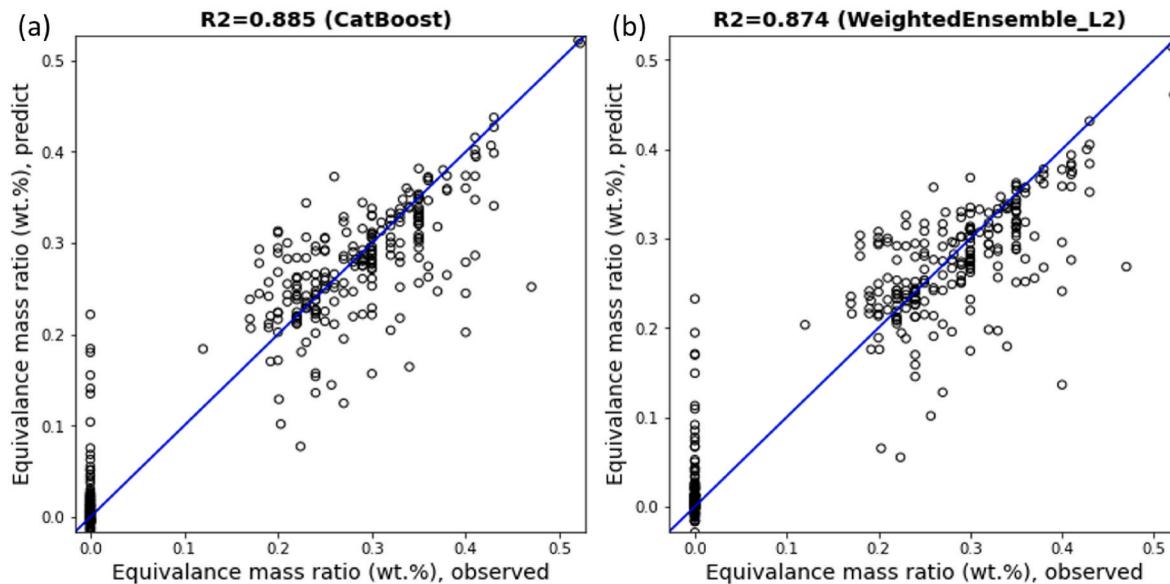
#### 3.1. Model evaluation and selection

The prediction accuracy data with the top five machine learning models in terms of average  $R^2$  and RMSE for input features are listed in Table 2. Note that each machine learning model is ranked by the average values of  $R^2$  and RMSE calculated by AutoGluon. The other performance results of machine learning are listed in Supplementary Information 2 Figures S1-S5, Tables S1 and S2. In terms of the average value of  $R^2$ , both CatBoost and WeightedEnsemble\_L2 showed the highest accuracy. Considering both  $R^2$  and RMSE, CatBoost was the best-performing machine learning algorithm in this dataset. Among the input features, ER showed the highest accuracy for all machine learning algorithms listed in Table 2, with  $R^2$  ranging from 0.860 to 0.886 and RMSE ranging from 0.051 to 0.055. Fig. 3 shows the prediction data of ER compared to the observed (test) dataset using CatBoost and WeightedEnsemble L2. The distribution of most data was on the parity plot, referring to the perfect prediction. These results indicate that the ER is the best feature to predict with the input features (i.e., syngas compositions, LHV, char and tar yield) using these two machine learning algorithms. Also, the ER, one of the significant variables for operating the fluidized biomass gasifier, can be well predicted by the final products after the gasification.

Following the ER, temperature also showed high  $R^2$  and low RMSE values. Temperature is another significant parameter affecting the syngas composition, LHV, char yield, and tar yield experimentally shown by previous researchers [14,45–47]. Superficial gas velocity ( $U_g$ ) showed the lowest  $R^2$  and the highest RMSE for all machine learning algorithms, with  $R^2$  ranging from 0.377 to 0.472 and RMSE ranging from 0.269 to 0.292. These results may be related to the indirect relationships of  $U_g$  with syngas composition, LHV, char yield, and tar yield.  $U_g$  affects the hydrodynamics of the system in a biomass fluidized bed gasifier, and may change the compositions of the syngas, as well as char and tar formation since the superficial gas velocity is related to the space velocity of the biomass [45]. However, only  $U_g$  itself cannot determine the syngas production. Compared to other features of gasification such as lignocellulosic composition, SBR, and temperature,  $U_g$  is more heavily dependent on the particle properties including size and density which are known to affect hydrodynamics [48,49]. As the reaction kinetics of biomass gasification is significantly affected by the temperature, biomass feedstock, and gasifying agent, the importance of  $U_g$  may not be high enough to affect the gas composition and char/tar yield. The feature importance and effect of  $U_g$  will be further treated in Section 3.2. The  $R^2$  and the RMSE values of the lignocellulosic compositions using

**Table 2**Top five machine learning models with the highest performance using average  $R^2$  and RMSE as statistical indicators.

Models	CatBoost		WeightedEsemble_L2		ExtraTreeMSE		NeuralNetFastAI		LightGBMLarge	
Indicators	$R^2$	RMSE	$R^2$	RMSE	$R^2$	RMSE	$R^2$	RMSE	$R^2$	RMSE
Cell.	0.753	0.044	0.774	0.042	0.714	0.048	0.720	0.047	0.667	0.051
Hem.	0.753	0.052	0.722	0.055	0.660	0.061	0.586	0.068	0.690	0.058
Lignin	0.684	0.049	0.675	0.050	0.671	0.050	0.626	0.053	0.641	0.052
Temp.	0.731	0.444	0.723	0.451	0.729	0.446	0.704	0.466	0.707	0.463
Pressure	0.651	0.471	0.704	0.434	0.629	0.485	0.659	0.465	0.570	0.522
ER	0.886	0.051	0.874	0.053	0.869	0.054	0.860	0.056	0.862	0.055
SBR	0.673	0.355	0.624	0.381	0.653	0.365	0.643	0.371	0.541	0.421
$U_g$	0.377	0.292	0.410	0.284	0.472	0.269	0.460	0.272	0.465	0.271
Average	0.689	0.220	0.688	0.219	0.675	0.222	0.657	0.225	0.643	0.237

Fig. 3. Parity plot of data distribution of observed  $ER$  vs. predicted  $ER$  for CatBoost and WeightedEnsemble\_L2 with  $R^2$  as a statistical indicator.

the top five models showing the highest performance were:  $R^2 = 0.667\text{--}0.774$  and  $\text{RMSE} = 0.042\text{--}0.051$  for cellulose (cell.),  $R^2 = 0.586\text{--}0.753$  and  $\text{RMSE} = 0.052\text{--}0.068$  for hemicellulose (hem.), and  $R^2 = 0.626\text{--}0.684$  and  $\text{RMSE} = 0.049\text{--}0.053$  for lignin. These variables (i.e., cell., hem., and lignin) are inextricably linked to each other to affect the syngas composition [50]. The relevance between these three variables will be further explained with the SHAP value in Section 3.2.

The classification accuracies for biomass feedstock types with machine learning algorithms are listed in Table 3. Similar to the performance results of the regression test shown above, WeightedEnsemble,

ExtraTrees, and CatBoost were the three models showing the highest performance for classification. Including these three models, machine learning using ensemble methods for optimization showed high accuracy performance. Because ensemble methods can achieve better performance with robustness than any single contributing model, the models applied to ensemble methods may show good accuracy for predicting the biomass feedstock types with the lignocellulosic compositions [51–53].

### 3.2. SHAP for feature importance and feature effects

Based on the features with the highest performance shown in Table 2, the SHAP values for the cellulose prediction based on each input feature (i.e., syngas composition,  $LHV$  and char/tar yield) with (a) CatBoost and (b) WeightedEnsemble\_L2 are plotted in Fig. 4. The order of the input features ( $LHV$ , syngas composition and char/tar yield) on the y-axis is from the highest to the lowest feature importance. Dots represent individual data points in the dataset. Color becomes red when each input feature positively affects more to predict cellulose, whereas blue represents the data when the input feature more negatively affects such prediction. Although CatBoost with cellulose as an output feature showed a slightly better performance than WeightedEnsemble\_L2 based on  $R^2$  and RMSE values ( $R^2 = 0.753$  and  $\text{RMSE} = 0.044$  for CatBoost and  $R^2 = 0.774$  and  $\text{RMSE} = 0.042$  for WeightedEnsemble\_L2, respectively), the results from these two models were similar based on the feature importance on the input feature. For both machine learning models,  $LHV$ ,  $CO_2$ , and tar yield were the three highest importance features.

**Table 3**  
Classification accuracy of each machine learning classifier algorithm according to biomass feedstock type.

ML models	Accuracy
WeightedEnsemble_L2	0.810
ExtraTreesEntr	0.810
ExtraTreesGini	0.805
CatBoost	0.793
RandomForestEntr	0.790
RandomForestGini	0.788
LightGBMXT	0.785
LightGBM	0.773
NeuralNetFastAI	0.765
LightGBMLarge	0.756
XGBoost	0.751
KNeighborsDist	0.717
NeuralNetTorch	0.700
KNeighborsUnif	0.620

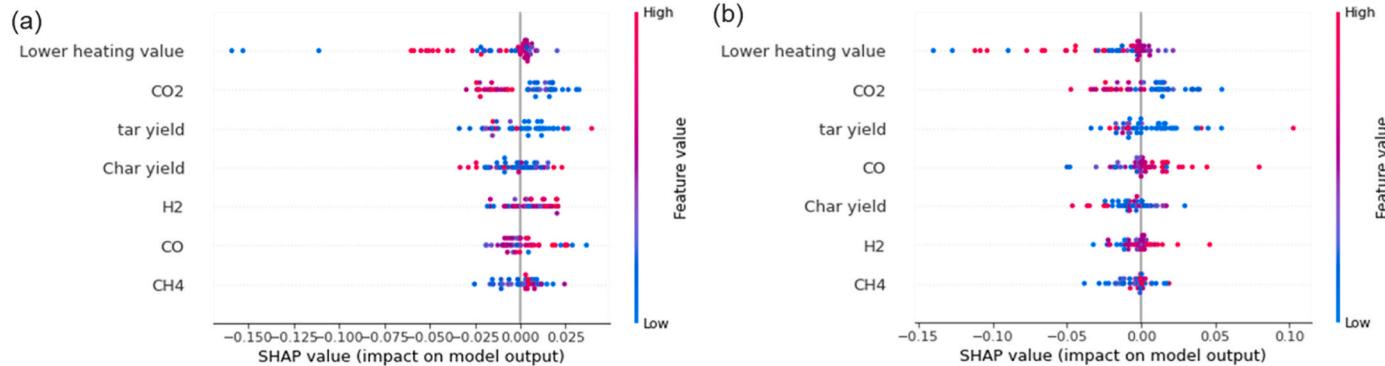


Fig. 4. SHAP value of the cellulose prediction with each input feature using (a) CatBoost and (b) WeightedEnsemble L2.

Another similarity between these two models is the dataset distribution of the *LHV* on cellulose. The cluster of positive data points using both CatBoost and WeightedEnsemble\_L2 showed values close to 0, or even slightly negative values, indicating that the *LHV* had mixed effects for predicting cellulose composition. As H<sub>2</sub>, CO, and CH<sub>4</sub> in the syngas affected the *LHV* whereas the CO<sub>2</sub> had no influence on the *LHV*, the *LHV* itself could not fully predict cellulose composition [54]. This has been experimentally proven by previous researchers [55,56]. On the other hand, CO<sub>2</sub> and tar yield clearly showed a positive or negative effect on cellulose composition. For both CatBoost and WeightedEnsemble\_L2, with low CO<sub>2</sub> composition and low tar yield, the SHAP value indicated that a higher cellulose composition was used for the biomass feedstock. As cellulose and hemicellulose are polysaccharides from sugar monomers, they are easier to be gasified compared to lignin. Specifically, lignin has different physico-chemical characteristics, with its aromatic polymers having high C/O ratio which favours char/tar formation during the gasification [57]. The SHAP value for the lignin also agreed well with the relationship between the lignin and the char/tar yield previously proven experimentally. In Fig. 5, the CatBoost and WeightedEnsemble\_L2 models explained that low char and tar yields were associated with a low lignin composition. Thus, the model prediction is well matched with the experimental data.

Among the operating conditions (i.e., temperature, pressure, *ER*, *SBR* and *U<sub>g</sub>*), *ER* and temperature showed the highest R<sup>2</sup> with CatBoost and WeightedEnsemble\_L2, indicating that the performances of these models were suitable. Figs. 6 and 7 show the SHAP values of the *ER* and temperature predictions with each input feature using (a) CatBoost and (b) WeightedEnsemble\_L2. Similar to the results shown in Figs. 4 and 5, the order of the feature importance for both *ER* and temperature between the CatBoost and WeightedEnsemble\_L2 was also similar, showing that both CatBoost and WeightedEnsemble\_L2 were suitable models. In Fig. 6, low CH<sub>4</sub> composition, low H<sub>2</sub> composition, and high CO<sub>2</sub>

composition were attributed to the high *ER* according to the SHAP value. During the gasification process, the *ER* is always less than 1. Increasing *ER* during gasification indicates that the biomass is prone to fully combust, producing more thermal energy than syngas production. Thus, the model can well explain the physical meaning of the temperature on the syngas composition during the gasification process [58]. In case of the temperature prediction, a clear trend attributed to the temperature was shown with CH<sub>4</sub> composition. As higher temperature is required for the combustion, these two models followed the physical meaning of the experimental results. The other prediction results based on each input feature with CatBoost and WeightedEnsemble\_L2 are plotted in Supplementary Information 2 Figures S6-S9.

As explained in Fig. 4, some input features showed co-dependent relationships, affecting the output feature. In order to find these relationships between the input features, SHAP dependence plots were used to explain the relationship between input features. Fig. 8 illustrates the relationship between syngas compositions affecting *ER* using CatBoost. *ER* was chosen because the highest R<sup>2</sup> was observed with CatBoost in Table 2. Also, since syngas compositions are reciprocally affecting each other, these dependence results were plotted. The color at the secondary y-axis indicates the value of another feature with which each input feature shows the highest interaction. Here, the positive value of SHAP means that the feature positively affects model predictions. From the dependence plot, higher *ER* predicts a decreasing composition of H<sub>2</sub> and an increasing composition of CO<sub>2</sub>. CO shows a concave down curve with the *ER*, indicating that the CO shows a mixed effect in predicting *ER*. The *ER* can decrease with higher CO because of biomass gasification. Although no clear relationship between H<sub>2</sub> and CH<sub>4</sub> was found in Fig. 8(a), decreasing CO<sub>2</sub> could increase CH<sub>4</sub> as shown in Fig. 8(c). These results are well matched with the gasification system. During the gasification, higher *ER* indicated that lower CH<sub>4</sub> and higher CO<sub>2</sub> were generated in the syngas composition because reactions are

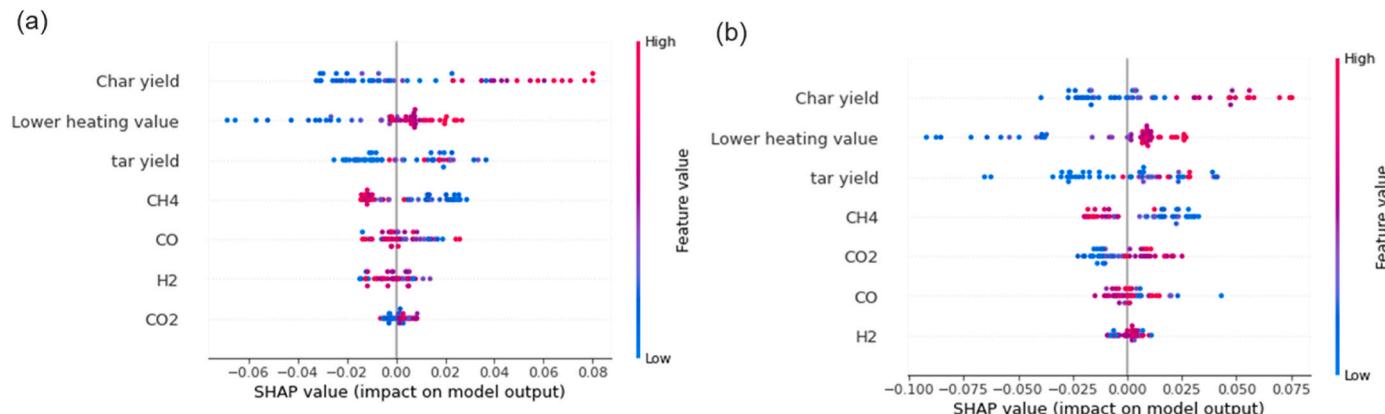
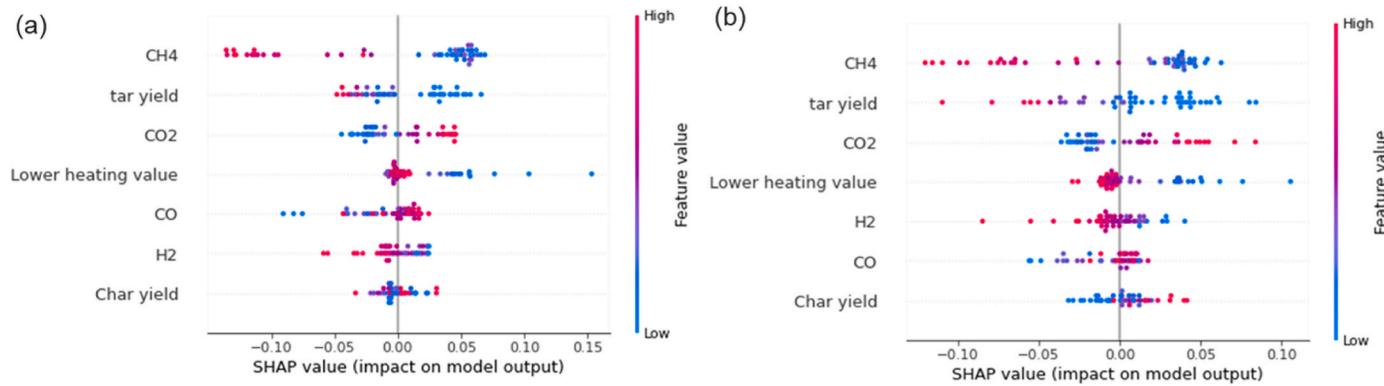
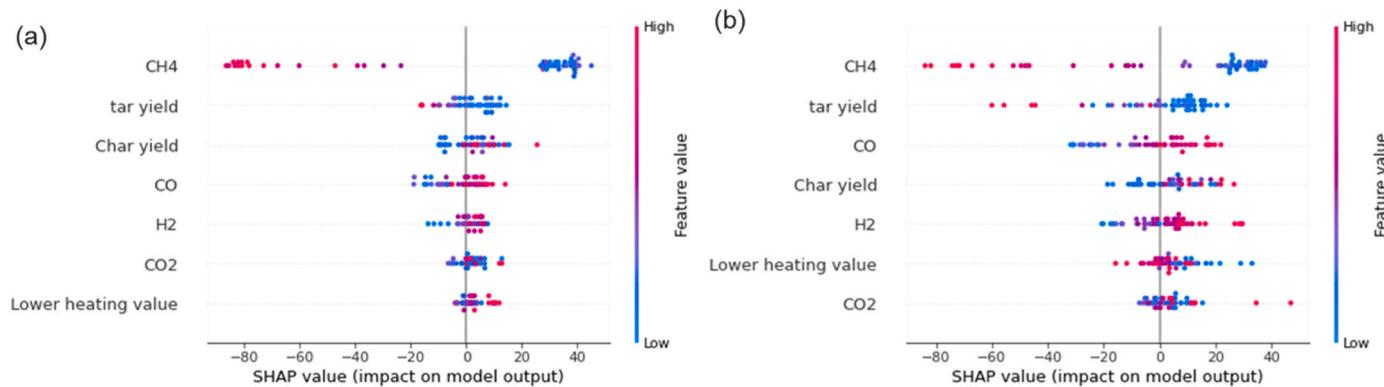


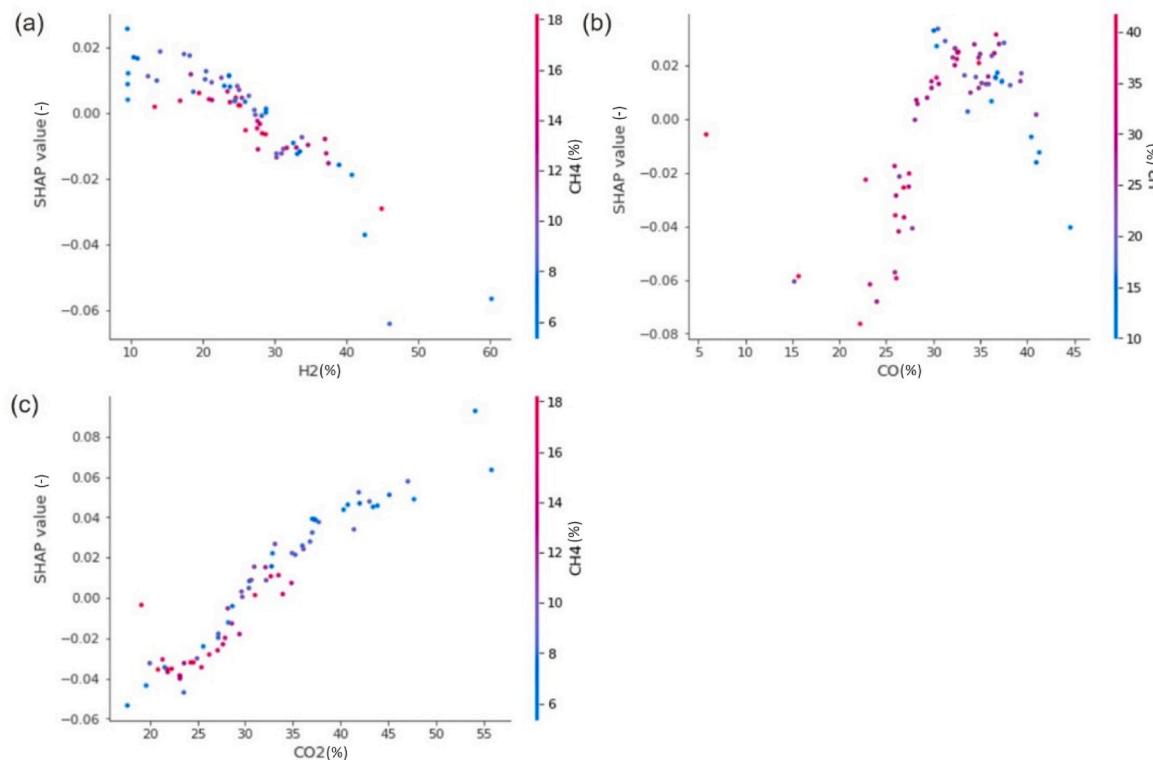
Fig. 5. SHAP value of the lignin prediction with each input feature using (a) CatBoost and (b) WeightedEnsemble\_L2.



**Fig. 6.** SHAP value of the ER prediction based on each input feature with (a) CatBoost and (b) WeightedEnsemble\_L2.



**Fig. 7.** SHAP value of the temperature prediction based on each input feature with (a) CatBoost and (b) WeightedEnsemble\_L2.



**Fig. 8.** SHAP dependence plots of syngas compositions affecting SHAP value for ER by CatBoost algorithm. (a) H<sub>2</sub> vs. CH<sub>4</sub>, (b) CO vs. H<sub>2</sub>, and (c) CO<sub>2</sub> vs. CH<sub>4</sub>. Dots represent each data point tested in this study.

move toward from gasification to combustion. In Fig. 8(b), the *ER* is decreased when the biomass is not likely to participate in either gasification or combustion [59]. The optimum *ER* according to this SHAP value is found when CO is approximately 35% and H<sub>2</sub> is 27%, which matches well with the optimum *ER* findings of experimental data at the temperature ranges of 700–1000 °C reported by Mohammed et al. [54].

#### 4. Conclusions

Based on the dataset from the fluidized bed biomass gasifiers with catalysts for tar elimination, AutoML was used to find the best suitable machine learning algorithm to predict the biomass composition and operating conditions with the syngas composition, *LHV*, and char/tar yield. Among machine learning algorithms, CatBoost and WeightedEnsemble\_L2 showed the highest R<sup>2</sup> (0.688 and 0.688, respectively) and the lowest RMSE (0.220 and 0.219, respectively) for the average value of overall input features. Also, the machine learning classification with AutoGluon was applied for the biomass feedstock. It was found that ensemble methods showed the highest accuracy. In order to explain such machine learning results, the cooperative game theory, i.e., SHAP, was used to analyze which input features were affected in a way to find the feature importance as well as the effect on the output feature prediction. From the SHAP value predicted with cellulose and lignin, CatBoost and WeightedEnsemble\_L2 showed similar feature importance and effect data, indicating that both machine learning models can well predict the desired biomass composition and operating conditions. SHAP dependence plots were adopted for co-dependent syngas compositions affecting the *ER*. Results were well matched with the previous experimental data. The cooperative game theory was decisive in setting the operating conditions of the biomass gasifier using a fluidized bed and in finding the optimum operating conditions based on the syngas and char/tar yield. The AutoML with cooperative game theory for explainable machine learning described here can be used by researchers to find biomass gasification conditions when only product dataset is available. In addition, the AutoML with SHAP method is a powerful tool which can be directly applied to the fluidized bed biomass gasifier system as a soft sensor to predict the biomass types and operating conditions from the syngas. We envision that this AutoML with SHAP method would give many researchers significant impact as a tool not only to optimize the machine learning effectively in various chemical engineering applications but also to utilize this method as a preliminary/fundamental research before modifying/upgrading machine learning algorithm.

#### Credit author statement

**Jun Young Kim:** conceptualization, methodology, formal analysis, writing-original draft, writing-review & editing, project administration, funding acquisition. **Ui Hyeon Shin:** software, methodology, validation, visualization. **Kwangsu Kim:** supervision.

#### Declaration of competing interest

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

#### Data availability

I have shared my datafile in the attached file

#### Acknowledgement

This study was supported by a grant (NRF-2018M3D3A1A01018009) from the National Research Foundation of Korea (NRF) funded by the Korea government. Also, this work was partly supported by Institute of Information & communications

Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No.2019-0-00421, AI Graduate School Support Program (Sungkyunkwan University)).

#### Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.energy.2023.128138>.

#### References

- [1] Xiang Y, Cai L, Guan Y, Liu W, He T, Li J. Study on the biomass-based integrated gasification combined cycle with negative CO<sub>2</sub> emissions under different temperatures and pressures. Energy 2019;179:571–80. <https://doi.org/10.1016/J.ENERGY.2019.05.011>.
- [2] Sung WC, Jung HS, Bae JW, Kim JY, Lee DH. Hydrodynamic effects on the direct conversion of syngas to methyl acetate in a two-stage fixed-bed/fluidized-bed combined reactor. J CO<sub>2</sub> Util 2023;69:102411. <https://doi.org/10.1016/J.JCOU.2023.102411>.
- [3] Abdellouahed L, Authier O, Mauviel G, Corriou JP, Verdier G, Dufour A. Detailed modeling of biomass gasification in dual fluidized bed reactors under Aspen Plus. Energy Fuels 2012;3840–55.
- [4] 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>.
- [5] Zhong H, Wei Z, Man Y, Pan S, Zhang J, Niu B, Yu X, Ouyang Y, Xiong Q. Prediction of instantaneous yield of bio-oil in fluidized biomass pyrolysis using long short-term memory network based on computational fluid dynamics data. J Clean Prod 2023;391:136192. <https://doi.org/10.1016/J.JCLEPRO.2023.136192>.
- [6] Cheng LB, Kim JY, Ebneayamini A, Li ZJ, Lim CJ, Ellis N. Thermodynamic modelling of hydrogen production in sorbent-enhanced biochar-direct chemical looping process. Can J Chem Eng 2022. <https://doi.org/10.1002/CJCE.24482>.
- [7] 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>.
- [8] 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>.
- [9] Ebneayamini A, Li ZJ, Kim JY, Grace JR, Lim CJ, Ellis N. Effect of calcination temperature and extent on the multi-cycle CO<sub>2</sub> carrying capacity of lime-based sorbents. J CO<sub>2</sub> Util 2021;49:101546. <https://doi.org/10.1016/j.jcou.2021.101546>.
- [10] 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>.
- [11] 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>.
- [12] 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>.
- [13] 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>.
- [14] Couto N, Rouboa A, Silva V, Monteiro E, Bouziane 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>.
- [15] Bae K, Kim JY, Go KS, Nho NS, Kim D, Bae JW, Lee DH. Bubble/micro-bubble regime transition in a pressurized bubble column of a low surface tension liquid system. Chem Eng Sci 2022;249:117191. <https://doi.org/10.1016/j.ces.2021.117191>.
- [16] Ruiz JA, Juárez MC, Morales MP, Muñoz P, Mendivil MA. Biomass gasification for electricity generation: review of current technology barriers. Renew Sustain Energy Rev 2013;18:174–83. <https://doi.org/10.1016/J.RSER.2012.10.021>.
- [17] Luo X, Wu T, Shi K, Song M, Rao Y. Biomass gasification: an overview of technological barriers and socio-environmental impact, gasification for low-grade feedstock. 2018. <https://doi.org/10.5772/INTECHOPEN.74191>.
- [18] Marda JR, DiBenedetto J, McKibben S, Evans RJ, Czernik S, French RJ, Dean AM. Non-catalytic partial oxidation of bio-oil to synthesis gas for distributed hydrogen production. Int J Hydrogen Energy 2009;34:8519–34. <https://doi.org/10.1016/j.ijhydene.2009.07.099>.
- [19] Sansaniwal 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>.
- [20] Mansaray AKG, Al-Taweel AM. Mathematical modeling of a fluidized bed rice husk gasifier: Part III - model verification. Energy Sources 2000;22:281–96. <https://doi.org/10.1080/00908310050014243>.
- [21] Fatoni R, Gajjar S, Gupta S, Handa S, Elkamel A. Modeling biomass gasification in a fluidized bed reactor. In: International conference on industrial engineering and operations management; 2014. p. 1047–56.

- [22] Natale G, Galgano A, di Blasi C. Modeling particle population balances in fluidized-bed wood gasifiers. *Biomass Bioenergy* 2014;62:123–37. <https://doi.org/10.1016/j.biombioe.2014.01.006>.
- [23] Kim JY, Li ZJ, Ellis NJ, Lim CJ, Grace JR. Model for attrition in sorption-enhanced chemical-looping reforming in fluidized beds. *Fuel Process Technol* 2021;213:106702. <https://doi.org/10.1016/j.fuproc.2020.106702>.
- [24] Pio DT, Tarelho LAC. Empirical and chemical equilibrium modelling for prediction of biomass gasification products in bubbling fluidized beds. *Energy* 2020;202:117654. <https://doi.org/10.1016/J.ENERGY.2020.117654>.
- [25] de Souza MB, Couceiro L, Barreto AG, Quítete 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>.
- [26] Kardani N, Zhou A, Nazem M, Lin X. Modelling of municipal solid waste gasification using an optimised ensemble soft computing model. *Fuel* 2021;289:119903. <https://doi.org/10.1016/j.fuel.2020.119903>.
- [27] Mutlu AY, Yucel O. 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>.
- [28] Yuan X, Ou C, Wang Y, Yang C, Gui W. A novel semi-supervised pre-training strategy for deep networks and its application for quality variable prediction in industrial processes. *Chem Eng Sci* 2020;217:115509. <https://doi.org/10.1016/j.ces.2020.115509>.
- [29] Erickson N, Mueller J, Shirkov A, Zhang H, Larroy P, Li M, Smola A. AutoGluon-tabular: robust and accurate AutoML for structured data. 2020. <http://arxiv.org/abs/2003.06505>.
- [30] Ge P. Analysis on approaches and structures of automated machine learning frameworks. In: Proceedings - 2020 international conference on communications, information system and computer engineering, CISCE, 2020; 2020. p. 474–7. <https://doi.org/10.1109/CISCE50729.2020.00106>.
- [31] Feurer M, Klein A, Jost KE, Springenberg T, Blum M, Hutter F. Efficient and robust automated machine learning. *Adv Neural Inf Process Syst* 2015;28.
- [32] Olson RS, Urbanowicz RJ, Andrews PC, Lavender NA, Kidd LC, Moore JH. Automating biomedical data science through tree-based pipeline optimization. *Lect Notes Comput Sci* 2016;9597:123–37. [https://doi.org/10.1007/978-3-319-31204-0\\_9/FIGURES/5](https://doi.org/10.1007/978-3-319-31204-0_9/FIGURES/5).
- [33] Thornton C, Hutter F, Hoos HH, Leyton-Brown K. Auto-WEKA: combined selection and hyperparameter optimization of classification algorithms. In: Proceedings of the ACM SIGKDD international conference on knowledge discovery and data mining; 2012. p. 847–55. <https://doi.org/10.48550/arxiv.1208.3719>. Part F128815.
- [34] Lundberg SM, Erion G, Chen H, DeGrave A, Prutkin JM, Nair B, Katz R, Himmelfarb J, Bansal N, Lee SI. From local explanations to global understanding with explainable AI for trees. *Nat Mach Intell* 2020;2(1):56–67. <https://doi.org/10.1038/s42256-019-0138-9>. 2 (2020).
- [35] 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).
- [36] 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>.
- [37] Horvat A, Pandey DS, Kwapisinska M, Mello BB, Gómez-Barea A, Fryda LE, Rabou LPLM, Kwapisinski 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>.
- [38] Zhang Y, Ning H. Irony and stereotype spreaders detection using BERT-large and AutoGulon. In: CLEF; 2022. p. 2746–52. <http://ceur-ws.org>.
- [39] Zöller MA, Huber MF. Benchmark and survey of automated machine learning frameworks. *J Artif Intell Res* 2021;70:409–72. <https://doi.org/10.48550/arxiv.1904.12054>.
- [40] 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, [https://doi.org/10.1007/978-3-030-10925-7\\_40](https://doi.org/10.1007/978-3-030-10925-7_40); 2018. 655–670.
- [41] Breiman L. Random forests. *Mach Learn* 2001;45:5–32. <https://doi.org/10.1023/A:1010933404324>.
- [42] Lundberg SM, Allen PG, Lee S-I. A unified approach to interpreting model predictions. <https://github.com/slundberg/shap>. [Accessed 18 October 2022].
- [43] Onsree T, Tippayawong N, Phithakkitnukoon S, Lauterbach J. Interpretable machine-learning model with a collaborative game approach to predict yields and higher heating value of torrefied biomass. *Energy* 2022;249:123676. <https://doi.org/10.1016/J.ENERGY.2022.123676>.
- [44] Murphy KP. Machine learning: a probabilistic perspective. 2012. [http://noiselab.ucsd.edu/ECE228/Murphy\\_Machine\\_Learning.pdf](http://noiselab.ucsd.edu/ECE228/Murphy_Machine_Learning.pdf). [Accessed 8 March 2023].
- [45] 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>.
- [46] Pradhan P, Arora A, Mahajani SM. A semi-empirical approach towards predicting producer gas composition in biomass gasification. *Bioresour Technol* 2019;272:535–44. <https://doi.org/10.1016/J.BIOTECH.2018.10.073>.
- [47] Serrano D, Golpouri 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>.
- [48] Fu S-Y, Feng X-Q, Lauke B, Mai Y-W. Effects of particle size, particle/matrix interface adhesion and particle loading on mechanical properties of particulate-polymer composites. *Compos B Eng* 2008;39:933–61. <https://doi.org/10.1016/J.COMPOSITESB.2008.01.002>.
- [49] 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>.
- [50] Liakkou ET, Vreugdenhil BJ, Cerone N, Zimbardi F, Pinto F, André R, Marques P, Mata R, Girão 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>.
- [51] Kim JY, Kim D, Li ZJ, Dariva C, Cao Y, Ellis N. Predicting and optimizing syngas production from fluidized bed biomass gasifiers: a machine learning approach. *Energy* 2023;263:125900. <https://doi.org/10.1016/J.ENERGY.2022.125900>.
- [52] Elmaz F, Büyükkıçakır B, Yücel Ö, Mutlu AY. Classification of solid fuels with machine learning. *Fuel* 2020;266:117066. <https://doi.org/10.1016/j.fuel.2020.117066>.
- [53] 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>.
- [54] Mohammed MAA, Salmiato A, Wan Azlina WAKG, Mohammad Amran MS, Fakhru’L-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>.
- [55] 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).
- [56] Herguido J, Corella J, González-Saiz J. Steam gasification of lignocellulosic residues in a fluidized bed at a small pilot scale. Effect of the type of feedstock. *Ind Eng Chem Res* 1992;31:1274–82. <https://doi.org/10.1021/ie00005a006>.
- [57] Baruah J, Nath BK, Sharma R, Kumar S, Deka RC, Baruah DC, Kalita E. Recent trends in the pretreatment of lignocellulosic biomass for value-added products. *Front Energy Res* 2018;6:1–19. <https://doi.org/10.3389/fenrg.2018.00141>.
- [58] Silva IP, Lima RMA, Santana HEP, Silva GF, Ruzene DS, Silva DP. Development of a semi-empirical model for woody biomass gasification based on stoichiometric thermodynamic equilibrium model. *Energy* 2022;241:122894. <https://doi.org/10.1016/J.ENERGY.2021.122894>.
- [59] Meng F, Meng J, Zhang D. Influence of higher equivalence ratio on the biomass oxygen gasification in a pilot scale fixed bed gasifier. *J Renew Sustain Energy* 2018;10:053101. <https://doi.org/10.1063/1.5040130>.